Journal of Experimental Psychology:
Learning, Memory, and Cognition
1994, Vol. 20, No. 6, 1322–1340

0278-7393/94/$3.00

Acquisition of Procedural Skills From Examples

John R. Anderson and Jon M. Fincham

Three experiments were run in which Ss first memorized examples of input–output pairs and then generated the outputs for a series of new inputs by analogy to the original examples. Ss first performed these mappings by explicit analogy to an example, but with practice they learned to make these input–output mappings directly without reference to the examples. Ss sped up as a power function of practice over a day (Experiment 1) or days (Experiments 2 and 3). Ss developed a directional asymmetry such that they were slower to calculate the input from the output than the output from the input (whereas initially they had not been). Ss showed similar speed up in their ability to recall the original examples but did not show the same directional asymmetry. Initially, there was some transfer from practicing the procedure to recalling the examples, but this diminished over days.

In this article, we report research on the acquisition of procedural skills and their origins in declarative memory for examples. Many research efforts have provided subjects with specific experiences and have looked for the emergence of procedures that are more general than the training examples. For example, Chi, Bassok, Lewis, Reimann, and Glaser (1989), Novick and Holyoak (1991), Reed (1987, 1989), Ross (1984, 1987, 1989), and Ross and Kennedy (1990) have looked at development of explicit procedures from examples. Berry and Broadbent (1984) have looked at development of implicit procedures from solving example problems. The work of Reber (1967, 1989) could be conceived of in these terms, as can some of the work in implicit memory and amnesia (e.g., Schacter, 1987; Squire, 1992).

The ACT (Adaptive Control of Thought) theory has been predicated on a distinction between declarative and procedural knowledge since its inception (Anderson, 1976). In the ACT* theory (Anderson, 1983) a specific proposal was put forward for how procedural knowledge derived from declarative knowledge. The claim was that all knowledge first came into the system in a declarative form. For instance, one might memorize the side–side–side theorem in geometry. With practice at using the knowledge in a particular context, production rules would develop, which embodied a procedural form of that knowledge. This learning process was called *proceduralization.* A prerequisite for proceduralization was that the declarative knowledge first be committed to long-term memory.

This view that declarative knowledge was a prerequisite to procedural knowledge has been criticized. The results from patients with amnesia indicate that procedural knowledge can be acquired without the ability to acquire declarative knowledge. Broadbent (1989) also argued that the results from the Berry and Broadbent (1984) task contradicted the ACT analysis. Berry and Broadbent showed that subjects can learn to successfully manipulate a rule-based system but cannot consciously state the rules. They interpret this to mean that procedural knowledge is acquired without first going through a declarative stage. Similarly, Nissen and her associates (e.g., Nissen, Knopman, & Schacter, 1987; Willingham, Nissen, & Bullemer, 1989) argued that their data indicate that procedural knowledge can be acquired independent of declarative knowledge. Partly in response to these criticisms, we have developed in ACT a somewhat different conception of the relationship between declarative and procedural knowledge.

Learning by Analogy to Examples

The conception of the relationship between declarative knowledge and procedural knowledge has changed since Anderson (1983) and receives its clearest statement in Anderson (1993), which describes a version of ACT called ACT–R (Adaptive Control of Thought—Rational). There have been two main changes: The first concerns the declarative origins of procedural knowledge. Originally, the emphasis had been on declarative memory for instructions, but now the emphasis is on declarative memory for examples of how the procedures should be executed. Although not going so far as to deny that other types of declarative knowledge might be sources for procedures, the emphasis has shifted to learning from examples. It is argued that initial use of these examples involves analogy and that production rules are compiled that summarize the analogy process. The major reason for this shift of emphasis to examples has been the research with acquisition of academic (mathematics, science, and computer programming) problem-solving skills and the evidence that subjects make heavy reference to examples in their initial attempts to solve problems in these domains (e.g., Chi et al., 1989; Pirolli, 1991; Pirolli & Recker, 1994; Reder, Charney, & Morgan, 1986).

Much of our research on this topic has taken place with respect to acquisition of programming skills in the language LISP. In this domain a student might see an example such as (+ 91 712) and observe that it computes the sum 803. Suppose

John R. Anderson and Jon M. Fincham, Department of Psychology, Carnegie Mellon University.

This research was supported by Grant NSF 91-08529 from the National Science Foundation.

We thank Wayne Gray, Marsha Lovett, and Lael Schooler for their comments on this article.

Correspondence concerning this article should be addressed to John R. Anderson, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. Electronic mail may be sent to ja@cmu.edu.

the subject is then presented with (* 4 3) and must predict what this will compute. The analogy process in ACT–R will (a) map + onto *, 91 onto 4, and 712 onto 3, (b) retrieve that $4 \times 3 = 12$, and (c) predict 12 as the answer. The ACT–R theory holds that as a consequence of this analogy, a production rule would be formed which encoded the mapping as follows:

IF the LISP expression is (=operator =arg1 =arg2)
  and =value is the value of =operator applied to =arg1 and =arg2
THEN the LISP expression will evaluate to =value

where the terms prefixed by the equal sign (=) are variables explicitly identifying the terms mapped in the analogy. This process of forming a production to represent an analogy is called *knowledge compilation.*[1]

The second major change in the ACT–R conception of the transition from declarative to procedural knowledge concerns the long-term status of the declarative knowledge. It is assumed that the student is working from a declarative representation of the example. However, it is not essential that the representation be permanent and retrievable from long-term memory. All that is required is that it be active in working memory during the analogy process. As often as not, students actually look up examples in resources such as a textbook without ever committing the examples reliably to long-term memory. In such a case, the examples are being maintained in working memory by the environment and no long-term memory trace is necessary.

This move to the requirement of an active (but not necessarily permanent) declarative representation of examples is consistent with much of the data showing that procedures can be acquired without a corresponding declarative representation. There is evidence in the Berry and Broadbent (1984) task that subjects base their behavior on memory for specific examples (Maresaux, Dejean, & Karnas, 1990). Phelps (1989) and Squire and Frambach (1990) have reported success at teaching patients with amnesia to perform the Berry and Broadbent task only if the patients were given concurrent access to their judgments on previous trials and did not have to remember them. Also, in sequence extrapolation tasks it has been shown that when subjects are distracted from forming a declarative representation of the sequence by using a dual-task procedure, they fail to show any procedural learning (Willingham, Nissen, Bullemer, 1989).[2] There is also evidence that performance in the Reber (1967) task depends on memory for fragments from examples (Dulany, Carlson, & Dewey, 1984; Servan-Schreiber & Anderson, 1990) as predicted by Brooks' theory (1978, 1987).

It is too strong to argue that procedural knowledge can never be acquired without a declarative representation or that the declarative representation always has to be in the form of an example that is used in an analogy process. Nonetheless, the research does indicate that this is a major avenue for the acquisition of procedural knowledge. The purpose of this article is to test the details of ACT–R theory about how this transition should take place. In many places, we have argued that learning by analogy is a type of one-trial learning (Anderson, Conrad, & Corbett, 1989). On the first occasion that an opportunity arises for analogy, the production is learned and after that point it is used. This is a rather dramatic proposal for the transition between the two bases for behavior. There were two reasons for such a radical proposal. The first was that this was the easiest way to implement it in a computer simulation. The second was that there was some data that suggested a first-trial discontinuity.

Data From Tutors

Because ease of simulation does not hold much force, it becomes important to look critically at the evidence for a first-trial discontinuity in acquisition of procedural skills. This data came from work with intelligent tutoring systems for programming (Anderson et al., 1989) and geometry (Anderson, Bellezza, & Boyle, 1993), which allowed tracking of specific production rules. These tutoring systems involve a student model consisting of a set of production rules that are capable of performing the skill that is being taught. The tutors can segment the interactions of students at the computer terminals into units that correspond to specific production rules and measure the accuracy and speed associated with each specific production rule. For instance, the LISP tutor has a separate production rule corresponding to the coding of each function in LISP. A problem-solving segment corresponding to the coding of that function is attributed by the tutor to that production rule. We collected two measures of performance of such rules. The first is how many errors a student made in the coding episode corresponding to the rule (students were allowed up to three). Thus, the error measure varied from 0 to 3. The second is when students made no errors, we measured the time it took them to code this segment. This time was measured from when the system was ready to receive their input until they completed typing it.

Figure 1 presents examples of that data from the LISP tutor (Anderson et al., 1989). The data are averaged over 32 production rules. The figure presents performance on a specific rule as a function of the number of opportunities there have been to practice that rule. It is typical of the distribution of skills in an academic subject that only a few skills get much repeated practice, and many skills get little practice. Therefore, we aggregated later trials to reflect the decreasing number of contributing productions.

The data have been plotted on a log–log scale to expose the power law of learning that has been reported for most procedural skills (Newell & Rosenbloom, 1981). This law states that if one plots time (T) as a function of number of trials of practice (N), one gets a power function of the form:

$$T = aN^{-b}.$$

If one takes logarithms of both scales one gets a linear function:

$$\log T = \log a - b \log N.$$

[1] This discussion describes approximately the analogy process and production compilation in ACT–R. For a precise description see chapter 4 of Anderson (1993).

[2] Note that Cohen, Ivry, and Keele (1990) found that simpler sequences can be learned with a distractor task.

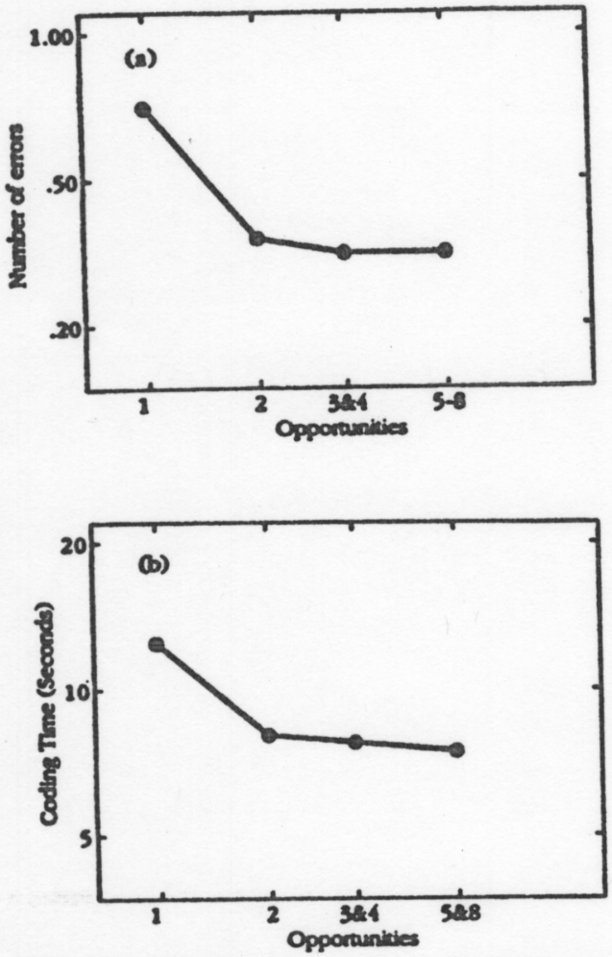

Figure 1. Data from the LISP tutor (Anderson et al., 1989): (a) shows the number of errors (maximum is 3) per production as a function of number of opportunities for practice; (b) shows the time to correctly code productions as a function of amount of practice.

