Journal of Experimental Psychology:
Learning, Memory, and Cognition
1997, Vol. 23, No. 4, 932–945

0278-7393/97/$3.00

The Role of Examples and Rules in the Acquisition of a Cognitive Skill

John R. Anderson, Jon M. Fincham, and Scott Douglass
Carnegie Mellon University

In 3 experiments, participants memorized 8 examples, each exemplifying a different rule. Participants were asked to extend these rules to new examples. They practiced applications of the rules to examples over a period of 4 days (Experiment 1) or 5 days (Experiments 2 and 3). Although these rules were bidirectional, an asymmetry gradually built up such that participants became more facile in using the rules in the practiced direction. Participants also showed an advantage when the initial study example was repeated or when test examples were repeated. It is argued that skill acquisition involves development of a complex set of strategies based on use of rules and retrieval of examples. Four overlapping stages of skill acquisition are described.

Two somewhat different types of theories have been offered about how examples are involved in the learning of skills. One theory proposes that learning progresses from reliance on examples to reliance on abstract rules. The other theory proposes that learning progresses in the reverse direction. We briefly review these two theoretical perspectives and then review the logic behind a series of experiments designed to evaluate whether one, both, or neither of these theories is correct.

Examples to Rules

A number of researchers (e.g., Anderson & Fincham, 1994; Novick & Holyoak, 1991; Pirolli, 1985; Ross & Kennedy, 1990) have suggested that initial problem solving involves explicitly referring to examples. Sometimes the examples are available in some physical medium like a textbook or in other cases they have to be recalled from memory. In either case the examples illustrate the solution of a similar problem and the problem solver analogically maps the solution of the example onto a solution for the current problem. With repeated practice, however, general rules develop and the specific example is no longer accessed. In the adaptive control of thought—rational (ACT–R) theory (Anderson, 1993), this is the principal means by which knowledge transitions from a declarative form (encoding of examples) to a procedural form (productions rules).

There are a number of lines of evidence offered for the example-to-rule account of the skill acquisition process. One is that participants stop looking to external examples and stop mentioning examples in concurrent protocols (Blessing & Anderson, 1996; Pirolli, 1985). A second is that participants' knowledge appears to become more general and less tied to the specifics of the example (e.g., Novick & Holyoak, 1991; Ross & Kennedy, 1990). A third is that participants develop a directional asymmetry in their use of the knowledge (Anderson & Fincham, 1994; Kessler, 1988; McKendree & Anderson, 1987; Pennington, Nicolich, & Rahm, 1995; Rabinowitz & Goldberg, 1995).

The directional asymmetry is important to the logic of the current experiments. Production rules in a theory like ACT–R are inherently directional. Consider the abstract fact that in the programming language LISP the function *car* retrieves the first element of a list illustrated by the example *(car'(abc)) produces the output a.* Although neither the abstract fact nor the example is inherently directional, different rules are required to use this knowledge for evaluation of code and generation of code:

Evaluation
IF the goal is to evaluate *car* applied to a list,
  and *x* is the first element of the list,
THEN the value will be *x*.

Generation
IF the goal is to get *x* from a list,
  and *x* is the first element of the list,
THEN use the function *car.*

Practice of one of these production rules will not transfer to the other. For instance, Kessler (1988) compared participants who practiced their knowledge of LISP in evaluation (going from code to result) with those who practiced generation (going from desired result to code). Participants were much more facile in the practiced direction. Although they replicated this asymmetry, Pennington et al. (1995) did find some transfer in the reverse direction.

Rules to Examples

An alternative theory (Logan, 1988) is that participants begin with a general procedure (perhaps implemented as production rules) but commit to memory specific examples of problem solutions. As examples repeat, these answers

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

This research was supported by Grant NSF 91-08529 from the National Science Foundation. We thank Marsha Lovett for her suggestions on the design of Experiment 3 and Helen Borek who ran Experiment 3. We would like to thank Marsha Lovett and Chris Schunn 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 via Internet to ja+@cmu.edu.

come to be retrieved, and participants come to enjoy the advantage of retrieval over computation. There are also a number of lines of evidence for this rule-to-example transition process. One of the most compelling is the advantage participants show for repetitions of problems on which they have been tested (Logan, 1988; Logan & Klapp, 1991; Rabinowitz & Goldberg, in press; Rothkopf & Dashen, 1995). This indicates that participants have memory for specific examples and can use this memory to facilitate processing. So, for instance, Logan (1988) found that participants responded more rapidly to repeated items in a lexical-decision task. Moreover, he showed that a model that assumed a race between retrieval of specific items predicted the power law decrease in the mean and variance of the response times.

In addition to showing effects of repetition of specific problems, Rabinowitz and Goldberg (1995) have shown symmetry in knowledge access for repeated problems. They used an alpha–arithmetic task in which participants were shown questions like $D + 3 = ?$ and had to respond with the letter that many digits (in this case *3*) forward in the alphabet (in this case the answer is *G*). Participants who practiced on just a few specific examples were quick when tested in the reverse direction for that specific example, $G - 3 = ?$. This is what would be predicted if participants committed to memory specific examples like $D + 3 = G$, which they could index with the *3* and the *G* and so retrieve the *D*. This contrasts with the asymmetry predicted on the basis of production rules.

Logic of Current Experiments

The research reported here is principally concerned with the relationship between two empirical signs of the development of a skill. One is the advantage for repeated examples, and the other is the development of asymmetries in access. Experiments in the past have focused on one or the other measure. Research looking at the repetition effect concluded that skill acquisition involved increased retrieval of examples. Research looking at the development of directional asymmetry concluded that skill acquisition involved the acquisition and strengthening of rules. The research reported in this article looks simultaneously at both sorts of effects.

All three experiments reported here used the paradigm and materials introduced by Anderson and Fincham (1994). In the first part of these experiments, participants committed to memory eight specific facts such as *Skydiving was practiced on Saturday at 5 p.m. and Monday at 4 p.m.* Although participants were not aware of it at the time, they were learning examples of rules about the time relationship between the two events for that sport. In this case, the rule was that the second skydiving event always occurs two days later and one hour earlier. We denote this rule as *+2, −1.* Only after memorizing these examples were participants given an explanation for the significance of the examples, and they were then tested with problems (which could be old or new) in an interface like that illustrated in Figure 1. Participants were given either the first or second time (day plus hour) and had to predict the other time. In the case in Figure 1 in which the first time is Friday at 3, they would have to predict that the second time was Sunday at 2. Participants made their prediction by clicking the relevant elements in the boxes below. We were interested in the speed and accuracy with which they could do this. The example in Figure 1 involved going from the first time to the second time, but Anderson and Fincham trained participants on eight examples, and half of them involved going from the second time to the first time. Anderson and Fincham found that after 40 or more trials of always going in the same direction for a particular rule, an asymmetry appeared such that participants were more accurate and faster in the practiced direction than in the reverse direction. One of the things we wanted to do in this research was to look at the emergence of this asymmetry. In this research we were interested in performance on problems that involved an example that had been studied or an example that had not been studied and in testing a rule in the direction that it had been practiced or in the reverse direction. Table 1 summarizes abstractly the various possible conditions and the different ways that participants might have answered the questions in these conditions. The possible test problems could be classified by whether the test problem was old or new and by the direction of testing. Two methods were applicable in all conditions. Analogy was one of these, and initially it was the only method available to the participant for new problems. The participant had to retrieve the memorized skydiving example and map it to the current problem. However, with practice it was possible for the participant to abstractly characterize the relationship—*Skydiving is two days later and one hour earlier.* The participant could retrieve this abstract, declarative representation and so circumvent the need to retrieve the study example.

The remaining two methods could only apply in certain conditions. One was use of production rules. For instance, if participants were practicing going from the second time to the first time for skydiving, they could form a production rule such as the following:

IF the question is about skydiving,
  and the second day is D2 and the second hour is H2,
  and D1 is two days before D2,
  and H1 is one hour after H2,
THEN the first day is D1 and the first hour is H1.

Unlike analogy or the abstract declarative rule, this production rule can be used only in the same direction as practiced, but it is indifferent to whether the test problem is old or new. Finally, the participant might have retrieved a specific example that matches the test problem. This reference example could be either the study example (Experiment 1) or a repeated test problem (Experiments 2 and 3). A retrieved example such as *Skydiving was practiced on Saturday at 5 p.m. and then Monday at 4 p.m.* provides the answer in either direction because it stores both the first time and the second time. However, without calling on analogy it provides no help for a new problem that has not been tested.

Thus, we could use degree of directional asymmetry to assess how much participants were relying on production

Test Phase Display:

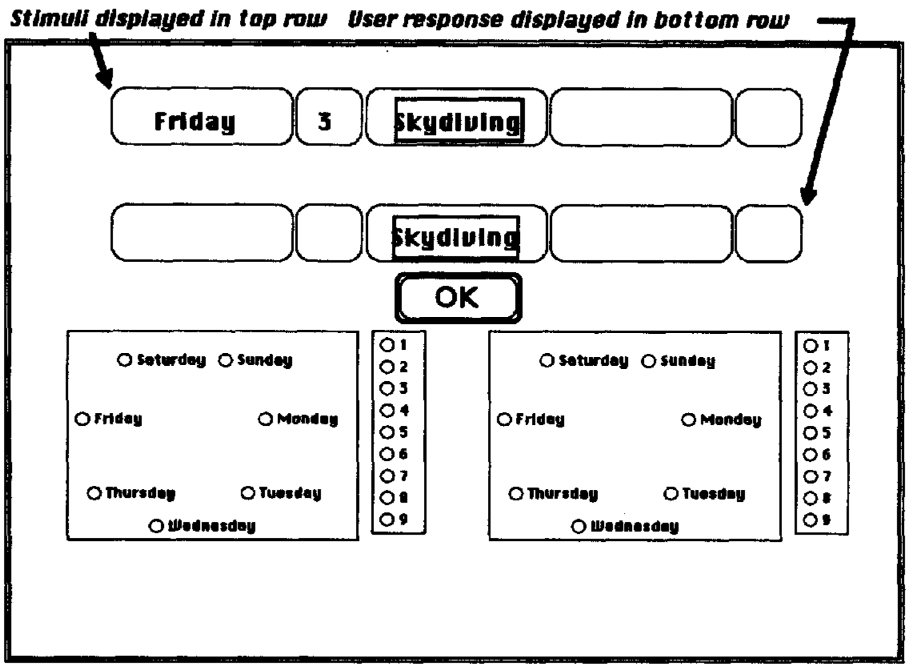

Figure 1. An example of the interface used in Anderson and Fincham (1994) and in the research reported here. From "Acquisition of Procedural Skills From Examples," by John R. Anderson and Jon M. Fincham, 1994, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* p. 1331. Copyright 1994 by the American Psychological Association.

rules and the advantage for repeated examples to assess how much they were relying on the retrieval of examples. In contrast, analogy and abstraction would produce no differences among the cells in Table 1. This would give us a fuller picture of skill development than has been painted by past research.

The interpretation of a directional asymmetry for old examples is somewhat ambiguous. Because one might argue that examples are preferentially retrieved in the direction of practice (but Anderson & Fincham, 1994, showed that there is little if any asymmetry in retrieval of examples), the strongest evidence for emergence of production rules is asymmetry with new examples.

In summary, this research used the same methodology as in Anderson and Fincham (1994) but addressed the relative contribution of rules and retrieval of examples to the emergence of a skill. It also provided evidence on the time course of the development of production rules and the development of reliance on retrieval. To foreshadow, it would provide evidence for a much more complex conception of skill acquisition than has been acknowledged heretofore. In particular, it would provide evidence for all four methods of responding that are identified in Table 1. We argue for what we call a *four-stage model of skill acquisition,* which involves four overlapping stages—the participant starts with analogy to study examples, develops abstract rules, and slowly moves to use of production rules and retrieval of specific examples.

Table 1
Methods Available to Solve Different Types of Problems

Direction of testing	Test problem: Old	Test problem: New
Practiced	Analogy Abstraction Production rule Retrieval of example	Analogy Abstraction Production rule
Reverse	Analogy Abstraction Retrieval of example	Analogy Abstraction

Experiment 1

On each of the 4 days, participants practiced eight rules in one direction. Two of these rules were also tested in the other direction right away on Day 1, two more starting with Day 2, two more starting with Day 3, and the final two starting with Day 4. We were interested in whether participants would be faster in the more practiced direction. We expected to see more asymmetry for rules that were reversed after more practice in one direction. We were also interested in how well participants would do on the specific examples they had originally studied. Therefore, each day we tested each of the original examples once. If participants were performing better for these original examples, then it would be evidence that they were solving these problems by means of retrieving the study example.

The other thing we wanted to investigate in this experiment was whether participants were aware of whether they were using a rule or an example. Therefore, we asked half of the participants to tell us after each question how they answered the question—whether by rule or by example. Reder and Ritter (1992) and Siegler (1987) have had some success in getting participants to report whether they were using retrieval or rule-based computation. It is possible that this self-reflection on strategy might change the results. It certainly disrupts the pace of the experiment, and we would not be surprised to see some overall effect on speed. As a control to determine whether strategy reflection changed the basic pattern in participants' behavior, we asked the other half of the participants to engage in a reporting task that did not require strategy reflection—to rate whether they thought they answered each question faster or slower than average. We refer to these two conditions as the rule and rate conditions.

Method

Participants. Twenty-eight Carnegie Mellon University undergraduates were recruited to participate in this 4-day experiment. The first session lasted 2 hr, whereas the remaining 3 sessions lasted between 45 min and 1 hr. Participants were paid $4 per session. In addition, they received between $8 and $16 bonus pay that was dependent on performance.

Materials. Table 2 gives the abstract structure of the eight rules. Each participant saw different randomly generated examples that embodied these rules. All four possible relations (−2, −1, +1, +2) between the two hours and days occurred twice in the eight rules. Direction in Table 2 refers to whether the participant predicted the second date from the first (Right) or the first from the second (Left). The two groups differed in terms of which rules were tested in which direction. Participants were randomly assigned to either Group 1 or Group 2. The rules were broken up into four pairs (denoted by letters in Table 2). One pair was randomly selected without replacement to be introduced in the reverse direction on each day of the experiment.

Eight study examples were randomly generated, one for each rule. For each day's training session, 15 new examples were generated for each direction in which each rule was tested. These training examples for each rule were different from one another and from the study example. However, there was no effort to avoid repetitions of examples across days.[1] These 15 examples plus the study example were used to provide 16 training tests for each rule.

Table 2
Abstract Structure of the Rules Used in the Experiment

		Direction practiced	
Pair	Day/hour	Group 1	Group 2
A	+1/+2	Right	Left
A	−2/−1	Left	Right
B	−1/+1	Left	Right
B	+2/−2	Right	Left
C	−1/−2	Right	Left
C	+2/+1	Left	Right
D	+1/−1	Left	Right
D	−2/+2	Right	Left

Table 3
Rules Tested During Each Block for Each of the 4 Days

Rule	Day 1	Day 2	Day 3	Day 4
1	Right Left	Right Left	Right Left	Right Left
2	Left Right	Left Right	Left Right	Left Left
3	Right	Right Left	Right Left	Right Left
4	Left	Left Right	Left Right	Left Right
5	Right	Right	Right Left	Right Left
6	Left	Left	Left Right	Left Right
7	Right	Right	Right	Right Left
8	Left	Left	Left	Left Right

Note. There were 16 blocks per day.

Procedure. The same basic interface illustrated in Figure 1 was used in all phases of the experiment. The first day began with an initial exposure to the eight study examples followed by a three-pass drop-out phase. During the initial exposure phase, participants were told to study each of the eight examples and to copy them from the top row to the bottom row. This gave them the opportunity to memorize the examples in addition to familiarizing themselves with the interface before beginning the drop-out phase. In the drop-out phase, participants were shown just the sport name and had to reproduce the two days and two times. In each pass of the drop-out phase, they were tested repeatedly over the items until they had correctly recalled the times for each sport name. As soon as they recalled the times for a name, it was dropped out of the pass. The pass stopped when there were no more items. Then participants were tested on all the items anew for another pass. The drop-out phase was followed by the training phase in which participants would see just the first day and hour and have to predict the second or vice versa. Subsequent days involved only the training phase.