Thus, a linear function on this log–log scale would be diagnostic of a power function. The latency[3] data appear to satisfy such a relationship except for the first point, which is decidedly higher than would be predicted by such a relationship. Most data on skill acquisition do not allow any definitive conclusions to be made about a possible first-point discontinuity. The first trial is often excluded because of warm up, averaged in with others, or quite noisy because a single observation is obtained per subject. It is a peculiarity of tutoring data that the first-trial opportunities provide the most data because all production rules are observed at least once (and usually twice). Similar first-trial discontinuities have been observed in the geometry tutor data (Anderson et al., 1993).[4]

This first-trial discontinuity is potentially significant with respect to the claim of single-trial learning of a production rule. The first trial would be when analogy is being used and later trials would depend on the production rule. The large improvement from first to second trial would reflect the compilation of the production rule, and the remaining power-law learning would reflect the accumulation of production strength. The ACT–R theory predicts that this strength accumulation should follow a power law.

Although the tutoring data have some unique advantages, one has to question the basing of such a strong conclusion on data collected within the tutoring paradigm. Students working with the tutor are in no way in a controlled experimental environment, and it is possible that the discontinuity may reflect something about that environment and not about the learning process per se. The most obvious possibility is that subjects might choose to look up examples the first time they need a rule and not at later times. Thus, the extra time on the first trial could reflect a search through a textbook. Therefore, we were motivated to pursue this issue in a more controlled environment.

In addition to these data from our tutors, a number of other researchers have presented data consistent with the proposal that subjects switch from example-based to rule-based problem solving after a single example. Pirolli (1985) found that subjects showed a dramatic drop in time spent referring to examples after solving a single problem. This suggests they were using rules and not examples after the first problem. Ross and Kennedy (1990) found a significant increase in generalizations after subjects solved a single problem by analogy. This suggests that subjects had extracted a problem-solving rule after solving one problem. Novick and Holyoak (1991) found that subjects' ability to state a rule after a problem-solving episode predicted ability to solve later problems independent of their success on the initial problem. This suggests that what is critical is the rule they extracted from the problem rather than the problem solution episode itself.

These experiments are all at least consistent with the suggestion that something special happens on the first trial of practicing a rule. Such a qualitative change is inconsistent with the continuous improvement implied by a power law. Despite all the data on power-law learning, nothing properly addresses the issue of the character of the function over the first few trials. There are two principal reasons for this empirical hole. First, because most experiments have relatively few first-trial observations, most reports of learning function average the first few trials together. Second, the relationship between what happens before practice and what happens in practice is not well controlled. Newell and Rosenbloom (1981) found it necessary often to estimate some amount of prior practice in fitting the learning curves and argued that in most experiments it was plausible to suppose some amount of prior practice before the experiment proper. We needed a situation in which the first trial of practice is the first opportunity on which the subject has had any opportunity to practice the skill. This required hiding the skill to be learned from the subject until that trial. This is what we did in our experiments. Thus, one goal of this research was to try to provide a relatively definitive characterization of the first few trials of the practice function.

[3] The error data seem to have the same form. In a later section, we discuss ACT–R predictions for the error data.

[4] Special statistics are gathered to deal with the disappearance of productions with serial position. Thus, the discontinuity is not due to the artifact of "harder" productions being concentrated in the first position. See Anderson et al. (1989) for details.

Declarative-Procedural Distinction: Asymmetry of Access

The assumption made earlier is that there is a change in the nature of the knowledge from procedural to declarative on the first trial. However, the experiments we referenced do not really show that there has been such a change in the nature of the knowledge. Indeed, the fact that subjects could verbally state the rule in the Novick and Holyoak (1991) study might be used to argue that subjects are switching from a declarative representation of the example to a declarative representation of the rule. The ACT–R theory predicts that there will be a significant difference between declarative and procedural knowledge in the way that it can be used. ACT–R also predicts an asymmetry in access to knowledge once it is proceduralized. For instance, suppose a subject formed the production rule (given earlier in this article) for evaluating arithmetic functions in LISP. Practicing it should improve ability to evaluate arithmetic functions such as

$$(/\ 42\ 7) = ?.$$

However, suppose a subject was presented with a piece of code such as

$$(/\ 42\ ?) = 6.$$

and must decide what the second argument should be. Although answering this question calls on the same knowledge as answering the evaluation problem, practicing evaluating should not improve answering this question because the evaluation production does not apply in this situation. This prediction has been tested in a number of contexts. Kessler (1988) and McKendree and Anderson (1987) have reported data supporting this prediction of no transfer among different procedures involving the same knowledge. More recently, however, Pennington and Nicolich (1991) have found some, but not total, transfer between the two tasks. The ACT–R theory does allow for some declarative transfer. The use of the knowledge in one way may practice or correct the declarative representation of the knowledge, which will facilitate performance when that declarative representation is used to compile a production for performance in the other direction. However, this new production cannot take any benefit from the other production and must be strengthened through its own practice.

ACT–R thus assumes that procedural knowledge is committed to a specific use, and one use cannot generalize to other uses. In contrast, declarative knowledge is flexible and can be used in any direction. The experiments discussed in this article are concerned with gathering evidence for this prediction of use specificity of procedural knowledge and whether it really is a feature that distinguishes it from declarative knowledge.

Experiment 1

We designed this experimental paradigm to expose the process of learning by analogy to examples. We felt that this was occurring in the tutoring environments, but we needed to study it in a more experimental context that would facilitate systematic manipulation and control as well as theoretical analysis. Inevitably, this meant transition to materials that are more artificial. However, we wanted material to which the ACT–R theory of analogy had a clear application. The key assumptions of the ACT–R theory of analogy are as follows.

1. The example can be decomposed into two parts, which are an antecedent and a consequent. In the case of the LISP example, the code for a function is the antecedent and the value of the function is the consequent.

2. ACT–R can extract some relationship that connects the antecedent and consequent. In the LISP example, the antecedent of (/ 42 7) is related to the consequent of 6 because $42 \div 7 = 6$. Usually, as in this case, this relationship involves retrieving some fact from memory.

3. The problem posed to the ACT–R analogy mechanism is either to find some antecedent that will produce a specified consequent or the consequent that will result from a specified antecedent. For instance, one can either be asked to write code to subtract 6 from 13 in LISP or be given the code (− 13 6) and predict how LISP will evaluate this.

4. ACT–R maps the part provided in the problem onto the corresponding part in the example and then tries to apply the relationship extracted from the example to predict the missing part in the problem.

Thus, our requirements for experimental materials are that they involve two distinguishable parts, that a relationship between the two be defined by some highly available long-term memory fact, and that this relationship can be extended by analogy to other examples. In addition, we wanted the rules that the subjects were learning to be equally novel to all subjects. Furthermore, we did not want the relationship to be immediately apparent to the subject on studying an example because we wanted the subject to perform the analogy only at time of test. Therefore, we came up with examples such as 35a44. Unknown to the subjects when they studied these examples, this exemplified the "a" rule, which was that the first digit on the left was incremented and the second digit decremented to get the number on the right. Thus, subjects might be tested with a problem such as 73a____ and they would have to respond 82. Or subjects might be tested with a problem such as ____a73 and they would have to respond 64.

We wanted to study subjects applying such rules over a sequence of trials and observe their learning functions for applying these rules. We wanted to make sure that the first trial was the one in which they first extracted the rule from the example and that there had not been covert practice of the rule in advance. Therefore, we had subjects memorize nine examples like those presented earlier that reflect the "a" through "i" rules but did not tell them they were rules. Rather, they were simply told this was an experiment in rote memorization. When they had successfully memorized the examples, they went to the next phase of the experiment in which we explained to them that the examples each reflected a rule, each rule involved a relationship between the digit on the left and the corresponding digit on the right, and that they would now be required to predict the number on one side given the number on the other side.

Method

Subjects. Twenty-four subjects were recruited from the Carnegie Mellon undergraduate population for an experiment that lasted 2 hr. They either received a base pay of $6 for participation in the experiment or 1 credit hour for a human subject requirement. In addition, all subjects received a bonus for their performance in the experiment. That bonus varied from $5 to $10.

Materials. The first digit on the left was either one less, the same, or one more than the first digit on the right. The similar relationships held between the second digits on the left and right. Combining these two factors created nine rules. Specific examples of the rules were created by randomly generating digits that exemplified these rules. All rule instances were unique for a rule and different from the original example.

Procedure. The experiment was run on a Macintosh II with a two-page black and white monitor. The experiment had four phases: an exposure phase, a dropout phase, a training phase, and a transfer phase. The subjects received the following instructions at the beginning of the exposure phase:

> In this experiment, you will be presented with a set of nine stimuli that you are required to memorize. Each stimulus will be a string of characters of the following form:
>
> **nnWnn**
>
> where each **n** is a numeric digit and the **W** will be some letter between *a* and *i*.
>
> Your goal is to memorize each of the nine strings in such a way that if you are given the letter, you can correctly recall all of the digits as they occurred on both sides of that letter in the studied string. During the memorization phase, you will be tested on the strings that you have studied. You will be presented with patterns similar to the following form:
>
> ___ ___**W**___ ___
>
> where **W** is a letter between *a* and *i*.

The first phase simply involved presenting the nine examples. The second phase of the experiment was a dropout learning phase in which the subjects committed to memory the nine examples. Subjects were prompted for their memory of an example by a letter, *a* through *i* , and had to recall (by typing) the two digits to the left and to the right. If the subject's recall was correct, that example was dropped out for that pass. Those examples that were incorrect were retested. When all examples had been correctly recalled, this dropout procedure was repeated for a second and third time.

The third phase was the main training phase. It involved 40 blocks of 9 trials, and in each block each rule instance was tested once. Thus, there were 360 trials in all. Within each block, rules were randomly ordered with the restriction that the last rule of one block could not be the first rule of the next block. Subjects were informed of the nature of the underlying rules by written instructions at the beginning of this phase. Half of the subjects were presented with a letter and a two-digit number on the left, and they were instructed to type the two-digit number for which that rule would create on the right. The other half of the subjects had to go right to left. The following are the instructions received by the subjects who went left to right:

> The *first* digit on the any given side was always one less than, the same as, or one more than the *first* digit on the opposite side. Similarly, the *second* digit on a given side was always one less than, the same as, or one more than the *second* digit on the opposite side. Thus, the nine different letters from *a* to *i* represent in random ordering the nine possible transformations of digits on one side of the letter to the digits on the opposite side (the nine transformations are: same,same; same,increase; same,decrease; increase,same; increase,increase; increase,decrease; decrease, same; decrease,increase; decrease,decrease).
>
> In this part of the experiment, we are going to test your ability to apply these rules you have learned. You will be presented with new numbers on the left-hand side of the letter representing the rule. These patterns will be similar to the following form:
>
> **nnW**___ ___
>
> where each **n** is a numeric digit and the **W** will be some letter between a and i, representing the appropriate rule to apply.
>
> You will need to type in the two digits on the right-hand side that would follow by applying the rule associated with the letter to the digits on the left-hand side. For example, had you memorized the following strings:
>
> **24m33**
> **68n78**
>
> and were tested with:
>
> **61m**___ ___
> **47n**___ ___
>
> you would type 70 as the response to the first test string since 7 is one more than 6 (as 3 was one more than 2 in the learned rule) and 0 is one less than 1 (as 3 was one less than 4 in the learned rule). Similarly, you would respond 57 to the second test string since 5 is one more than 4 (as 7 is one more than 6 in the learned rule) and 7 is the same as 7 (as 8 is the same as 8 in the learned rule).

When subjects made errors they were so informed, shown the correct answer, and given unlimited time to study it. Subjects were given points for their speed and accuracy. For each correct response 2 points were awarded, whereas for each incorrect response, the score was decremented by 10 points. An additional point was received for every half second the response time was under 5 s. Thus, if a response was correct and executed in 4 s, the score was incremented by 4 points. The points awarded and total number of trials were displayed after every trial. These points were converted into a dollar amount that was given as the bonus pay for each subject upon completion of the experiment. This served to keep subjects engaged and make the experience more enjoyable.

The fourth phase was the transfer phase. It consisted of 10 blocks in which each rule was tested in both directions. Again, rules were randomly ordered; there were 180 trials in all. The same payoff schedule held for these trials as in the training trials.