Table 3 illustrates the basic training pattern for the eight rules across the 4 days. Participants received 16 training blocks per day. Each block consisted of one instance of each of the eight rules in the practiced direction (four left to right and four right to left). In addition, at the onset of each new session, another two of the eight rules were also trained in the reverse direction on each block. Thus on Day 1, participants trained on blocks of 10 stimuli, containing eight instances corresponding to the practiced direction of the eight rules and two instances in the unpracticed or reverse direction. Because there had been no prior training, the designation of practiced and reverse direction was arbitrary for these two rules.

[1] There were 56 possible examples that tested a rule with a 1-hr difference and 49 that tested a rule with a 2-hr difference. Over the course of the experiment, there were between 75 (items reversed on Day 4) and 120 (items reversed on Day 1) tests with nonstudy examples for a rule. This meant, depending on the rule and the day it was reversed, an example would be expected to be tested between 1.36 and 2.50 times for that rule by experiment's end. Because examples were randomly generated from day to day, the actual repetitions varied from example to example.

Similarly, on Day 2 each block consisted of eight practiced and four reverse; on Day 3 each block contained eight practiced and six reverse; and finally on Day 4, all rules were trained in both directions. Thus, once a rule was tested in the reverse direction it continued to be tested in both directions throughout the experiment.

For each of the rules the study example was presented as a training test once during Blocks 7 through 10 instead of as a new instance. If a particular rule was being trained in both directions, both instances were replaced with the corresponding study example. These study examples were randomly placed in one (and only one) of the Blocks 7–10 on each day.

Fifteen participants were in the rule condition, and 13 were in the rate condition. Participants in the rule condition were asked to characterize their problem-solving process after each trial. They reported whether they used the initial study example to solve the problem or simply used the corresponding rule. Participants in the rate condition rated whether they had been faster or slower than usual on that trial. They answered these questions by clicking one of two boxes. The actual prompt for the rule participants was as follows:

> Please click the button that most accurately describes your most recent problem solving . . .
>
> [Button 1] I'm referring to the original EXAMPLES to solve the problems.
> [Button 2] I'm referring only to the underlying RULES to solve the problems.

In contrast the rate participants saw the following prompt:

> Please click the button that most accurately describes your most recent problem solving . . .
>
> [Button 1] I'm performing WORSE now than I've been doing previously.
> [Button 2] I'm performing BETTER now than I've been doing previously.

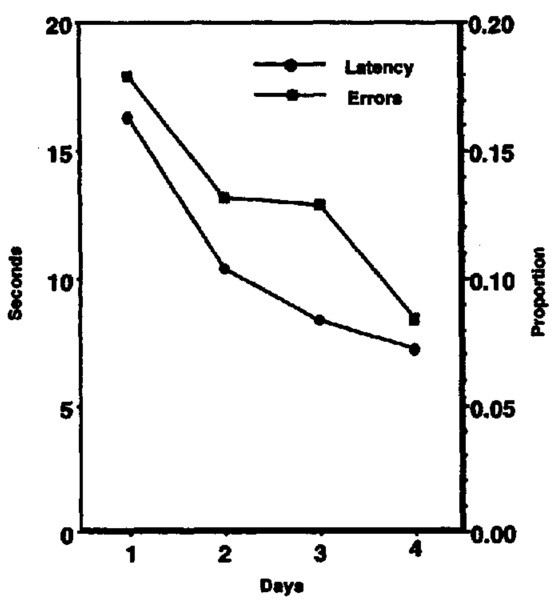

Figure 2. Decrease in latency and error rate in Experiment 1 as a function of day. Data are averaged over the two rules that are reversed that day and over the two directions.

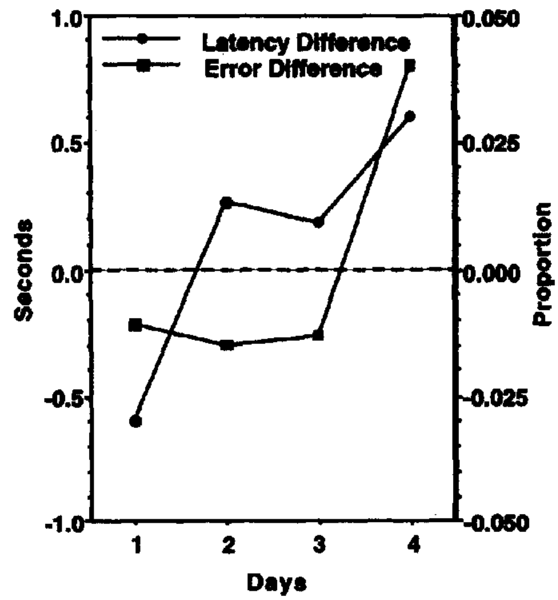

Figure 3. Directional difference, reverse − practiced, in Experiment 1 as a function of day. Data come from the two rules that are reversed that day.

Results

Throughout the article we report all effects from analyses of variance (ANOVAs) that are significant at the .05 level or lower.

We performed a number of analyses on different subsets of the data. To test for the emergence of the directional asymmetries predicted by the examples-to-rules theory, we computed ANOVAs focusing on just those two rules that were being used for the first time in both directions on each day (the main diagonal in Table 3). We took the average performance measured for these rules over the 16 blocks of that day. ANOVAs were performed on both latencies and error rates in which the variables were day (1–4), direction of testing (practiced or reverse), and report (rule or rate).[2] Given that latencies and errors are on such different scales, we report the *MSE*s for errors to an extra decimal place of precision. There was no effect of report condition on either measure, nor any significant interaction with report condition. There were significant effects of day: $F(3, 72) = 91.96$, $MSE = 32.39$, $p < .001$, for latency; $F(3, 78) = 3.05$, $MSE = 0.060$, $p < .05$, for error rate. Figure 2 illustrates the improvement over days for both measures. There were not significant effects of direction for either measure: $F(1, 24) = 0.39$, $MSE = 6.08$, for latency; $F(1, 26) = 0.03$, $MSE = 0.008$, for error rate. However, there were significant day-by-direction interactions for both measures: $F(3, 72) = 2.95$, $MSE = 4.51$, $p < .05$, for latency; $F(3, 78) = 5.45$, $MSE =$

[2] Two participants were excluded from the latency analysis because they had unusually high error rates and so mean response latencies could not be calculated for all conditions. This left us with 13 participants in both the rule and rate conditions.

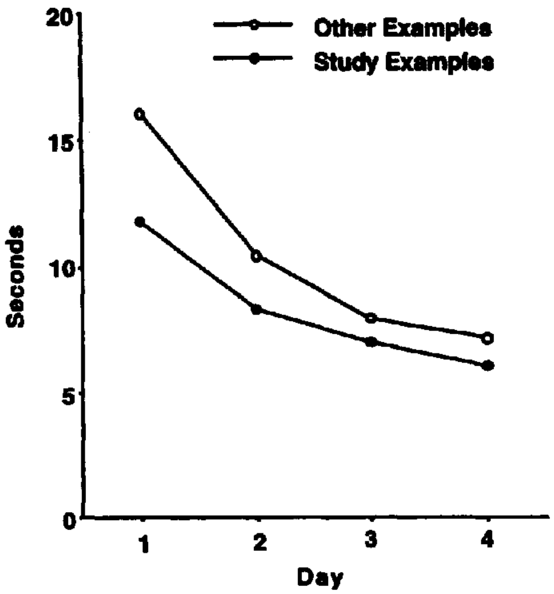

Figure 4. Average time for study examples versus other examples in Experiment 1 as a function of day.

0.007, $p < .005$, for error rate. Figure 3 plots the difference between the two directions, reverse − practiced, as a function of day. Both latencies and errors display the expected increase in the directional asymmetry, with the latency effect becoming positive on Day 2 and the error effect becoming positive on Day 4. We have no explanation for why these effects are negative on Day 1. The distinction is perfectly arbitrary for that day and must reflect the accidents of assignment. However, the shift over days is consistent with a move to rule-based processing and is a result that would not be predicted if skill acquisition depended only on retrieval of examples.

The data in Figures 2 and 3 represent only a quarter of the data—those items that are being introduced for the first time in the reverse direction on a particular day. We also did an analysis of all items on Day 4 (the last column in Table 3) to see whether there were any effects being hidden by this selection process. We would expect the directional effect to be larger for items reversed on later days. ANOVAs were performed in which the variables were day the rule was introduced in the reverse direction, direction of testing, and report condition. Again there were no significant effects or interactions of report condition on either errors or latency. With respect to latency, there were significant effects of direction, $F(1, 26) = 10.92$, $MSE = 0.11$, $p < .01$, and an interaction between day of introduction and direction such that the effect of direction was larger if the rule was reversed on a later day, $F(3, 26) = 9.12$, $MSE = 0.22$, $p < .01$. The directional effect was −0.04 s for items reversed the first 2 days, but 0.39 s for items reversed the last 2 days. With respect to errors, there was a main effect of day of introduction such that participants were more accurate the longer they had been practicing the rules in both directions, $F(3, 26) = 3.06$, $MSE = .0043$, $p < .05$, and an interaction between day of reversal and direction, $F(3, 26) = 3.11$, $p < .05$. The directional effect was −0.3% for items reversed the first 2 days and 2.8% for items reversed the last 2 days. This analysis of the last day is consistent with the analysis in Figure 3 of the effect associated with the rules reversed for the first time each day. Both indicate that directional asymmetry increases with differential practice in one direction.

To assess whether participants were retrieving specific examples, we looked at performance on the study example that was a highly practiced example. The effect of testing the study example was analyzed with respect to the middle segment (Blocks 7–10), in which participants either saw the study example or did not see the example. Here we looked at data from all examples that appeared in the segment. We separated mean time for the study example from mean time for all the other examples that appeared in those trials. The latency data are displayed in Figure 4. There was a significant effect of whether the item was the study example, $F(1, 27) = 41.04$, $MSE = 5.94$, $p < .001$; of day, $F(3, 81) = 45.61$, $MSE = 12.87$, $p < .001$; and an interaction between the two variables, $F(3, 81) = 14.98$, $MSE = 2.25$, $p < .001$. The main effect of study example is consistent with the view that participants are solving the study example problems by retrieval of the examples.

The data from the middle segment (Blocks 7–10) of each day were also analyzed for frequency with which participants in the rule condition reported use of examples. Figure 5 shows the frequency with which participants reported that they used examples as a function of day. An ANOVA was performed in which the variables were day and whether the problem was the study example. Participants were overall more likely to report use of examples on the study example, $F(1, 14) = 8.39$, $MSE = 0.025$, $p < .05$. The overall effect of days was not significant, $F(3, 42) = 1.13$, $MSE = 0.111$, $p >$

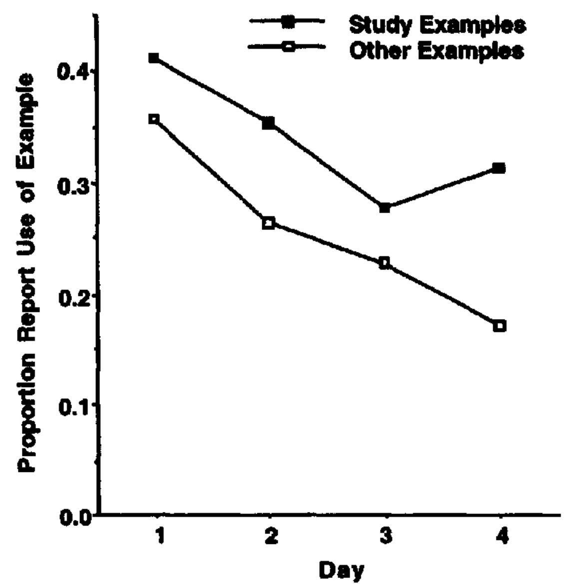

Figure 5. Proportion of times participants reported they solved a problem by reference to an example.

.3. However, a specific contrast looking for a linear trend was marginally significant, $t(42) = 1.76, p < .05$, one-tailed. The interaction between the variables was not significant, $F(3, 42) = 0.85$, $MSE = 0.016$. This reduction in use of examples would be predicted if participants were switching from analogy to rule-based processing. Although the effects in Figure 5 are in the expected directions, they seem rather small. We are suspicious about how well participants were able to give accurate reports. One participant reported 100% use of examples throughout the experiment, 1 participant reported 0%, and 3 participants reported under 5% on all days. These extremes strike us as implausible, suggesting that at least these participants were not giving veridical reports.

One of the curious results was that even on Day 1, participants were reporting a majority of rule use even though directional asymmetries only developed on later days. This suggests that on Day 1 they were using "rule" to refer to something nonprocedural, such as a declarative representation of the transformation—for example, *The second time for hockey is 1 day later and 2 hours earlier than the first time.* These are the "abstractions" referred to in Table 1. Such declarative representations of the rule need not have any directionality. It is not clear that participants should be able to report use of production rules. According to the ACT–R (Anderson, 1993) theory, production rules represent implicit knowledge and so are not reportable. Perhaps, the rather bizarre reporting of some participants reflects the fact that they did not have conscious access to the basis for their decisions.[3]

Although the overall frequency of rule use was high and the overall frequency of example use never fell below 20% on any day, only 6 of the 15 participants reported a nonzero frequency of both uses on all 4 days. The latency data from these 6 participants are displayed in Figure 6. These data are

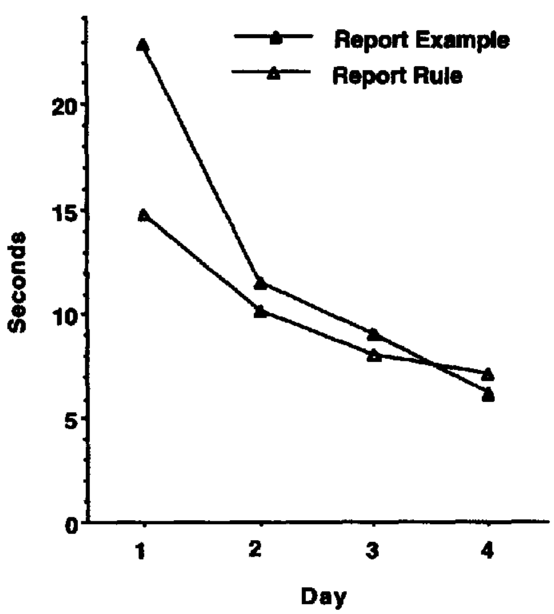

Figure 6. Time to respond as a function of whether rule or example is reported.

from all blocks of the experiment, not just the middle segment as in Figures 4 and 5, and therefore the study examples contribute very little to these data. There were significant effects of method reported, $F(1, 5) = 17.58$, $MSE = 3.85, p < .01$, with participants being faster if they reported using a rule; a significant effect of days, $F(3, 15) = 30.77$, $MSE = 11.13, p < .001$, and a significant interaction between the two effects, $F(3, 15) = 8.18$, $MSE = 5.54, p < .001$. The fact that participants were faster when they reported rule use supports the claim that this is more efficient than analogically extending the example. The interaction may reflect a switch in what participants mean when they say they are using examples. Originally, they may be retrieving the original example and analogically extending it. However, the test examples can repeat across days, and later participants may be more and more recalling having answered that specific example. Then they could simply recall the answer and so circumvent the need to analogically extend the study example or use a rule.

Discussion

In Table 1 we listed four methods for responding to the problems. There is evidence in the experiment that participants were at different times using all four methods. To review, the four methods in order of increasing efficiency are (a) *analogy to examples:* retrieval of the study example and analogical extension of the example to the current problem, (b) *declarative abstractions:* after a few applications participants probably consciously identify the rule associated with the sport and apply it, (c) *production rules:* with extensive practice participants develop a procedural embodiment of the rule, and (d) *retrieval of examples:* retrieval of an example that matches the target problem and simple readout of the answer.

We propose that in our experiment participants started out using the retrieval method *d* for the study example and the analogy method *a* for the other examples. As the experiment progressed, they moved to the rule methods *b* and *c* and then to the retrieval method *d* for nonstudy but previously presented examples. However, the test examples repeated only infrequently and haphazardly across days. Therefore, retrieval of test examples may not be a significant variable in this experiment.

The following is a review of the data that are consistent with this proposal:

1. Participants were slower on problems for which they reported example use. This is consistent with the claim that the analogy method *a* is slower.

2. Participants reported a majority of rule use even on the first day. However, directional effects did not appear until later days. This suggests that performance is initially mediated by declarative abstractions as in method *b*. Rule reports increased over days suggesting a shift away from analogy.

3. An asymmetry in the application of the knowledge appeared on later days, which is consistent with moving to the production rules as in method *c*.

[3] We thank one of the reviewers for pointing this out to us.