Results

Training data. Analyses of variance (ANOVAs) were performed for the training phase on the error rates and mean latencies for correct responses. Variables in the ANOVAs were whether subjects were trained left to right or right to left, rule (nine values), and serial position (eight values: Trials 1, 2, 3, 4–5, 7–10, 11–15, 16–25, and 26–40). We represented the first serial positions separately because we wanted trial-by-trial data for the first serial positions to assess whether there is a first trial discontinuity. Values are collapsed on later trials to allow for more even plotting on a log–log scale. There was no effect of direction of training, generate digits on the right from left or vice versa, $F(1, 22) = .13$ for latency, $MS_e = 793.94$; $F(1, 22) = 1.86$ for errors, $MS_e = .601$, nor did it participate in any significant interactions.[5] There were significant effects of trials, $F(7, 154) = 35.17$, $MS_e = 87.95$; $F(7, 154) = 12.08$, $MS_e = .104$ for errors for latency ($ps < .0001$).[6] We discuss effects of rule type in a later section.

The training results are shown in Figure 2 for latency and Figure 3 for error rate. The top half of each figure displays the data in normal coordinates and the bottom half displays the results in log coordinates. The abscissa in these graphs is number of trials on particular rules. The bottom half of Figure 2 shows that the latency data closely correspond to a power

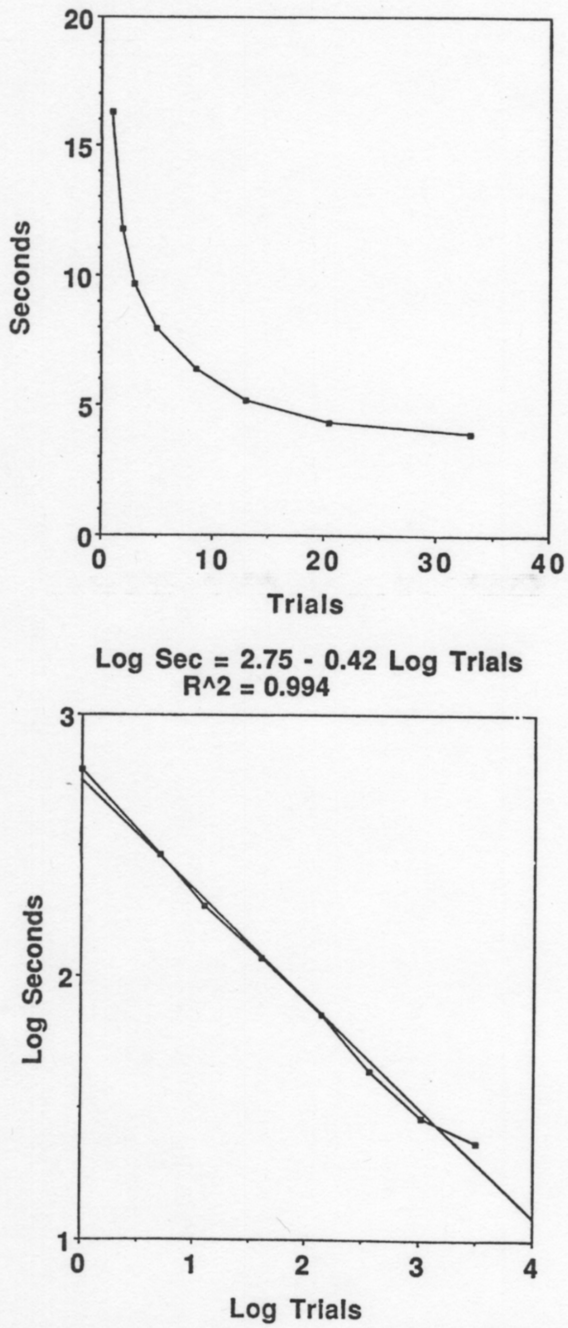

Figure 2. Latency as a function of trials of practice: The top half shows a normal scale; the bottom half shows log–log scale.

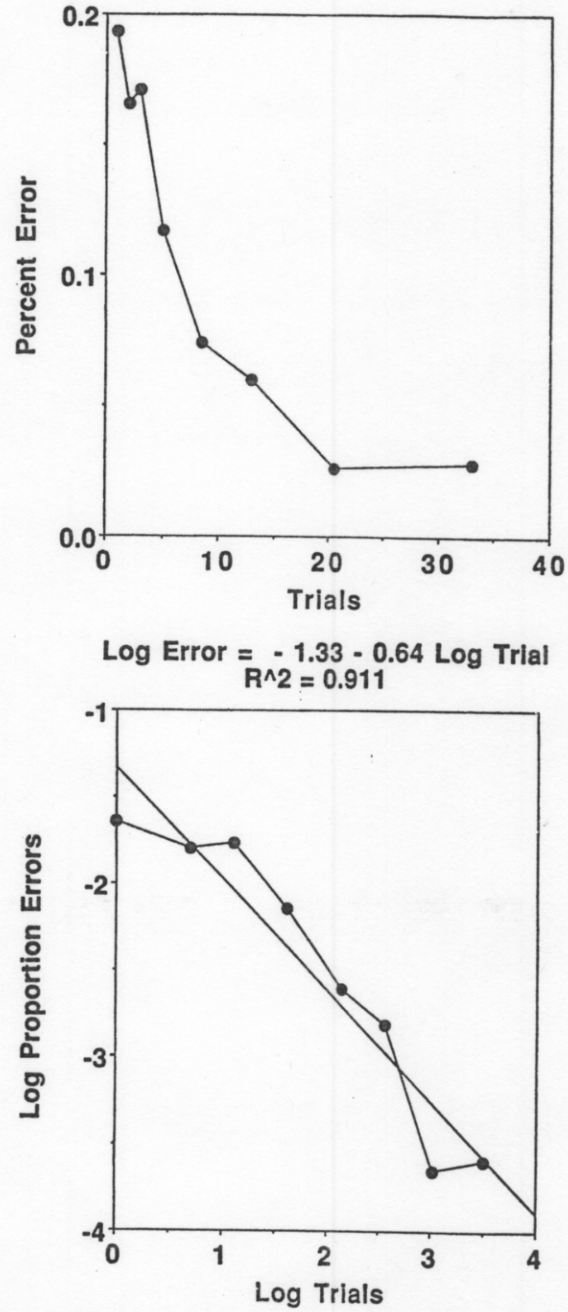

Figure 3. Proportion errors as a function of trials of practice: The top half shows a normal scale; the bottom half shows a log–log scale.

[5] No effects were expected. This is just a counterbalancing variable.

[6] Throughout we report standard probability values. No results change if we were to report the more stringent Geisser-Greenhouse probability values.

function and the bottom half of Figure 3 shows the error data at least approximate a power function. Neither measure shows any evidence for a first-trial discontinuity as had been found in the tutor data.

It is an interesting question why the error functions are even approximately power functions. The ACT–R theory predicts that one kind of error should decrease as a power function. These are retrieval errors where the subject fails to retrieve the correct example or rule on a particular trial. Such errors reflect long retrieval times and should obey a power function (see Anderson & Schooler, 1991). However, the problem is that this is not the only reason subjects may display errors. Particularly at the beginning of the experiment they may misunderstand instructions. There are other random sources of errors that can occur such as misreading problems or mistyping answers. There is no reason to suppose these sources obey power functions. We assume that the approximation to a power function reflects the contribution of the retrieval errors.

Most research on the power function has emphasized latencies and ignored error rates. In part, this is because at high levels of practice it is hard to detect effects in error rates. The general attitude is that as long as the error rates are low, or show no effect, or show the same general effects as the latencies, they can be ignored because there is no evidence of a speed–accuracy trade-off. Thus, usually the only concern is whether latency corresponds to power functions. We present the error data throughout this article, although we do not place as great an emphasis on these data because they involve a mixture of sources. In contrast, we assume as the rest of the field has that latencies of correct responses are rather pure measures of improvement in the performance of the skill and are better measures for testing the power law.

Transfer. We performed ANOVAs for the transfer phase on error rates and mean latencies for correct responses. Variables were direction of training, whether the test matched the training direction, rule type, and serial position (three values: 1–3, 4–6, 7–10). Effects of rule type are discussed later. There were no significant effects of direction of training and it did not enter into any significant interactions.

Figures 4 and 5 show the data with respect to the transfer phase of the experiment. The figures collapse together the first three trials on a particular rule, the next three, and the final four. Figure 4 displays latency data and Figure 5 displays the error rates. With respect to latency, there are significant effects of whether test direction matched study direction, $F(1, 22) = 39.9$, $p < .001$, $MS_e = 2.72$; trial of practice, $F(2, 44) = 13.47$, $p < .001$, $MS_e = 1.79$; and the interaction between the two, $F(2, 44) = 4.07$, $p < .05$, $MS_e = 1.19$. Subjects show a strong advantage when tested in the same direction as they had practiced, an improvement over trials, and some tendency for the directional advantage to diminish with trials as the subjects get training in the reverse direction. This directional asymmetry in latency is what ACT–R predicts for knowledge once it has been proceduralized.

With respect to errors, there is only a significant effect of whether test direction matched, $F(1, 22) = 9.36$, $p < .01$, $MS_e = .043$; and not of trial, $F(2, 44) = .35$, $MS_e = .018$; or their interaction, $F(2, 44) = .12$, $MS_e = .016$. Although all the significant latency effects are not significant in the error data,

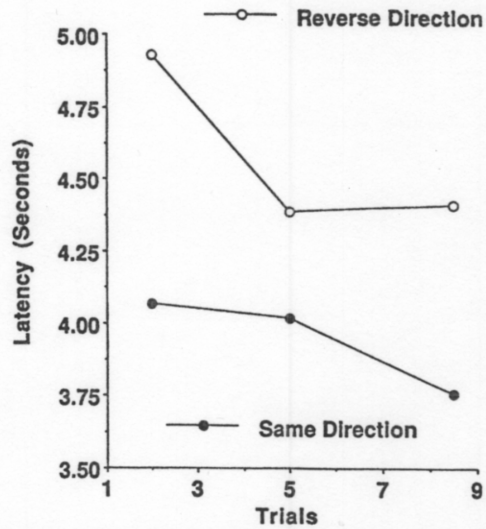

Figure 4. Latency in the transfer phase as a function of trials of practice.

the error data do not contradict any of the conclusions on the basis of the latency data. We present an analysis of what ACT–R's transfer predictions might be for errors in the *Results* section in Experiment 3.

Rule type. Finally, we were interested in whether the nine types of rules were of equal difficulty. The effects of rule type on latency were significant in both training, $F(8, 176) = 7.56$, $p < .061$, $MS_e = 106.17$; and transfer, $F(8, 176) = 53.45$, $p < .001$, $MS_e = 4.19$. Similarly, rule type had significant effects on errors both for training, $F(8, 176) = 5.27$, $p < .001$, $MS_e = .078$; and for transfer, $F(8, 176) = 4.76$, $p < .001$, $MS_e = .043$. Subjects found especially easy the rule that involved two identity operators and somewhat easier rules that involved one identity operator. The rule type also entered into a number of

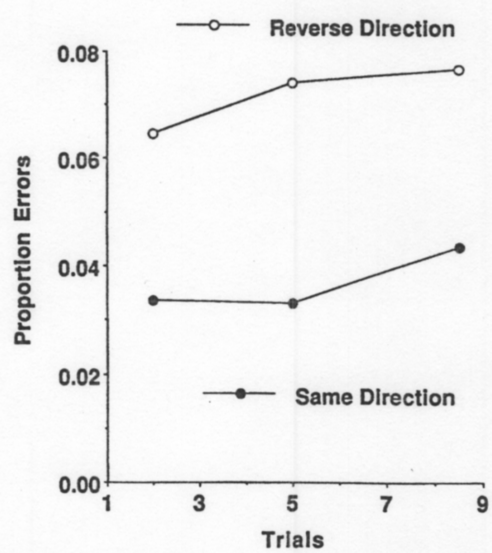

Figure 5. Proportion errors in the transfer phase as a function of trials of practice.

latency interactions such that the effects were smaller (but not reversed) in the case of an identity operator.

Discussion

With respect to the issues that motivated the experiment, we make a number of conclusions. First, there is no evidence in the learning function for any discontinuity associated with moving from the purported analogy-based processing on the first trial to a rule-based processing on later trials. This suggests that the discontinuities observed with the tutor data may reflect uncontrolled activities such as students looking up examples.

One conclusion from the lack of discontinuity might be that subjects never ceased to use the example analogically throughout the 40 trials. However, this is disconfirmed by the strong directional asymmetry that built up over those 40 training trials. Subjects were much faster in the same direction that they had practiced than they were in the reverse direction. Thus, they had built up something that was specific to the rule as they were using it. Also, some subjects spontaneously reported shifting from the example to a rule. However, there was no behavioral discontinuity associated with that change in performance: It seems to be a gradual change.

Although there was asymmetry in using the rule in the reverse direction, subjects were showing some transfer in using the rule in this new direction because they were much faster in their initial use of the rule in the new direction in the transfer phase than they were in their initial use of the rule in the original direction. Their average time on the first three trials was 4.93 s in the new direction during transfer, whereas it had been 11.38 s in original direction during training. It is unclear how much of this difference reflects increased facility with the general procedures of the experiment and increased practice of the examples. It is also unclear whether any of this advantage might reflect generalization in the use of the rule in one direction to the use of the rule in the reverse direction. This is one issue that we address in the subsequent experiments.

Experiment 2

We had a number of goals for the second experiment. First, we wanted to see whether we could replicate the results from the first experiment. Second, we wanted to get more data on the nature of the directional asymmetry that was observed in the first experiment. Third, we wanted to find a set of materials that would not display such large differences in the difficulty of individual rules. Although all rules contributed to all conditions in Experiment 1, we wanted to have a more homogenous rule set. Fourth, we wanted to investigate what would happen when we looked at a more extensive procedural practice.

We also wanted to deal with a potential criticism of the first experiment, which was that subjects had extracted the rules from the example before training. There are at least two versions of this criticism. One is that subjects would extract these rules at study to help them learn the examples. The other is that when they heard the test instructions they mentally reviewed examples, extracting the rules before the test began. We thought it unlikely that subjects were doing either, but we decided to interview subjects after the experiment to find out what they were in fact doing.

We also wanted to avoid one other peculiarity of reversed rules in Experiment 1. For instance, subjects studied both a +1, +1 and a −1, −1 rule. When +1, +1 was reversed, it became effectively −1, −1, that is, both of the digits were decremented. Thus, subjects had been practicing the transformation they would need to produce in the reverse direction. Some of the improvement for subjects in the reverse direction may reflect the fact that they had practiced the transformation with another rule. To deal with this, we needed rules in one direction that did not correspond to any rule in the reverse direction.

As another design concern, we were bothered by the fact that rules involved two transformations of the same type. This meant that in practicing the transformation on one digit for one rule, subjects were practicing the same transformation on the other digit for another rule.

We decided to use a different set of transformations that would (a) avoid the identity transformation, (b) give us more rules to avoid practicing reverse transformations, and (c) have different operators for the first and second element. Subjects were asked to memorize examples such as "Hockey was played on Saturday at 3 and then on Monday at 1."

Although subjects did not know it while they were memorizing the examples, the second day and time were related by a transformation to the first day and time. The 2 days were either 1 or 2 days apart and the hours were 1 or 2 hr apart. Thus, the above example reflects the rule that the second game of hockey is 2 days later and 2 hr earlier. Subjects could be asked to apply their rule left to right to a novel item (If the first game of hockey was Wednesday at 8 when was the second game?) or right to left (If the second game of hockey was Friday at 6 when was the first game?). We can denote the rules for the day and or the hour according to the four possible transformations: −2, −1, +1, +2. Crossing these creates 16 possible rules. The example above would be denoted +2, −2, in which +2 refers to the fact that the day is incremented by two and −2 refers to the fact that the hour was decremented by two.

A final question that we wanted to answer in this study was what impact procedural practice would have on memory for the examples. We speculated that early in training subjects would be making reference to the example and so practicing the procedure would facilitate declarative memory for the example. However, with time the subjects would not refer to the example in applying the rule and so would no longer rehearse the example. To assess relative accessibility of the example, the experiment included occasional tests for memory of the original examples.

Method

Subjects. Twenty-two undergraduates from Carnegie Mellon University were recruited to participate in this experiment, which lasted 4 days, but one subject was excluded from the analysis because that error rate was too high and so mean times could not be calculated for certain conditions. The first session lasted 2 hr whereas the remaining three sessions lasted between 45 min and 1 hr. Subjects were either paid $16 for participating or given $8 and one credit for the human subject requirement in an introductory psychology course. In addition, all subjects received a bonus for performance, which varied from \$6 to \$16.

Procedure. The instructions used for this experiment were analogous to those used in Experiment 1 with appropriate modifications to reflect the differences in material. The experiment involved an initial learning phase, a training phase, and a transfer phase:

1. The first day began with the initial exposure phase followed by the three-pass dropout phase as in Experiment 1. During the initial exposure phase, subjects were told to study each of the eight examples and copy them from the top row to the bottom row. This gave them the opportunity to memorize the exemplars in addition to familiarizing themselves with the interface before beginning the three-pass dropout phase. In the dropout phase, they were shown just the sport name and had to reproduce the two days and two times.

2. After the dropout phase, subjects went into the training phase. This phase involved a memory test, a rule test, and then another memory test. The next two days repeated the training phase sequence of memory test, rule test, and memory test. In the rule test of each training phase, subjects were tested on two of the rules in the left-to-right direction (i.e., generate second day and time given first) and the other two rules in the right-to-left direction. The other four rules were not tested at all. The rule test involved 40 passes in which each of the four rules was tested once for a total of 160 trials. In the rule test, we were interested in how closely the learning function corresponds to a power function. The memory test required the subject to recall the two days and two times, given the sport name. On all days, the memory tests involved two passes through the eight examples for a total of 16 trials. Here we were interested in differential memory for trained and untrained examples.

3. The last day involved the transfer phase, in which the rule test was different. There were 10 passes in which all eight rules were tested in both directions for a total of 160 trials. The same memory tests were administered in the transfer phase before and after the rule test, just as in the training phase. Here we were interested in the degree of transfer in the practiced and unpracticed direction for the trained rules relative to the untrained rules.

Both the dropout phase and the memory test used the same cued-recall procedure. The sport name was presented and the subject had to recall the original example by using the mouse to click on the two days and two times. To make the rule test more comparable to the memory test, subjects had to click on both of the days and both of the times in the rule test. This meant that in the rule test, the subject simply copied the day and time presented while having to compute the other day and time according to the rule. Subjects received feedback on both the rule and memory tests.

As in the previous experiment, subjects were given points for their speed and accuracy. For each correct response 2 points were awarded, whereas for each incorrect response the score was decremented by 10 points. An additional point was received for every 0.5 s the response time was under 11 s. Thus, if a response was correct and executed in 7 s, the score was incremented by 10 points. The points awarded and total number of trials were displayed after every trial. These points were converted into a dollar amount that was given as the bonus pay for each subject upon completion of the experiment.

Materials. The fundamental structure of the experiment was determined by the rules. Each subject saw different randomly generated examples that embodied these rules. Two sets of eight rules were created. This would allow us to test whether the results obtained depended on the particular rules chosen. All possible relations (−2, −1, +1, +2) between time and day occurred twice in a set. The specific rules are shown in Table 1. Direction in Table 1 refers to whether the subject predicted the second date from the first (left) or the first from the second (right). Subjects were randomly assigned to either Rule Set A or Rule Set B. Within a rule set, half of the subjects practiced applying the first half of the rules (Set A1 or B1) during the first 3 days and never practiced applying the second half of the rule set. The remaining subjects practiced the second half (Set A2 or B2) of the rules on the first 3 days and never practiced the first half. Within each half set subjects practiced each of the four possible transformations on days and numbers.[7]

Specific examples for training and test were created by randomly assigning days and hours to the rules. Within a given training session, all 40 instances of a rule were unique and, in addition, were different from the initial exemplar for the rule. The rules were randomly assigned to eight sports that served to index the rules. Figure 6 illustrates the interface that was used in all phases of the experiment. Subjects selected days and times by clicking on the appropriate items.

Table 1
Rules Used in Experiment 2

Set	Apply rule	Direction
A		
Part 1	+1, +2	L
	+1, −1	R
	−2, +2	R
	−2, −1	L
Part 2	−1, −2	L
	−1, +1	R
	+2, −2	R
	+2, +1	L
B		
Part 1	+1, +1	L
	+1, −2	R
	−2, +1	R
	−2, −2	L
Part 2	−1, −1	L
	−1, +2	R
	+2, −1	R
	+2, +2	L

Note. L = left; R = right.

Results

Subject reports. At the end of the experiment, subjects were questioned within the next 2 days by E-mail as to how they had remembered the examples and how they had performed in the rule test. Eighteen of the 22 subjects responded with free-form answers. Ten subjects spontaneously reported using mnemonics to learn the examples. Eleven subjects spontaneously reported that there came a point in the training phase where they switched to the rule and bypassed recalling the example. One subject reported never abandoning use of the examples in the rule test. Only one subject reported extracting the rule before the rule test. Thus, it seems that most subjects did indeed recall the examples during the beginning of the training phase and switched directly to using rules as the training phase progressed.

Training data. There were no effects of which set of material that the subjects were trained with and so we averaged over this factor. Figures 7 and 8 show the improve-

[7] The first two rules in set A1 are +1, +2, and +1, −1, and so both involve a relation (+1) where the second day is one advanced from the first day. However, the second +1 relation is applied in the right-to-left direction and so becomes a −1 operation.

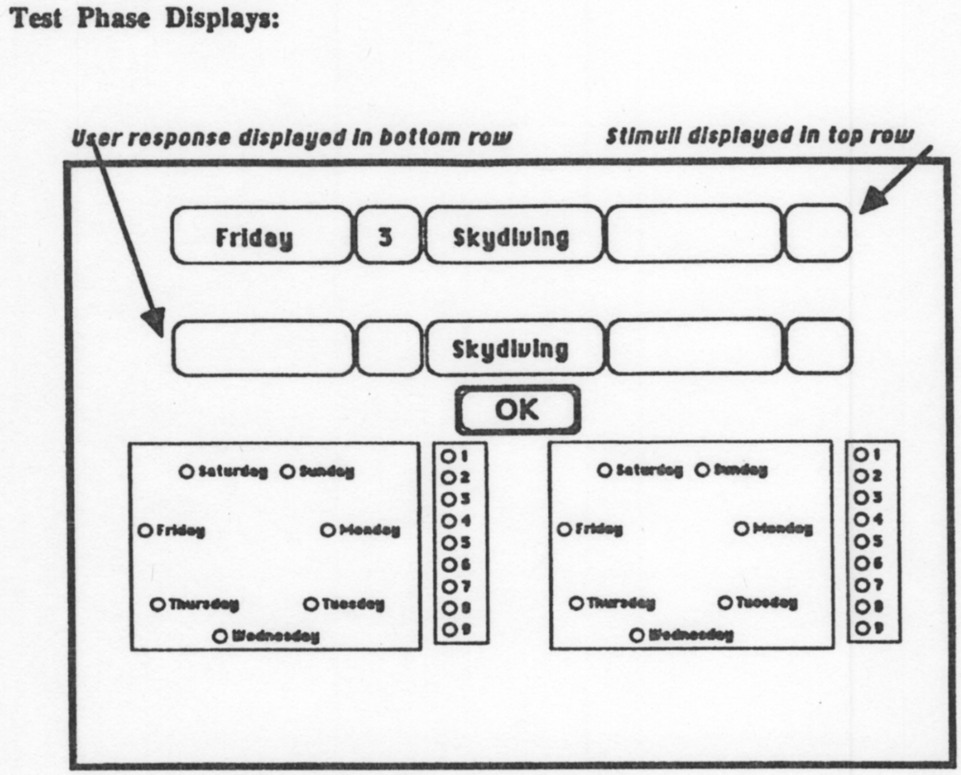

Figure 6. An example of the interface used in Experiments 2 and 3.

ment over the course of the experiment during the training phase for latency and accuracy. These figures also present the best fitting power functions. Note these data are plotted over 3 days. The first three trials on each day are plotted individually. The data are aggregated for Trials 4–6, 7–10, 11–15, 16–25, and 26–40 on each day as in Figures 1 and 2. (Trial m on day n is plotted as $40(n - 1) + m$ in the figures.) The power functions are fit to all the data points except the first 3 trials on Days 2 and 3. Excluding these points, the fit is very good for the latencies ($R^2 = .99$) and modest for accuracy ($R^2 = .79$). There are two noteworthy aspects about these learning data. First, as in Experiment 1, there is no evidence for a first-trial discontinuity. Second, there is no permanent forgetting from day to day. Rather, the effect of a delay is to introduce some initial slowness and increased error rate from which the subjects recover after three trials and then resume where they left off on the learning curves.