4. Participants reported less rule use for the study example, which is consistent with using the retrieval method *d* for those problems. Participants were faster on the study examples, which is consistent with the claim that the retrieval method is faster.

5. The interactions with day in Figures 4 and 6 are consistent with the view that participants are switching to answering more problems by retrieval of test examples, resulting in a loss of advantage for the study examples (Figure 4) and a loss of advantage for cases in which rules are reported (Figure 6). However, this may be a floor effect, and the haphazard repetition of test examples in this experiment might not have been sufficient to produce such effects.

One implication of these results is that the speed-up in rule execution cannot be totally a result of retrieval of examples as suggested by Logan (1988). Participants are getting faster while they are developing increasing directional asymmetry. Carlson and Sohn (1996) have also shown that participants get faster at executing procedures. They use the fact that different procedures maintain their relative difficulty with practice to rule out any simple retrieval explanation of this speedup.

The complementary implication of these results is that speedup cannot be entirely a result of increased rule use. Participants were achieving their most rapid processing by direct retrieval. The next two experiments provide further evidence for the importance of retrieval of specific examples to skill acquisition.

It is interesting that at the beginning of the experiment the participants were reporting a majority of rule use but not showing any asymmetries in their use of these rules. This we took as evidence for the use of declarative rules such as *The second time for hockey is 1 day later and 2 hours earlier than the first time.* Although such declarative statements might seem to have a directionality in their statement, they can be used just as easily in either direction. Retrieving and using such declarative rules in ACT–R requires *interpretative* production rules such as

Forward
IF the question is about a sport,
  and the first day is D1 and the first hour is H1,
  and the second time for that sport is *n* days later
    and *m* hours earlier,
  and D2 is *n* days after D1,
  and H2 is *m* hours before H1,
THEN the second day is D2 and the second hour is H2.

Backward
IF the question is about a sport,
  and the second day is D2 and the second hour is H2,
  and the second time for that sport is *n* days later
    and *m* hours earlier,
  and D2 is *n* days after D1,
  and H2 is *m* hours before H1,
THEN the first day is D1 and the first hour is H1.

These interpretive production rules match the declarative rule in their clause *the second time for that sport is n days later and m hours earlier* and simply match patterns differently going forward or backward. It is no more difficult to go in one direction than another. Moreover, retrieval of such declarative rules eliminates the need to engage in analogy producing the advantage displayed initially in Figure 6.

Because such interpretative production rules have to retrieve the declarative rule, they are slower than a production rule, such as the following, that especially encodes the rule for a sport:

Forward-Hockey
IF the question is about hockey,
  and the first day is D1 and the first hour is H1,
  and D2 is 1 day after D1,
  and H2 is 2 hours before H1,
THEN the second day is D2 and the second hour is H2.

If participants practice such rules in only one direction, they will show a directional advantage for that direction over the unpracticed direction. In the unpracticed direction, participants would have to fall back on retrieval of the declarative rule by means of interpretative production rules.

Experiment 2

Experiment 1 is consistent with the view that skill progresses from analogical use of a specific example to rule-based performance to retrieval of specific examples and as such is consistent with a synthesis of the ACT–R rule compilation and the Logan (1988) exemplar models. However, there were a couple of weak points in the empirical evidence. First, although there was a significant day-by-direction interaction, the main effect for direction did not reach significance in the cross-day analysis (Figure 3).[4] In part this may have occurred because a quarter of the items, those assigned to both forward and reverse directions on Day 1, never received one-way practice. Therefore, we decided to extend this experiment to 5 days and to introduce items in the reverse direction only on the second day. On Day 1 participants would practice all items in only one direction. Also to enhance the effect, we doubled the number of training trials per day.

Another problem was that the evidence was weak for the direct retrieval of anything but the study example. The evidence for direct retrieval of other examples was the disappearance of effects across days in Figures 4 and 6. This could be interpreted as increased retrieval of test examples that would eliminate the effects of the variables in Figures 4 and 6. However, it could also be interpreted as a floor effect. Test examples could only repeat across days in Experiment 1; thus, repetition was infrequent and haphazard. What is needed is a designed manipulation in which some examples are selected to be repeated frequently and others are not. As we noted in the introduction, this has been used in other research to provide evidence for the exemplar model. Therefore, Experiment 2 repeated specific test examples and never repeated the study examples.

Another difference between Experiment 2 and Experiment 1 is that we looked at participants' memory for the study examples 5 days later at the end of the experiment. We

[4] However, the main effect was significant in the analysis of all rules on Day 4.

suspected that in this experiment, in which the study example was not being repeated at all during the 5 days, participants might forget the original study examples. This would provide compelling evidence that they had lost the declarative origins of their skill and could no longer be responding by analogy to the example.

A final difference between this experiment and the previous one is that we decided to forgo gathering of reports. Although the report data were consistent with the overall picture, we felt that many participants were not able to give useful reports of what they were doing.

Method

Participants. Twenty-two Carnegie Mellon University undergraduates were recruited to participate in this 5-day experiment. The first session lasted 2.5 to 3.0 hours, whereas the remaining four sessions lasted between 1 hr and 1 hr 30 min. Participants were paid $4 per session. In addition, they received between $8 and $16 bonus pay that was dependent on performance.

Procedure. The same interface (Figure 1) and material structure (Table 2) were used as in Experiment 1. Participants had 32 training blocks per session. Each block consisted of one instance of each of the eight rules in the practiced direction (four left to right and four right to left). In addition, at the onset of each new session beginning with the second session, a new pair of the eight rules was also tested in the reverse direction. Thus on Day 1, participants trained on blocks of eight stimuli corresponding to the eight rules. On Day 2, participants trained on blocks of 10 stimuli, containing eight instances corresponding to the practiced direction of the eight rules and two instances corresponding to the pair selected for training in the reverse direction. Similarly, on Day 3 each block consisted of eight practiced and four reverse; on Day 4 each block contained eight practiced and six reverse; and finally on Day 5, all rules were trained in both directions.

The instances were constructed such that one fourth of the training instances of a particular rule were identical (however, different from the study example for that rule). Thus, if a rule was tested in just one direction, 8 of the 32 tests would involve this repeated instance, and if a rule was being trained in both directions, the same instance would be repeated 8 times in both directions for a total of 16 repetitions. None of the other examples repeated within days, and the study example was never repeated. Within these constraints, the materials were randomly generated. Thus, there would be some repetition across days, but this would be minor compared with the amount of practice of the examples selected for repetition.

At the end of Day 5, a posttest was administered to test participants' memory of the initial study exemplars learned on Day 1. There were four blocks. In each block, participants were tested on each of the eight exemplars, cued with the sport name, and asked to recall the initial exemplar. They were not given feedback.

Results

An advantage of Experiment 2 was that it was possible to look for both directional asymmetries and effects of example repetition in the same statistical analysis to test the logic set forth in Table 1. ANOVAs were performed on the latency and error data of the experiment for Days 2–5, looking at those items that were tested in the reverse direction for the first time that day. The variables in the ANOVAs were day, direction (practiced vs. reverse), and stimulus type (repeated or unique). Figure 7 shows the effect of days for latency and error rate. The effect of day was significant for latency, $F(3, 63) = 17.45$, $MSE = 31.76$, $p < .001$, but not for error rate, $F(3, 63) = 1.08$, $MSE = 0.017$, $p > .3$. Although the overall effect in the errors was not significant, a test for the linear trend was marginally significant, $t(63) = 1.71$, $p < .05$, one-tailed. The effects are weaker in Figure 7 than in Figure 2 because this analysis excluded Day 1. The effect of direction (shown in Figure 8) was also significant: $F(1, 21) =$

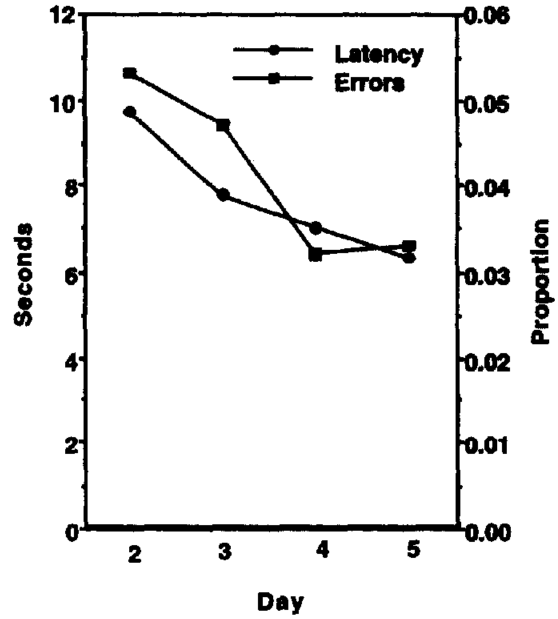

Figure 7. Decrease in latency and error rate in Experiment 2 as a function of day. Data are averaged over the two rules that are reversed that day and over the two directions.

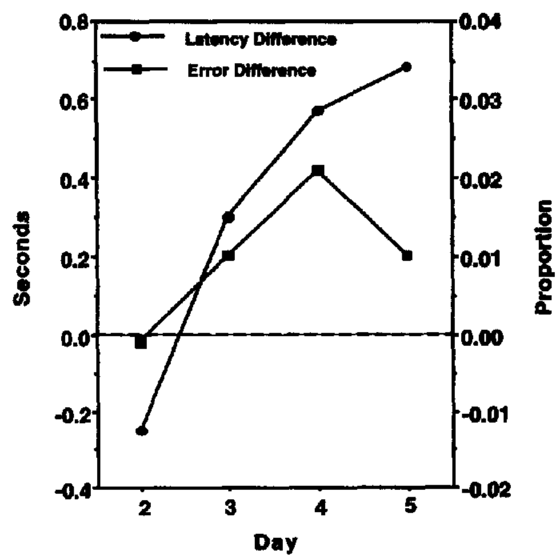

Figure 8. Directional difference, reverse − practiced, in Experiment 2 as a function of day. Data come from the two rules that are reversed that day.

8.75, $MSE = 4.30$, $p < .01$, for latency; $F(1, 21) = 6.09$, $MSE = 0.007$, $p < .05$, for error rate. Thus, the increased number of trials and the extra day of practice has produced a significant directional effect overall. As in the previous experiment, the day-by-direction interaction was significant for latency, $F(3, 63) = 3.88$, $MSE = 3.90$, $p < .05$. However, in contrast to the previous experiment, it was not significant for error rate, $F(3, 63) = 0.81$, $MSE = 0.009$. The latency effects showed a clear trend to increasing asymmetry that was significant, $t(63) = 3.38$, $p < .001$, whereas the error trend was not significant, $t(63) = 0.97$, although it was in the expected direction. Basically, we have replicated Experiment 1's evidence for the emergence of directional asymmetry that supports the contribution of production rules to skill development.

The main effect of stimulus type was also significant for both latency (8.01 s for unique and 7.74 s for repeated), $F(1, 21) = 9.61$, $MSE = 2.64$, $p < .01$, and for error rate (.051 for unique and .041 for repeated), $F(1, 21) = 5.78$, $MSE = 0.007$, $p < .05$. Thus, the data also indicated that participants showed an advantage for repeated examples supporting the contribution of retrieval of examples to skill acquisition. Although the effects were significant, they were not very large. Also, the interaction between direction and stimulus type was not significant: $F(1, 21) = 1.32$, $MSE = 3.82$, for latency; $F(1, 21) = 0.92$, $MSE = 0.005$, for error rate. One might have expected a directional effect only for unique stimuli and not for repeated stimuli. If participants have stored the repeated example, they should be equally capable of retrieving the first or second day from it. The lack of a significant interaction and the small effect of repetition suggests that participants were often responding by rule even for the repeated stimuli. There were no other significant interactions.

As in Experiment 1, we did a complete analysis of performance on the last day. The variables in the ANOVA were day on which the rule was introduced in reverse direction (Days 2–5), direction (practiced or reversed), and stimulus type (repeated or unique). With respect to latency, there were significant effects of direction, $F(1, 21) = 12.09$, $MSE = 0.72$, $p < .001$; of stimulus type, $F(1, 21) = 24.95$, $MSE = 0.81$, $p < .001$; and a significant interaction between day and direction, $F(3, 63) = 3.16$, $MSE = 0.81$, $p < .05$. The interaction was such that the directional effect increased with how recently the rule had been reversed: an average 0.04 s difference for items reversed on Days 2 and 3 and an average 0.60 s for items reversed on Days 4 and 5. With respect to error rate, there is only a significant effect of stimulus type, $F(1, 21) = 6.77$, $MSE = 0.002$, $p < .05$. Table 4 shows the data classified by direction and stimulus type. Although there appear to be main effects of both variables on both dependent measures, there is very little suggestion of an interaction. The absence of an interaction is counter to the expectation derived from retrieval of examples that the directional effect should occur for unique but not for repeated examples.

The other data we looked at involved participants' ability to recall the original example. At the end of Day 5, they were unable to recall 37.2% of the original examples. These errors appeared to be random, and there was no tendency for participants to misrecall the repeated example as the study example.[5] This poor recall occurred while they were making only 3.0% errors in answering questions in the training phase. This constitutes clear evidence that participants were no longer making all their responses by reference to the original study examples. That is, they had forgotten the declarative origins of some of their knowledge. The correlation across subjects between percent errors in the posttest recall and percent errors in rule training (overall trials on Day 5) was negative and low ($r = -.249$). If we split the 22 participants into two groups, the 11 with less than a 40% error rate in recall and the 11 with greater than 40% errors, the former make 3.7% errors in the rule training on the last day and the latter 2.4% errors, $t(20) = 0.60$. Of the 22 participants, 15 had examples that they could recall and others that they could not. For those participants there was no significant effect of ability to recall the example on latency in the rule training, $t(14) = 0.75$, $MSE = 0.74$, and at best a marginal effect on error rate, $t(14) = 1.70$, $MSE = 0.0009$, $p > .05$, two-tailed. The error effect takes the form of participants making fewer errors (2.2% vs. 3.2%) on rules for which they can recall the example than on rules for which they cannot. However, this is opposite of the direction of the between-subject effect.

Table 4
Latency (in Seconds) and Error Rates (in Percentages) on Day 5

Test problem	Practiced direction: Latency	Practiced direction: Error	Reverse direction: Latency	Reverse direction: Error	*M*: Latency	*M*: Error
Repeated	6.14	2.5	6.51	2.5	6.33	2.5
Unique	6.67	3.3	6.93	3.9	6.80	3.6
M	6.40	2.9	6.72	3.2	6.56	3.0

Discussion

This experiment provides further evidence for the four-stage model of skill acquisition. Over the course of the experiment, there was a gradual increase in the amount of asymmetry. The asymmetry appears roughly on the third day or after 64 trials, which agrees with the amount of practice used by Anderson and Fincham (1994). This is evidence for a transition to a production-rule-based processing. Moreover, participants appear to have forgotten a great many of the study examples by the end of the experiment. Participants do show a speed advantage for repeated examples, which suggests that some of these they are solving by explicit recall. This is the direct evidence for the existence of the fourth stage of responding for which we only had indirect evidence in the previous experiment. As a final comment, the gradual emergence of effects and the existence of directional effects for repeated examples suggests that

[5] Probably much of the forgetting is due to interference (but not intrusion) of the many test problems.

these stages are not disjoint but substantially overlap. There is a gradual shift from one method of responding to another method.

Experiment 3

The results of Experiment 2 were generally positive with respect to the four-stage model offered in this article. The effects shown in Table 4 are exactly what would be expected from Table 1 if participants were responding by both production rules and retrieval of examples. However, we did not feel entirely satisfied with the results of the experiment. The size of the effect of the repeated examples was smaller than we expected, and the failure to find any interactions of repetition with the direction manipulation was somewhat surprising. One might have expected the repeated problems to be solved by retrieval and so not show the directional effect attributed to production rules.[6] Therefore, we decided to do another experiment to obtain a larger effect of example repetition. In Experiment 3, the example was repeated on 50% of the trials for a rule rather than just 25%.

The small size of the effect associated with repeated examples in Experiment 2 (compared with other research—e.g., Rabinowitz & Goldberg, 1995; Rothkopf & Dashen, 1995) may also have been related to the fact that we used a within-subjects manipulation (in contrast, between-subjects manipulations were used in the other research) with some items unique and some items repeated. Participants may adopt special strategies when all of the items are unique or repeated. To assess this possibility we compared three conditions. One condition (the mixed condition) was as in Experiment 2 in which half of the trials for a rule involved the repeated example and the other half of the trials were unique. A second condition (the unique condition) involved only unique examples. The third condition (the repeated condition) involved only repeated examples in which each rule was tested throughout the experiment with just two examples.