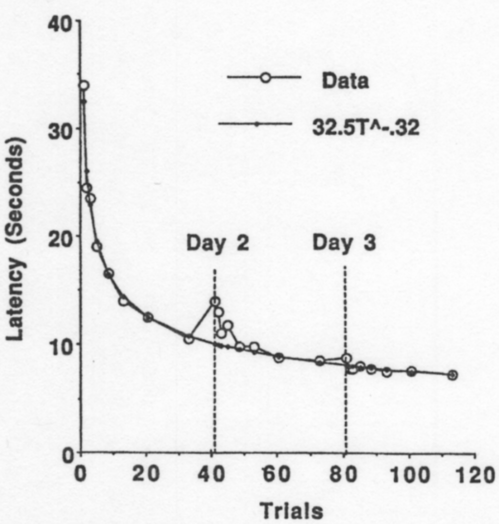

Figure 7. Improvement in latency over the course of Experiment 2.

It is both remarkable and unexpected that there appears to be no permanent forgetting. The motor-learning literature frequently finds that there is very little permanent forgetting and the initial loss is transient (Schmidt, 1988). The rapid improvement on retesting is referred to as a *warm-up effect* and the transient poor performance at the beginning of retesting is referred to as a *warm-up decrement* (Adams, 1961). In the verbal-learning literature, it has been noted that some of forgetting may be due to a warm-up decrement (Adams, 1961; Postman, 1969; for reviews). There have been demonstrations that the warm-up decrement might be nonspecific. For instance, Irion (1949) found that simply naming colors that appeared in a memory drum improved retention of paired adjectives. The argument has been that doing any experiment-like task that gets the subject back in the context will eliminate the warm-up decrement even if it does not involve the specific material being retained. There has been a mixed history as to whether such general practice effects are obtained with results like those of Irion failing to be replicated (see Adams, 1961 and Postman, 1969 for reviews). However, outside of the motor skills literature it seems that no one has actually given extensive retraining, as in Figure 7 and 8, to see whether there is any permanent forgetting. This is probably related to the tendency to use accuracy measures in many studies of retention that make it difficult to trace out extended practice functions. The latency data in Figure 7 make the lack of permanent forgetting particularly clear.

Memory test data. We performed ANOVAs on the memory test data that preceded and followed the procedural practice. Variables were day (four values), and time of day (two values: pre- vs. postrule test), and whether the rule was practiced (two values). With respect to latency there were significant effects of day, $F(3, 60) = 34.29, p < .0001, MS_e = 14.84$; of time of day, $F(1, 20) = 46.27, p < .0001, MS_e = 10.98$; a significant interaction between these effects, $F(3, 60) = 3.12, p < .05, MS_e = 9.46$. The character of these effects were that subjects got faster as they proceeded through the experiment, were faster in posttest, and the posttest advantage diminished over days. Although subjects were faster at recalling examples associated with practiced rules (8.54 s vs. 9.08 s), this was not significant overall, $F(1, 20) = 2.06, MS_e = 11.88$. The only other significant effect was a three-way interaction between days, time of day, and whether the material received procedural practice or not, $F(3, 60) = 4.84, p < .01, MS_e = 3.64$. The uninteresting part of this interaction was that there was no difference between practiced and unpracticed material before the first practice on Day 1. The more interesting aspect is how the practice effect emerged thereafter. This is illustrated in Figure 9, which aggregates the posttest after day n with the pretest before day $n + 1$ as "after practice on day n" for Days 1–3. The effect is much larger after the first day's rule practice than after the second or third day's practice. A contrast that specifically asks if the difference after Day 1 is larger than the difference after Day 3 is quite significant, $t(60) = 3.22, p < .001$.[8]

The analysis of error rates found significant effects of day, $F(3, 60) = 10.57, p < .0001, MS_e = .019$; time of day, $F(1,20) = 12.38, p < .01, MS_e = .004$; and whether the rules were practiced, $F(1, 20) = 6.73, p < .05, MS_e = .023$. The only significant interaction was the three-way interaction among the three variables, $F(3, 60) = 8.23, p < .0001, MS_e = .006$. Again it has the same interpretation. The interesting aspect of this interaction is illustrated in Figure 10 where it can be seen that the effect of practice is larger after Day 1 than after Days 2 and 3. Again, a specific contrast for this effect is significant, $t(60) = 4.19, p < .001$. Figures 9 and 10 indicate that the procedural practice has the largest benefit on memory for the examples early and later practice does not benefit them as much. This is evidence that subjects are using examples during Day 1 procedural training but not on later days.

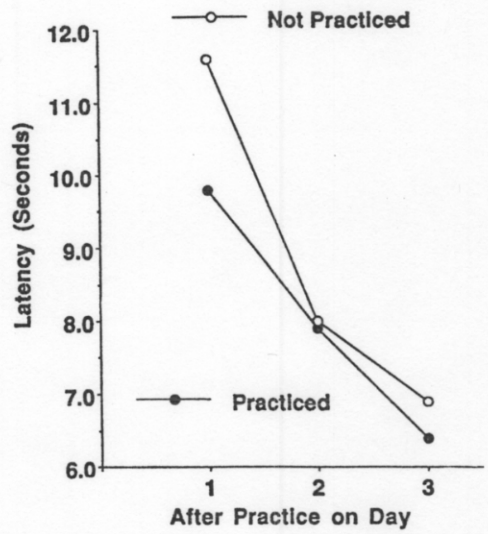

Figure 9. Latency in memory test for those items which received rule practice and those which did not as a function of number of days of practice.

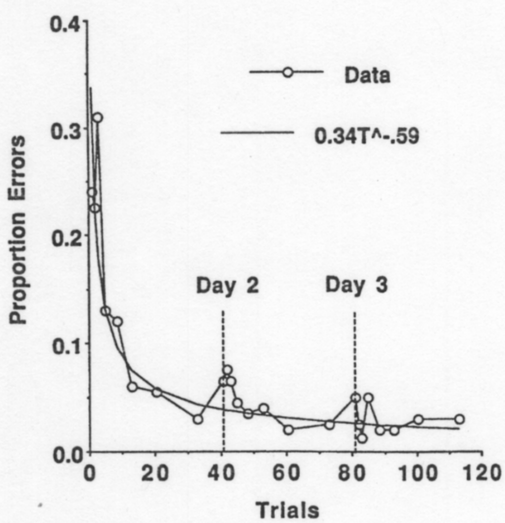

Figure 8. Improvement in proportion errors over the course of Experiment 2.

Transfer data. The final part of the data concerns the fourth day when all the rules were trained in both directions. We performed ANOVAs on both the latency and accuracy data. The variables were direction of the rule (same as training or reversed), whether the rule was practiced during training (two values), and amount of practice in the transfer phase (five values: the data were broken down into means for passes 1, 2, 3, 4–6, and 7–10). As for latency, there were main effects of direction, $F(1, 20) = 6.58, p < .05, MS_e = 7.91$; whether the rule had procedural practice, $F(1, 20) = 33.13, p < .0001, MS_e = 29.44$; and amount of practice in the transfer phase, $F(4, 80) = 14.14, p < .0001, MS_e = 18.97$. As in Experiment 1, there was a significant practice by direction interaction, $F(4, 80) = 2.77, p < .05, MS_e = 6.70$, such that the direction effect diminished with practice. The only other significant interaction was between whether the rule was practiced during training and direction of practice, $F(1, 20) = 9.95, p < .01, P < .01, MS_e = 19.40$, and it is displayed in Table 2. There is only an effect of direction for the rules that have been practiced, which is as expected. The direction variable is a dummy variable in

[8] Figure 9 shows improvement over days because of the learning benefits from earlier days.

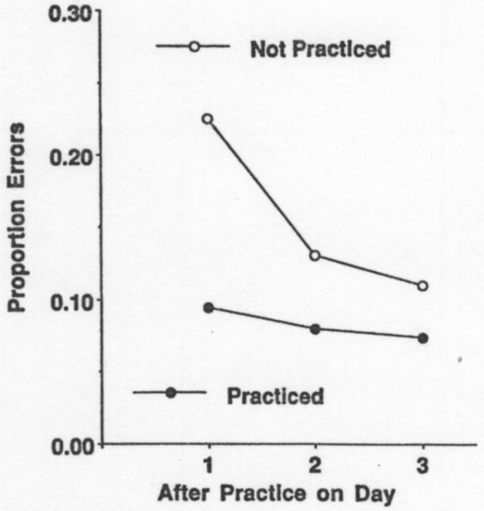

Figure 10. Proportion errors in memory test for those items that received rule practice and those that did not as a function of number of days of practice.

the case of rules that have not received procedural practice. This latency interaction is evidence for the procedural asymmetry predicted by the ACT-R theory.

The results for errors are comparable with significant effects of whether the rules have had procedural practice, $F(1, 20) = 8.14$, $p < .01$, $MS_e = .028$; amount of practice during the transfer phase, $F(4, 80) = 4.77$, $p < .01$, $MS_e = .022$; and a training-by-direction interaction, $F(1, 20) = 7.45$, $p < .05$, $MS_e = .017$. Table 2 also displays the error data for the training-by-direction interaction. There was also a training-by-practice interaction, $F(4, 80) = 3.26, p < .05, MS_e = .018$, such that the effect of prior training diminished with practice.

Discussion

The results confirm and extend the conclusions from the first experiment. Again there is no first-trial discontinuity and there is an asymmetry in access to the knowledge that has been practiced in a procedural test. The learning curves provide detailed data about the acquisition and retention of rules. The latency data is particularly important. We see that the learning curve is a good fit to a power function from the first trial. The retention results are truly surprising. What might appear as forgetting on the first few trials of the next day is better characterized as a rustiness that can be eliminated by a few warm-up trials.

Table 2
Effects of Direction of Practice and Whether the Rule has Been Trained: Latency (in Seconds) and Error Rates (Proportion)

	Rule			
	Practiced		Not Practiced	
Direction	Latency	Errors	Latency	Errors
Same	8.9	0.07	13.3	0.15
Reverse	10.9	0.12	12.6	0.13

There was a benefit of procedural practice on the memory test on the first day, and this diminished on subsequent days (Figures 9 and 10). This indicates that subjects were using the declarative example on the first day, presumably in an analogy process. This use of the examples in analogy strengthened the examples and gave them an advantage in the memory test after the first day. The transition to rule-based performance on later days allowed the example to be bypassed and, hence, led to a diminished memory difference between examples corresponding to practiced and nonpracticed rules.