Experiment 3 involved one other modification of Experiment 2. At the end of Day 5 and before the beginning of the test for memory for the study examples, we inserted a final test in which all rules were tested in both directions with only unique examples. For participants who had been tested with unique examples, this was no change. However, we expected performance would be disrupted for participants who had only been tested with repeated examples. This would be evidence that they had adopted bases for responding differently than participants in the other groups.

Method

Participants. There were 45 participants recruited from the Carnegie Mellon University undergraduate population and from a local high school (ages older than 16) for inclusion in the experiment. There were 13 university students and 2 high school students assigned to each of the three conditions.

Procedure. The procedures and materials were identical to those in Experiment 2 with the following exceptions:

1. There were three conditions—all repeated, all unique, and mixed.

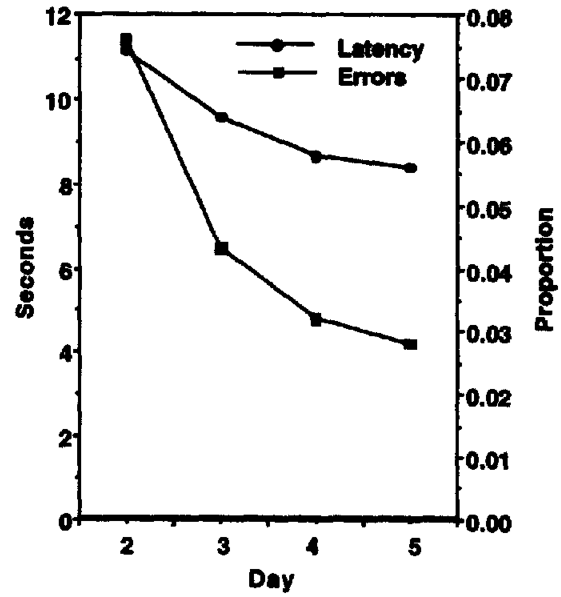

Figure 9. Decrease in latency and error rate in Experiment 3 as a function of day. Data are averaged over the two rules that are reversed that day and over the two directions.

2. Repeated examples were used on half of the trials (rather than on a quarter of the trials) for a rule—one example in the mixed condition and two examples in the repeated condition.
3. At the end of Day 5 there was a final phase involving 16 blocks in which each rule was tested with only unique examples in both directions.

Results

Again, to assess the relative contributions of rules versus retrieval, we performed ANOVAs on the latency and error data of Experiment 3 for Days 2–5, looking at those items that were tested in the reverse direction for the first time that day. The variables in the ANOVAs were condition, day, direction, and stimulus type (repeated or unique). For the all-repeated and all-unique conditions, the variable of stimulus type was a dummy variable.[7] Figure 9 shows the effect of day for latency and error rate. The effect of day was significant for latency, $F(3, 126) = 27.32$, $MSE = 23.94$, $p < .001$; and for error rate, $F(3, 126) = 10.27$, $MSE = 0.034$, $p < .001$. The effect of direction was significant: $F(1, 42) = 6.61$, $MSE = 5.89$, $p < .05$, for latency; $F(1, 42) = 2.72$, $MSE = 0.005$, $p < .1$, for error rate. Figure 10 shows the difference in direction as a function of day. The overall interactions between day and direction did not reach significance for either measure: $F(3, 126) = 1.61$, $MSE = 3.31$, for latency; and $F(3, 126) = 0.93$, $MSE = 0.004$, for error; however, the increasing linear trends were significant for

[6] However, it is possible retrieval would also show a directional effect (but this was not found by Anderson & Fincham, 1994).

[7] That is, to achieve a factorial design we arbitrarily designated half of the items in these conditions as unique and the other half as repeated.

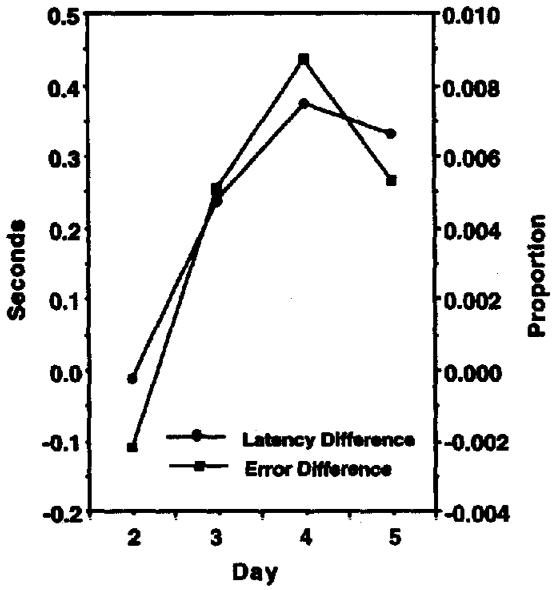

Figure 10. Directional difference, reverse − practiced, in Experiment 3 as a function of day. Data come from the two rules that are reversed that day.

both measures: $t(126) = 2.24$, $p < .05$, for latency; and $t(126) = 1.69$, $p < .05$, for error rate. In summary, the results from Experiment 3, reflected in Figures 9 and 10, approximately replicated the results of Experiments 1 and 2, reflected in Figures 2 and 3 and 7 and 8. In particular, the evidence for increased directional asymmetry with days and thus for production rules was replicated.

The effects involving stimulus type (repeated or unique) are relevant to assessing the contribution of retrieval of examples. There were highly significant interactions between condition and stimulus type: $F(2, 126) = 30.66$, $MSE = 11.37$, $p < .001$, for latency; and $F(2, 126) = 9.75$, $MSE = 0.023$, $p < .001$, for error rate. For errors there was also a significant interaction of these two with direction, $F(2, 126) = 12.05$, $MSE = 0.002$, $p < .001$, but this was not significant for latency, $F(2, 126) = 0.84$, $MSE = 2.62$. Table 5 displays these effects. Participants were generally faster and more accurate for repeated items and tended to show a smaller directional effect. There was some tendency for this to be enhanced in the mixed condition. For repeated items, participants were much faster in the mixed condition than in the pure repeated condition. For the unique items participants showed many more errors in the mixed condition than in the pure unique condition. This might suggest some shift of resources to the repeated items in the mixed condition. Participants might have adopted somewhat of a tendency to respond with the one repeated example in the mixed condition without checking if it was appropriate for the current problem. The elimination of the test for appropriateness would have yielded a speed advantage for the repeated example because they would have retrieved the answer for the repeated example without testing if it was appropriate. However, this strategy would result in errors for the unique example because participants would treat the unique items as if they were the repeated examples and give the wrong answer. In fact, 3.7% of the day errors and 4.3% of the hour errors to unique items for a rule in the mixed conditions were intrusions of the terms from the repeated item for that rule. This compares with 2.6% and 0.9% intrusions in the repeated conditions of answers from one of the repeated problems for a rule to the other problem for that rule. The difference in intrusions of the day, although in the expected direction, was not significant, $t(42) = 1.08$, but the difference in intrusions of the hour was significant, $t(42) = 3.56$, $p < .001$. In general, the effects were consistent with the view that participants tended to process repeated items by retrieval. Experiment 3 obtained a diminished directional effect with repeated items (which was highly significant for errors), unlike Experiment 2. In that experiment an item was repeated on one quarter of the trials for each rule, whereas in Experiment 3 it was repeated on one half of the trials.