The various results of these two experiments converge in suggesting that subjects start out solving by analogy to examples but gradually (not in a single trial) switch to responding by procedural rules. This gradual change would either take the form of specific items changing their probability of rule-based responding or different items switching in an all-or-none manner on different trials. Among the results consistent with this proposal are (a) this is what subjects report, (b) there is no first-trial discontinuity, (c) only on the first day is there a large effect of procedural practice on memory for examples, and (d) by the end of the experiments there is a large directional asymmetry. A number of other researchers have proposed that new procedures only appear gradually, increasing their probability of applying with repeated practice (Siegler & Jenkins, 1989, Van Lehn, 1991).

As a final point, note that in Table 2 subjects are faster in using the practiced rules in the reverse direction than they are in using the rules that are not practiced. One might want to interpret this as evidence for positive transfer between the procedures in the two directions. However, to the extent subjects used the examples in the transfer-rule test, the declarative representations might have received greater practice for the practiced rules and this may be the cause of the advantage. Figures 9 and 10 indicate that the examples receiving practice in the rule test were more available. The next experiment will try to provide data to decide between these two interpretations.

Experiment 3

Experiment 3 in part was motivated to replicate and extend the results of the previous experiments. We wanted to look at the learning curve for a procedural skill one more time, both to confirm the lack of a first-trial discontinuity and to see whether we could replicate the surprising warm-up effects at the beginning of the subsequent days.

The most important purpose of this experiment was to get more definitive data on the asymmetry that has been found in the procedural skills. We have used this asymmetry as an argument that procedural knowledge is different from declarative knowledge, but we have not performed a comparable test for declarative asymmetry (i.e., if subjects were trained to recall Part b of the example given Part a, how would they do at retrieving Part a, given Part b?). Also, we wanted to get data that would allow us to determine whether the advantage of the practiced rules used in the reverse direction over the nonpracticed rules (in Table 2) reflected, as hypothesized, extra practice of the example during the rule test.

To achieve these purposes, we introduced a task that would be the declarative analog of the procedural rule test used in the previous two experiments. We wanted a test in which a subject would be prompted with the sport name and two of the other terms and have to retrieve the other two terms just as in the rule test. To achieve this, we used an example test in which subjects were either prompted with the two days from the example and had to retrieve the two times or vice versa. This task could be solved by retrieving the example and so could be done declaratively.[9]

The paired-associate literature can be consulted for evidence on the issue of whether there is asymmetry in declarative learning (for reviews see Ekstrand, 1966; Horowitz, Norman, & Day, 1966; Paivio, 1971, pp. 276–285). More often than not, it is found that forward associations (stimulus to response) are more available than backward associations (response to stimulus), but much of this is due to variables such as greater learning of the responses (which are often nonsense syllables). Circumstances can easily be set up where the stimulus is more available than the response (e.g., when stimulus is a digit and response a nonsense syllable or stimulus a concrete noun and the response an abstract noun). In these cases backward recall typically exceeds forward recall. After reviewing those experiments that try to equate for stimulus and response availability, Ekstrand (1966) concludes, "It does appear that the difference between forward and backward associative learning has been drastically overestimated, and that if symmetry is not the rule, asymmetry will be very small" (p. 60). In our situation, we are using common digits and days and the response medium is clicking the mouse. Thus it seems unlikely that there will be much difference in response availability. Therefore, we should see near symmetry in the declarative tests.

Method

Subjects. Thirty-seven undergraduates from Carnegie Mellon University were recruited to participate in this experiment that lasted 4 days. Three subjects were excluded from the analysis because their error rates were so high that they left latencies undefined for certain conditions. Subjects were either paid $16 for participating or given $8 and one credit for the human subject requirement. In addition, they received a bonus for performance, which varied from $8 to $16.

Materials. The materials were the same as in the previous experiment, with the exception that only Sets A1 and A2 were included because the previous experiment had not revealed an effect of materials.

Procedure. The basic procedure involved first having subjects learn a set of examples as in Experiment 2. Then subjects practiced either applying the rule implicit in the example, or retrieving the example, or doing both. Here we were interested in the learning curves for different types of training (rule application or example retrieval). Interspersed throughout the experiment were transfer tests in which we assessed how well the subject could retrieve the example or apply the rule in the direction opposite to training. Here we were interested in degree of asymmetry in transfer in both rule application and example retrieval. Appendixes A and B provide an example of the phases of the experiment, conditions, and the experimental material. Specifications of the experiment are contained in the following description.

The first day started with an initial exposure and then a three-pass dropout learning phase as in the first two experiments. Each day involved a rule-training test, an example-training test, a final rule-transfer test, and a final example-transfer test. In either the rule-training or rule-transfer tests, the subject would either be presented with the left day and time and be asked to generate the appropriate right day and time or vice versa. In either the example-training or example-transfer test, the subject would be presented either with the two days from the original example and have to recall the original two times or vice versa. In all tests the sports names were presented. In both tests subjects would have to select all four items by clicking the mouse. As in the previous experiment, this meant they copied two items from the probe and produced the other two items.

There were in total four examples that were tested in the example-training test and four rules that were tested in the rule-training test. Two of the rules were assigned to be tested only in the example-training test, two to be tested only in the rule-training test, two in both training tests, and two in neither training tests. The rules in each pair were tested in different directions in the training tests. On each day, each of the training tests involved 32 blocks in which each rule or example was tested once. Thus, on each day there were 128 trials in the example training and 128 trials in the rule training. Both the example- and rule-transfer tests involved two passes in which all of the eight examples or eight rules were tested in both directions. Thus, each of these involved 32 trials. In the rule tests, all the instances for a rule on a given day were unique, and no two tests of the same rule appeared successively. As in the previous experiments, order was random within each block.

Subjects were given points for their performance as per the formula in the previous experiment. On successful completion of the experiment, we used the point total to determine the bonus dollars the subject received. Sixteen of the subjects were tested in the order of example training, rule training, example transfer, and rule transfer, and for 18 subjects the example and rule sections were reversed.

Results

Training. Figures 11, 12, 13, and 14 show the training data for the rule-training and example-training tests for the dependent measures of latency and accuracy. The R^2s for goodness of fit exclude the first three trials on subsequent days. There are two significant aspects to these learning curves. First, we see the latency data is fit quite well by power functions. Again, there is no evidence for a first-trial discontinuity in the rule-training data. There is a larger than predicted drop in the example training on the first trial, but we had no reason to predict this effect for example training. Second, the data continue to show warm-up effects on subsequent days. The observation of these effects for the example training is significant because, as observed earlier, such warm-up effects have been mainly observed in more procedural tasks.

We performed four ANOVAs on these data in the figures in which the variables were order (rule-training first or example-training first), rule type (either uniquely used in that test or in both tests), days (4 values), and trials (13 values: 1, 2, 3, 4–5, 6–8, 9–11, 12–14, 15–17, 18–20, 21–23, 24–26, 27–29, 30–32). Although the effects of days, of trials, and their interaction were always significant, there were no significant effects or

[9] Of course, subjects could form a production rule to retrieve the specific answer but we assumed they would not be motivated to do so because it was unnecessary. In contrast, on the rule test, the subject's task of generating the appropriate day and time for that sport was made much easier by using a rule rather than having to perform an analogy to the original example each time.

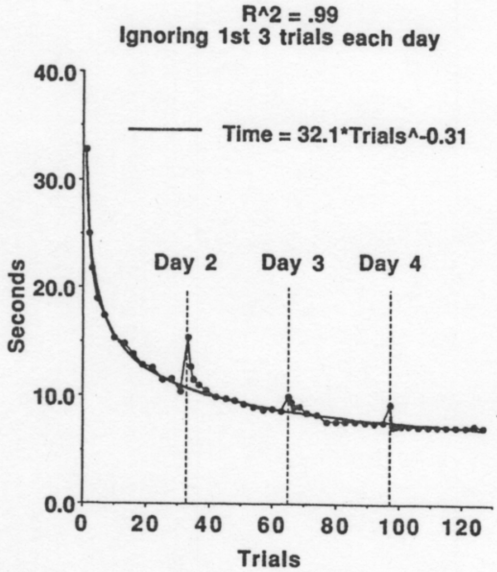

Figure 11. Latency in rule training as a function of amount of practice.

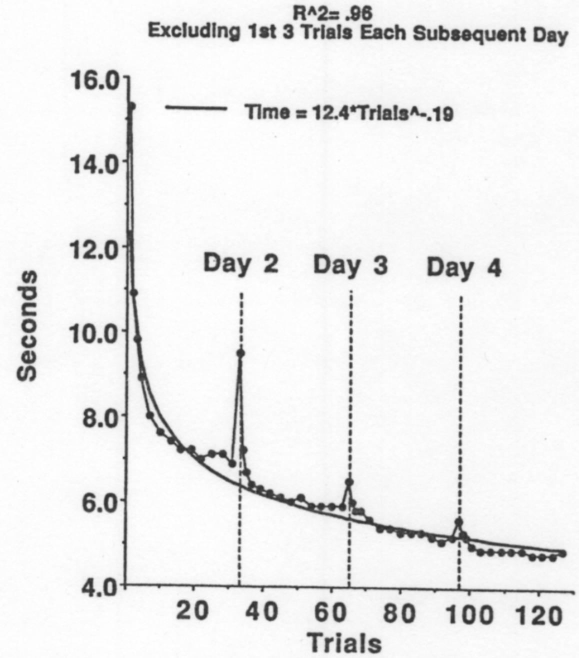

Figure 13. Latency in example training as a function of amount of practice.

interactions of order or rule type (at least using the more stringent Geisser-Greenhouse probability) except for a Trial × Order × Rule Type interaction for the errors in rule training, $F(12, 396) = 2.62$, $p < .01$, Geisser-Greenhouse $p < .01$, $MS_e = .018$. The effect here was that subjects were somewhat less error prone in early trials in the rule test in the both-test condition if it had been preceded by the example training.

There was no significant difference in rule-training test performance between those items that received only rule training and those that also received example training, 10.57 s for unique test and 10.74 s for both tests, $F(1, 33) = .64$, $MS_e = 39.34$; .067 errors for unique test and .064 errors for both tests,

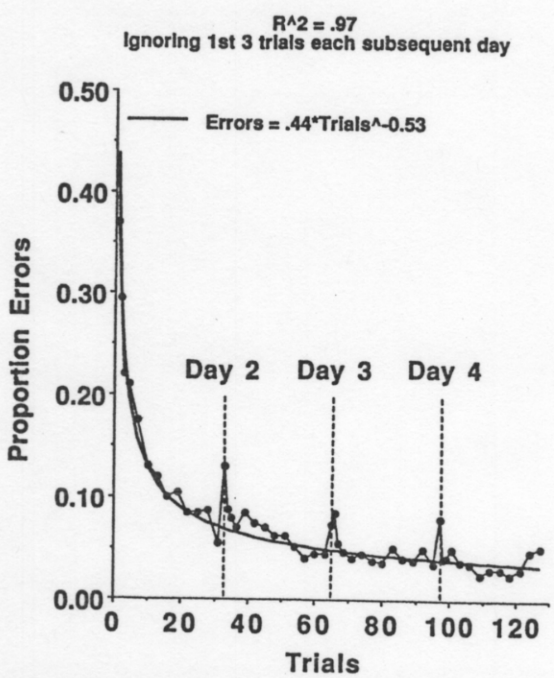

Figure 12. Proportion errors in rule training as a function of amount of practice.