To check the conclusions of this analysis, we also did a complete analysis of the training performance on the last day. The variables in the ANOVA were condition, day on which the rule was introduced in reverse direction (Days 2–5), direction (practiced or reverse), and stimulus type (repeated or unique). There were significant effects of direction: $F(1, 42) = 17.22$, $MSE = 4.50$, $p < .001$, for latency; $F(1, 42) = 6.29$, $MSE = 0.004$, $p < .001$, for error. Direction did not enter into significant interactions with the day that the rule was introduced: $F(2, 126) = 1.15$, $MSE = 9.07$, for latency; $F(2, 126) = 2.54$, $MSE = 0.005$, for error. However, the effects were in the expected direction—the mean directional difference was 0.32 s and 0.0% errors for items reversed on Days 2 and 3, whereas it was 0.34 s and 1.3% errors for items reversed on Days 4 and 5. The interaction between condition and stimulus type was highly significant: $F(2, 126) = 38.85$, $MSE = 9.18$, $p < .001$, for latency; and $F(2, 126) = 11.46$, $MSE = 0.009$, $p < .001$, for error rate. The interactions of these variables with direction were not significant: $F(2, 126) = 0.91$, $MSE = 6.20$, for latency; and $F(2, 126) = 1.49$, $MSE = 0.004$, for error rate; however, the effects were in the right direction (the directional differences were 0.31 s and 0.5% errors for repeated items and 0.37 s and 0.9% errors for unique items. One

Table 5
Latency (in Seconds) and Error Rates (in Percentages) in Experiment 3 as a Function of Condition, Stimulus Type, and Direction

Problem and direction	Mixed		Pure repeated		Pure unique	
	Latency	Error	Latency	Error	Latency	Error
Repeated						
Practiced	7.61	3.8	9.19	3.7	—	—
Reverse	7.73	4.1	9.32	3.1	—	—
Unique						
Practiced	9.70	7.7	—	—	10.11	3.3
Reverse	10.03	10.3	—	—	10.46	3.7

Note. Data are only for items reversed on that day. Dashes indicate that these problem types were not used for participants in these conditions.

might expect the directional effect to be somewhat washed out because of averaging over materials that had been reversed on earlier days.

We did analyses of the final transfer phase on Day 5 in which all participants solved only unique items. With respect to latency, there was only a marginally significant effect of condition, $F(2, 42) = 2.50$, $MSE = 59.58$, $p < .10$, with participants taking longer in the pure repeated condition (11.19 s) than in the pure unique condition (9.58 s) or in the mixed condition (9.00 s). The corresponding error rate effect was not significant, $F(2, 42) = 0.90$, $MSE = 0.151$, but participants on average made many more errors in the pure repeated condition (10.1%) than in the pure unique (4.0%) or in the mixed conditions (4.5%). These data do suggest that participants in the pure repeated condition were not developing as general a procedural skill for applying the rules. Planned comparisons were performed comparing the pure repeated condition with the average of the pure unique and mixed conditions. The comparison was significant in the case of latency, $t(42) = 2.21$, but not in the case of errors, $t(42) = 1.36$. The failure to get a significant effect reflects the high variability in the final error rates. The directional effect was significant for errors, $F(1, 42) = 10.78$, $MSE = 0.0034$, $p < .01$—7.2% versus 5.2%—but not for latency, $F(1, 42) = 0.55$, $MSE = 1.42$—9.95 s versus 9.85 s. There were no other significant effects.

Finally, we looked at ability to recall the original examples at the end of the fifth day. The overall error rate in recall was 36.7%, which was very similar to the result of 37.2% from Experiment 2. It did not differ significantly as a function of group, $F(2, 42) = 0.65$, $MSE = 1.04$, although participants in the pure repeated group did do worse (44.8%) than participants in the mixed group (34.8%) or the pure unique group (30.6%). Participants on average are failing to recall more than 35% of the original examples while they are making only 3.3% errors in rule use during training on the last day. Clearly, they are often applying their knowledge without reference to the original examples from which it was derived.

Discussion

This experiment did succeed in its intention of increasing the effect of the repetition manipulation. It also largely replicated the findings of the previous experiments. The results of this experiment support all four of the methods for responding that were reviewed at the end of the first experiment: (a) The lack of directional asymmetries early in the experiment is consistent with participants responding by analogy to examples and by declarative rules that lack any inherent direction, (b) growth of asymmetry and the forgetting of the original examples is consistent with development of asymmetric production rules, and (c) reduced directional effect for repeated examples and the difficulty that participants in the pure repeated condition had in the final transfer test is consistent with the development of response by retrieval of specific examples.

The results associated with repetition are quite similar to results reported by Rabinowitz and Goldberg (1995). They were looking at the alpha–arithmetic task (Logan & Klapp, 1991). They found reduced directional asymmetry when a few items were used in comparison with when many were used. In their case, participants were asked to transfer from answering *F* in response to $C + 3 = ?$ to answering *C* in response to $F - 3 = ?$. They also found that participants who practiced with a small set of items performed more poorly when transferred to a final set of totally new items.

General Discussion

It is worth reviewing the general course of skill acquisition to which we think these experiments point. First, participants memorize specific examples without any idea that the examples represent a rule. This is a somewhat peculiar requirement of our paradigm that is motivated to have participants start with pure example-based processing. However, other researchers (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Novick & Holyoak, 1991; Pirolli, 1985; Reed, 1987, 1989; Ross, 1984, 1987, 1989; Ross & Kennedy, 1990) have found that participants make extensive reference to examples even when they are initially taught the rules and principles.

Our assumption is that an example is encoded as a declarative structure. When participants are tested on their first problems, they have two possible ways to respond. If the example matches the problem they were trained on, they can simply retrieve the answer. However, if it does not match, they must analogically extend the example. We assume that this is a relatively difficult process. This accounts for the advantage when the study example was used in training in Experiment 1.

Each time a problem occurs a participant has an opportunity to codify abstractly the relationship between the first and second day. This initial coding is declarative as well. Our assumption is that a declarative encoding like *The second day is 2 days later and 1 hour earlier than the first* is not inherently directional. This accounts for the fact that participants were reporting a majority of rule use even on the first day of Experiment 1 but were showing no directional effects.

However, with somewhere around 50 applications it appears that a directional asymmetry does occur in the rule use. We take this as evidence for the emergence of production rules such as the following:

IF the question is about skydiving,
  and the second day is D2 and the second hour is H2,
  and D1 is two days before D2,
  and H1 is one hour after H2,
THEN the first day is D1 and the first hour is H1.

These production rules allow more rapid responses but only work in one direction, producing the directional asymmetry.

This phase of declarative and procedural rule use corresponds to Logan's (1988) algorithmic stage. We show that this rule use speeds up without any repetition of examples. However, continued practice allows the participants to experience more and more examples that we assume are also stored in declarative memory. This offers the participant another way of responding that is by retrieval of the example. This is even more direct and faster than production rule use. This corresponds to Logan's example stage. We assume it is responsible for the advantage of the repeated examples in Experiments 2 and 3 and the diminished directional effect for repeated examples in Experiment 3.

A final point worth stressing is that these four stages are not strictly sequenced. At any point in time a participant's responses may reflect a mixture of these methods with the proportions changing over time.

We view that the principal message of this article is that performance in a skilled task can reflect a complex mixture of processes. It involves using examples in two ways (analogy and simple retrieval) and using two types of rules (abstract declarative and procedural). Thus, we judge as implausible any theory that attributes skill acquisition to a single learning mechanism. Although we have made these investigations and conclusions in the context of a relatively simple and artificial task, we think the same mix of learning processes are involved in more complex problem-solving tasks. Indeed, the task used in these experiments was originally motivated to study more systematically the phenomena we observed in LISP programming (Anderson, Conrad, & Corbett, 1989) where we have found similar effects of directional asymmetry and problem repetition.

References

Anderson, J. R. (1993). *Rules of the mind.* Mahwah, 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., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1322–1340.

Blessing, S., & Anderson, J. R. (1996). How people learn to skip steps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 576–598.

Carlson, R. A., & Sohn, M. H. (1996, November). *Rule-application skills have internal structure.* Paper presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles, CA.

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.

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., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 179–195.

McKendree, J. E., & Anderson, J. R. (1987). Frequency and practice effects on the composition of knowledge in LISP evaluation. In J. M. Carroll (Ed.), *Cognitive aspects of human–computer interaction* (pp. 236–259). Cambridge, MA: MIT Press.

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

Pennington, N., Nicolich, R., & Rahm, J. (1995). Transfer of training between cognitive subskills: Is knowledge really use specific? *Cognitive Psychology, 28,* 175–224.

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

Rabinowitz, M., & Goldberg, N. (1995). Evaluating the structure-process hypothesis. In F. Weinert & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 225–242). Hillsdale, NJ: Erlbaum.

Reder, L. M., & Ritter, F. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 435–451.

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.

Rothkopf, E. Z., & Dashen, M. L. (1995). Particularization: Inductive speeding of rule-governed decisions by narrow application experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 469–482.

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

Received December 22, 1995
Revision received August 28, 1996
Accepted August 28, 1996 ■