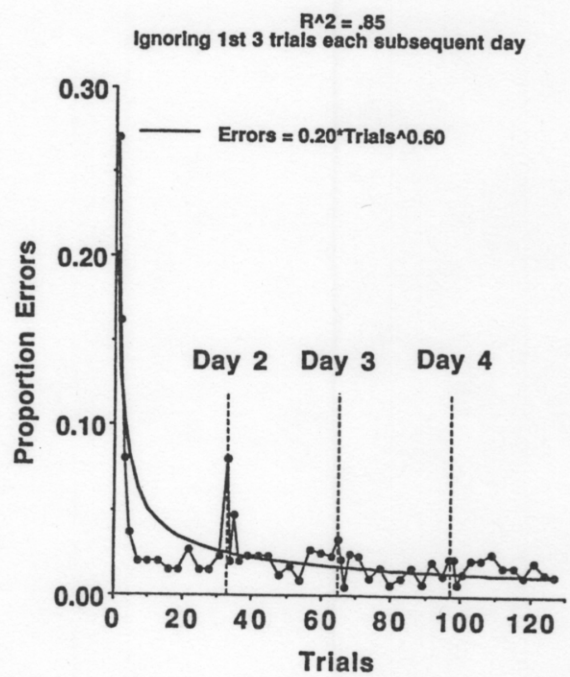

Figure 14. Proportion errors in example training as a function of amount of practice.

$F(1, 33) = .13$, $MS_e = .064$. Similarly, there was no difference in example-training test performance between those items that received only example training and those that also received rule training, 6.40 s for both and 6.37 s for only, $F(1, 33) = .29$, $MS_e = 8.69$; .028 errors for both and .024 errors for only, $F(1, 33) = 1.36$, $MS_e = .009$. Thus, practice of the knowledge in one type of training test did not transfer to the other type of training test. This lack of transfer supports the distinction between declarative knowledge (example training) and procedural knowledge (rule training).

Rule transfer. We performed ANOVAs on the rule-transfer data in which the variables were how the rule had been trained (both rule test and example test, just rule test, just example test, neither), direction of testing (same vs. reverse), and days (four values). There was no significant difference between rule-only and both, $t(33) = .66$, $MS_e = 32.51$ for latency; $t(33) = .67$, $MS_e = .132$, for errors). Therefore, performance on rule-only and both training conditions have been aggregated. However, performance has been divided for these two conditions into tests in the same direction as practiced and tests in the reverse direction as defined in Appendix B. Figures 15 and 16 plot performance on trained rules in the same direction, trained rules in the reverse direction, rules that received only example practice, and rules that received no practice.

We performed an ANOVA on the latency transfer data in which the variables were training type (same, reverse, example training only, no training) and day. There were significant effects of training type, $F(3, 99) = 30.54$, $p < .001$, $MS_e = 15.30$, day, $F(3, 99) = 58.98$, $p < .001$, $MS_e = 24.20$, and an interaction between the two, $F(9, 297) = 2.34$, $p < .05$, $MS_e = 7.73$. With respect to the effect of training type, the no-training condition was significantly slower than example-only training, $t(99) = 2.62$, $p < .01$; example-only training was slower than reverse training, $t(99) = 2.10$, $p < .05$, and reverse training was much slower than same training, $t(99) = 4.52$, $p < .001$. The basic character of the Day × Condition interaction is that the difference disappears over days among the no training, example-only training, and reverse-training condition (from a difference of more than 3 s on Day 1 to less than 1 s on Day 4), whereas the same condition maintains its 2-s advantage over the reverse condition throughout the experiment.

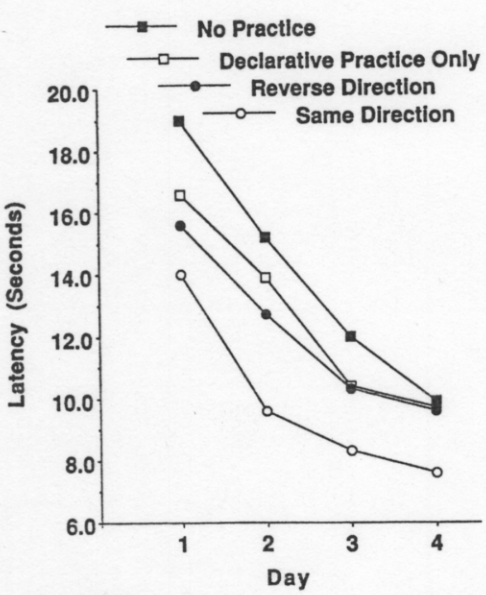

Figure 15. Latency in rule transfer test as a function of day.

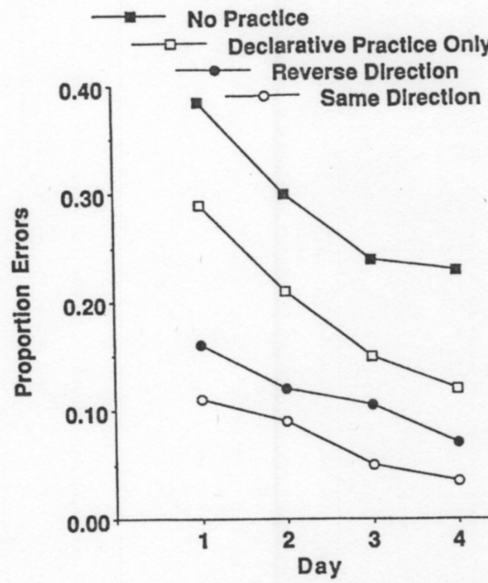

Figure 16. Proportion errors in rule transfer test as a function of day.

The data are consistent with the prediction of asymmetry in procedural knowledge. Everywhere the reverse condition is closer to the example-only condition than it is to the same condition. The small initial advantage of the reverse direction over example training may reflect some difficulty in correctly identifying the underlying relationship (e.g., the second day is one more than the first and the second hour is two less). Once this is cleared, there is no advantage. Thus, the data in the figure provide strong evidence for the ACT–R claim of asymmetry in procedural knowledge.

With respect to error rate, the reverse direction is closer to the same direction than example training. However, the error data shows the same interaction as the latency data in that the difference between example training and reverse direction diminishes while the difference between reverse direction and same direction maintains itself. This is consistent with the interpretation of the latency interaction as reflecting some initial benefit of practice in either direction on identifying the rule.

Although ACT–R is unambiguous in its prediction of asymmetry in latency, its predictions for errors are more complex. One has to consider all the ways in which errors can occur in transfer. There are at least four bases for errors to occur:

1. The subject might make an error for the same reasons already discussed with respect to the training data (see discussion of retrieval errors vs. other errors in the *Results* section of Experiment 1).

2. The subject may have forgotten the example. This seems particularly possible in the no-training condition.

3. The subject may misunderstand the rule embedded in the example. This seems particularly possible in the example-only and no-training conditions.

4. The subject may get confused about the direction in which to apply the rule. This seems particularly possible in the reverse condition.

Table 3 provides an analysis of the errors in transfer relevant to this classification. We scored separately whether the subject got the hour and the day right and then averaged these two scores. Thus, each trial offers two opportunities for an error, and Table 3 reports average error rates over these two opportunities. We have classified errors into directional errors in which the subject has applied the right rule but in the wrong direction (e.g., produced a -2 transformation in day or hour when a $+2$ transformation was required) and nondirectional errors that includes all other errors in production of the day or hour.[10] Even though directional errors represent only one incorrect choice and nondirectional errors sum all other choices, there are slightly more directional errors overall. Thus, the fourth type of error is a significant factor. With respect to nondirectional errors, there is no difference in error rate between the reverse direction and the same direction. With respect to directional errors, there is a higher error rate in the reverse case, $t(108) = 1.82, p < .05$, one-tailed. This reflects a greater propensity for the fourth type of error in the reverse case. There is a large difference between the example-training-only condition and the conditions where there is procedural practice. Presumably, this difference reflects the contribution of the third type of error in the declarative-only condition in which the subject has not had training in identifying the rule. There is also a large difference between the example-training-only condition and the no-practice condition reflecting in part the contribution of the second type of error.

There is no difference in nondirectional error rates for the same and reverse conditions. On the other hand, there are increased directional intrusions in the reverse direction condition. This is consistent with a view that holds that subjects first extract a declarative representation of the rule and then compile it into a production. Thus, acquiring a production in either direction will guarantee the rule is extracted and so give the same low rate of nondirectional errors. However, practicing that production rule will result in specific strengthening of the rule that will show up in shorter latencies and directional errors because the highly practiced rule will sometimes intrude when the other is required.

Table 3
Classification of Rule-Transfer Errors

Type of practice	Type of error: Directional errors	Nondirectional errors
No training	.103	.094
Example-training only	.057	.063
Reverse direction	.042	.020
Same direction	.020	.020
Average	.056	.049

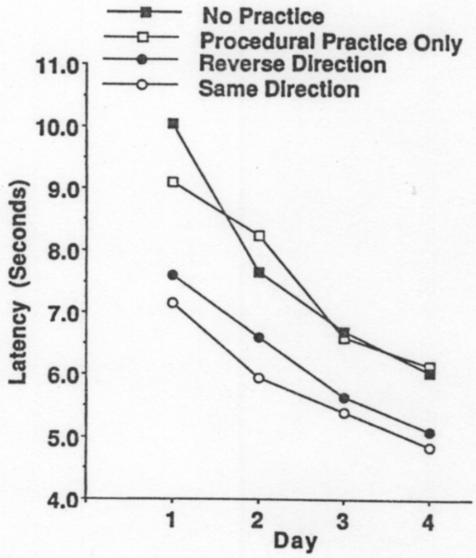

Figure 17. Latency in example transfer test as a function of day.

Example transfer. Figures 17 and 18 show comparable data for the example-transfer test using the analogous aggregation scheme (i.e., the role of example-training and rule-training switch in the analogy). For definitions of the conditions displayed in Figures 15–18, see Appendix B. Comparable ANOVAs were performed on these data as on the data in Figures 15 and 16. In this case, there is no difference between items that received just example training or example training and rule training, $t(33) = .86, MS_e = 15.62$, for latency; $t(33) = .71, MS_e = .073$ for errors. Therefore, these two conditions were collapsed and then divided according to whether the testing was in the same direction as training.

The corresponding ANOVA was performed on the declarative latency data as was performed on the procedural latency data. There were significant effects of type of training, $F(3, 99) = 19.62, p < .001, MS_e = 7.62$; of days, $F(3, 99) = 70.76, p < .001, MS_e = 3.57$; and a significant interaction between training and days, $F(9, 297) = 6.47, p < .001, MS_e = .98$. The no-training condition was not significantly slower than rule-training-only condition, $t(99) = 1.26, p > .05$; rule-only training was significantly slower than reverse training, $t(99) = 3.99, p < .001$; and reverse training was not significantly slower than same-direction training, $t(99) = 1.30, p > .05$. The nature of the Day × Condition interaction appears to be that only on the first day is there an advantage of rule training over no training. A contrast, specifically testing this, is significant, $t(297) = 2.81, p < .01$. This replicates the interaction from Experiment 2 and is consistent with the view that examples are used in the rule task only on the first day. After the first day, problem solving by analogy to the examples drops out.

[10] Note these are not the only logical ways subjects could make errors. They could also make errors in reproducing the day and time with which they were prompted, but such errors were rare.

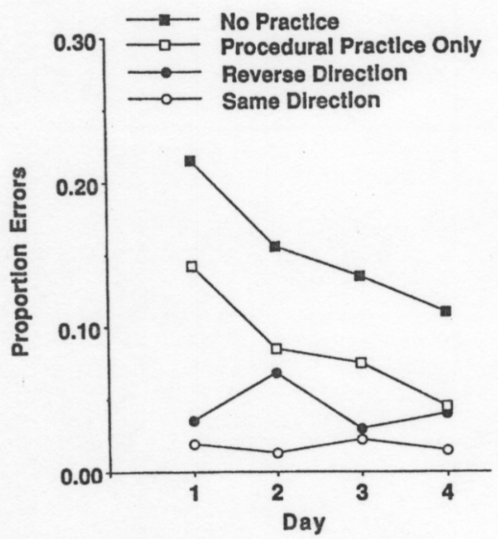

Figure 18. Proportion errors in example transfer test as a function of day.

The results are consistent with the paired-associate literature in showing relatively little asymmetry when effects of response availability are minimized. The small nonsignificant differences that remain may reflect small effects of response availability. That is, subjects trained on a specific example that involves recalling Saturday would get some extra practice in retrieving that day. It is also possible that subjects developed rules that generated the examples. That is, there is no reason why subjects could not form production rules for performing the example retrieval task. It is just that there is no reason for them to have done so.

Comparison of procedural and declarative asymmetry. Qualitatively, as predicted, in the latency data (Figures 15 vs. 17) there appears to be large asymmetry in the rule transfer and negligible asymmetry in the example transfer. However, it would be comforting to have some statistical test to support appearances. It is somewhat problematical to compare the size of the effects because they are occurring on a smaller base latency in the case of example transfer. Therefore, we developed the following measure of asymmetry:

$$\frac{\text{reverse direction} - \text{same direction}}{\text{other training} - \text{same direction}}$$

where these conditions are defined in Appendix B with other training being either example training for a rule test or rule training for an example test. This is a measure that should vary from 0, when there is complete transfer to the reverse direction, to 1, when there is no better performance in the reverse direction than is obtained by the other kind of training. This is a way of measuring asymmetry that normalizes for the fact that the rule test judgments are longer. These proportion asymmetries are .81 for procedural asymmetry (rule transfer test) and .26 for declarative asymmetry (example transfer test). We calculated for each subject whether the asymmetry was greater for the rule or the example test. For 28 of the 34 subjects, the asymmetry was greater in the procedural case. This is highly significant by a sign test ($p < .001$).

Summary of Experiment 3

This experiment replicates and extends the picture from the previous experiments. The significant results are as follows:

1. Rule training, particularly under a latency measure, corresponds to a power function from the very first trial. This disconfirms ACT–R's prediction of a first-trial discontinuity.
2. There appears to be no permanent retention losses for either rule training or example training. The initial loss is really just a warm-up decrement. This result is unexpected on all theoretical perspectives.
3. There is a strong asymmetry in rule transfer that does not appear in example transfer. This supports the declarative-procedural distinction in ACT–R.
4. There is no transfer between rule training and example training (i.e., no advantage of training on both over just on the target activity). This is also consistent with the procedural-declarative distinction.
5. There is only a benefit of the first day of procedural practice on example retrieval and no benefit thereafter. This suggests that the example is only being used in analogy on the first day and after that the rule is directly used.

General Discussion

These results clearly call for changing one assumption in the ACT–R theory. This is that the subjects switch from an example-based processing to a rule-based processing in a single trial. There is no first-trial discontinuity in any of the data. Rather, there seems to be a gradual shift from example-based processing to rule-based processing. Perhaps each trial gives subjects another opportunity to encode the rule. Or perhaps rule-based processing and example-based processing compete as alternative means of answering the procedural questions. After a day of practice, subjects had largely switched to rule-based processing. Evidence for this is the strong procedural asymmetries that emerge and the diminishing transfer to a declarative test after the first day. Novick and Holyoak (1991) also concluded the shift from analogy to examples to rule-based processing is gradual. Similarly, Vokey and Brooks (1992) argued that performance in the Reber task is a mixture of example-based performance and more abstract rule-based performance.

With practice on the procedural-rule task, subjects extract a declarative representation of the rule and compile a production to embody it. Extracting a declarative representation of the rule provides some benefit in the reverse direction, especially with respect to error rate. However, extensive practice should strengthen only the production, producing directional asymmetry particularly with respect to latency.

The latency data in Experiment 3 showed another asymmetry between procedural and declarative learning when measured by latency. There was large positive transfer from example-only training (relative to no training) to the procedural rule-transfer test, whereas there was only little transfer (and only on the first day) from rule-only practice to the declarative example-transfer test. This indicates that subjects can use the examples to perform the procedural task but not the rules to perform the declarative task. Thus, increased speed at retrieving the examples speeds analogical solving of the procedural task but increased speed at applying a rule does not facilitate retrieving an example. One only gets procedural-to-declarative transfer on the first day when subjects are still using the examples in analogy.

It is interesting to consider Logan's (1988, 1990) theory of skill acquisition in light of these results. He has emphasized a process of learning by which subjects go from rules to specific examples. According to his theory, the power law of learning is due to increased retrieval of the specific examples. The results of our study do not fit this theory well. For instance, because we plot trial-by-trial initial learning we are seeing large, dramatic power-law learning when there are no examples being repeated. An example was never repeated from the first trial to the second trial in Experiments 2 and 3, but subjects decreased from an average of more than 33 s to less than 25 s for a decrement of more than 25%.[11] Clearly, the improvement being observed on Trial 2 is not due to retrieving the example from Trial 1. Indeed, none of the learning on Day 1 can be due to example repetition because no examples are repeated.

In contrast to Logan's (1988) proposal of rule-to-instance transition, the data indicate an instance-to-rule transition of a sort. That is, subjects are switching from using an example in analogy to using a rule. It is possible that further along subjects might memorize the application of the rule to specific examples and retrieve these as well. However, it is clear that the power-law learning cannot be exclusively due to retrieval of specific examples of applying the general rule.

[11] It is true that, although the whole example would not repeat, either the day or time might repeat. There was approximately a one-quarter probability of either the day or time (but not both) repeating from example to example. Even if repetition of this half-reduced problem solution time on it to zero, it would only produce an average speed up of one eighth in contrast to the more than one-quarter speedup observed.

References

Adams, J. A. (1961). The second facet of forgetting: A review of warm-up decrement. *Psychological Bulletin, 58,* 257–273.

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

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

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bellezza, F. S., & Boyle, C. F. (1993). The geometry tutor and skill acquisition. In J. R. Anderson (Ed.), *Rules of the mind* (pp. 165–181). Hillsdale, NJ: Erlbaum.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science, 13,* 467–506.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology, 36A,* 209–231.

Broadbent, D. E. (1989). Lasting representations and temporary processes. In H. L. Roediger, III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 211–228). Hillsdale, NJ: Erlbaum.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development* (pp. 141–174). Cambridge, MA: Cambridge University Press.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–182.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 17–30.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General, 113,* 541–555.

Ekstrand, B. R. (1966). Backward associations. *Psychological Bulletin, 65,* 50–64.

Horowitz, L. M., Norman, S. A., & Day, R. S. (1966). Availability and associative symmetry. *Psychological Review, 73,* 1–15.

Irion, A. L. (1949). Retention and warming-up effects in paired associate learning. *Journal of Experimental Psychology, 39,* 669–675.

Kessler, C. M. (1988). *Transfer of programming skills in novice LISP learners.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

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

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22,* 1–35.

Marescaux, P., Dejean, K., & Karnas, G. (1990). *Acquisition of specific or general knowledge at the control of a dynamic simulated system: An evaluation through a static situations questionnaire and a transfer control task* (KAUDYTE Rep. No. 2PR2GK). Bruxelles: Universite Libré de Bruxelles, Laboraitoire de Psychologie Industrielle et Commerciale.

McKendree, J. E., & Anderson, J. R. (1987). Effect of practice on knowledge and use of basic LISP. In J. M. Carroll (Ed.), *Interfacing thought: Cognitive aspects of human-computer interaction* (pp. 236–259). Cambridge, MA: MIT Press.

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

Nissen, M. J., Knopman, D. S., & Schacter, D. L. (1987). Neurochemical dissociation of memory systems. *Neurology, 37,* 789–794.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 398–416.

Paivio, A. (1971). *Imagery and verbal processes.* Hillsdale, NJ: Erlbaum.

Pennington, N., & Nicolich, R. (1991). Transfer of training between programming subtasks: Is knowledge really use specific? In J. Koenemann-Belliveau, T. G. Moher, & S. P. Robertson (Eds.), *Proceedings of the Fourth Workshop on Empirical Studies of Programmers* (pp. 156–176). Norwood, NJ: Albex.

Phelps, E. A. (1989). *Cognitive skill learning in amnesics.* Unpublished doctoral dissertation, Princeton University, Princeton, NJ.

Pirolli, P. (1985). *Problem solving by analogy and skill acquisition in the domain of programming.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Pirolli, P. (1991). Effects of examples and their explanations in a lesson on recursion: A production system analysis. *Cognition and Instruction, 8,* 207–259.

Pirolli, P., & Recker, M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction, 12,* 235–275.

Postman, L. (1969). Experimental analysis of learning to learn. In G. Bower (Ed.), *Psychology of Learning and Motivation* Vol. 3. (pp. 239–252).

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology, 118,* 219–235.
Reder, L. M., Charney, D. H., & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. *Memory & Cognition, 14,* 64–78.
Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 124–139.
Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology, 81,* 532–540.
Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16,* 371–416.
Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639.
Ross, B. H. (1989). Distinguishing types of superficial similarities: Effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 456–468.
Ross, B. H., & Kennedy, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 42–55.
Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501–518.
Schmidt, R. A. (1988). *Motor control and learning.* Champaign, IL: Human Kinetics.
Servan-Schreiber, E., & Anderson, J. R. (1990). Chunking as a mechanism of implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 592–608.
Siegler, R. S., & Jenkins, E. A. (1989). *How children discover new strategies.* Hillsdale, NJ: Erlbaum.
Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99,* 195–231.
Squire, L. R., & Frambach, M. (1990). Cognitive skill learning in amnesia. *Psychobiology, 18,* 109–117.
Van Lehn, K. (1991). Rule acquisition events in the discovery of problem-solving strategies. *Cognitive Science, 15,* 1–47.
Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 328–344.
Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1047–1060.

Appendix A

Procedure and Conditions in Experiment 3

Materials

1. Randomly assign rules from A1 and A2 to Rules 1–8.
2. Randomly create examples 1–8 to instantiate Rules 1–8.

Day 1

a. Study Examples 1–8.
b. Three-pass dropout learning phase for Examples 1–8.
c. Rule training: Rules 1–4 tested 32 times each.
d. Example training: Examples 3–6 tested 32 times each.
e. Rule transfer: Rules 1–8 tested both ways twice.
f. Example transfer: Examples 1–8 tested both ways twice.

Days 2–4

Phases c–f repeated.

Sample Materials

1. Rule 3: +2, −2
2. Example 3: Might Be
   Saturday 3 Hockey Monday 1
3. Rule 3 might be trained with queries like
   ___ ___ Hockey Friday 3
   for which the correct answer is Wednesday, 5.
4. Example 3 might be trained with queries like
   Saturday ___ Hockey Monday ___
   for which the correct answer is 3, 1.
5. A reverse test of rule 3 would be
   Thursday 7 Hockey ___ ___
   for which the correct answer is Saturday 5
6. A reverse test of example 3 would be
   ___ 3 Hockey ___ 1
   for which the correct answer is Saturday, Monday.

Appendix B

Conditions in Figures 15–18

Rule Transfer

Same direction: Rules 1–4 tested in direction practiced.
Reverse direction: Rules 1–4 tested in direction not practiced.
Example training: Rules 5 and 6 tested in either direction.
No training: Rules 7 and 8 tested in either direction.

Example Transfer

Same direction: Examples 3–6 tested in direction practiced.
Reverse direction: Examples 3–6 tested in direction not practiced.
Rule training: Examples 1 and 2 tested in either direction.
No training: Examples 7 and 8 tested in either direction.

Received April 14, 1993
Revision received December 20, 1993
Accepted December 30, 1993 ■