

Comparing Multiple Paths to Mastery: What is Learned?

Timothy J. Nokes^a, Stellan Ohlsson^b

^a*Beckman Institute, University of Illinois at Urbana-Champaign*

^b*University of Illinois at Chicago*

Received 10 October 2003; received in revised form 31 January 2005; accepted 7 February 2005

Abstract

Contemporary theories of learning postulate one or at most a small number of different learning mechanisms. However, people are capable of mastering a given task through qualitatively different learning paths such as learning by instruction and learning by doing. We hypothesize that the knowledge acquired through such alternative paths differs with respect to the level of abstraction and the balance between declarative and procedural knowledge. In a laboratory experiment we investigated what was learned about patterned letter sequences via either direct instruction in the relevant patterns or practice in solving letter-sequence extrapolation problems. Results showed that both types of learning led to mastery of the target task as measured by accuracy performance. However, behavioral differences emerged in how participants applied their knowledge. Participants given instruction showed more variability in the types of strategies they used to articulate their knowledge as well as longer solution times for generating the action implications of that knowledge as compared to the participants given practice. Results are discussed regarding the implications for transfer, generalization, and procedural application. Learning theories that claim generality should be tested against cross-scenario phenomena, not just parametric variations of a single learning scenario.

Keywords: Psychology; Instruction; Problem solving; Representation; Skill acquisition and learning; Human experimentation

1. Introduction

Correct, effective, or successful behavior vis-à-vis a task can be generated in different ways. For example, walking from Point A to Point B in a city can be accomplished on the basis of a memorized route, a sequence of turns, or a mental map—a representation of the relevant spatial relations. We propose that qualitatively different learning scenarios vary with respect to what is learned and how that knowledge is represented, even when the resulting performance

level, as measured by accuracy, is the same. Furthermore, the fact that two learning scenarios produce the same performance level does not imply that behavior in those two scenarios is equal along other quantitative dimensions. Different knowledge structures will cause differences in the underlying cognitive processes, which in turn have consequences for processing time, strategy use, and transfer.

How do the knowledge structures created in learning scenario X differ from those created in scenario Y? What are the behavioral consequences of the differences between the knowledge structures? A general theory of human learning should be able to predict, for a given learning scenario, which type of knowledge structure will be created and, for a pair of such scenarios, how the resulting knowledge structures will differ and what differences they will cause in behavior.

Most contemporary work on learning is not directed toward answering such questions. A typical cognitive learning theory describes a single learning mechanism (e.g., analogy). To provide empirical support for its psychological reality, the researcher studies a learning scenario for which that mechanism has a high degree of face validity (e.g., to verify that people learn via analogy, give subjects analogs to a target problem and verify that performance improves). Further support for the hypothesized learning mechanism is accumulated by showing that it can explain the behavioral effects of various parametric variations of that scenario (e.g., number of prior analogs). Although this research strategy has been successful in deepening our understanding about the processes and properties of the particular learning mechanism under investigation, the question of how people learn in situations in which that mechanism is not applicable or plausible is left unanswered (e.g., how do people learn when they have no useful analog in memory?).

A useful complementary research strategy is to compare qualitatively different learning scenarios with respect to what is learned. We are interested in the differences between the knowledge structures acquired by learners who perform at the same level but who arrived at that performance level along different learning paths (for a similar approach, see Klahr & Nigam, 2004). Empirically documented differences between the knowledge structures acquired in different but effective scenarios constitute phenomena that a general learning theory ought to be able to explain.

There are two broad classes of learning scenarios that, taken together, account for a significant proportion of human learning in both formal and informal contexts. One class contains situations in which the learner is presented with oral or written discourse that explicitly expresses the target knowledge that he or she is supposed to acquire. Lectures and textbooks exemplify this type of instruction. The learning in this type of scenario is sometimes described as *learning by being told* (Carroll, 1968), but we will refer to this type of scenario as *direct instruction*, or just *instruction* when the context presents ambiguity. Instruction scenarios vary in the length and conceptual depth of the material presented. They also vary in whether or not worked out examples or illustrations are provided. The defining characteristic of direct instruction is that the instructing agent—person or machine—communicates the target knowledge in explicit form, usually via discourse.

A second important class of learning scenarios contains situations in which the learner engages in some activity that approximates the desired target performance. The learning in this type of scenario is usually called *learning by doing* (e.g., Anzai & Simon, 1979). Scenarios of

this type differ with respect to the variability of the practice problems. The terms *drill*, *practice*, and *analogical learning* refer to scenarios with greater and greater variability between the successive practice problems. The common feature of this class of learning scenarios is that the target knowledge is not communicated explicitly. Instead, the learning activities—the practice tasks—are designed in such a way that, by attempting to perform them, the learner is prompted to construct the target knowledge, or so the instructor hopes.

If two learners master one and the same task in these two ways, it is plausible that they will acquire different knowledge structures. Cognitive science provides a vocabulary of useful concepts for discussing knowledge structures (Markman, 1999). We focus on two: declarative versus procedural knowledge and abstract versus specific knowledge. The particular specification of these dimensions for a given knowledge structure has implications for transfer, generalization, and application of that knowledge.

1.1. Declarative versus procedural knowledge

Declarative knowledge is knowledge about the way the world is; it is descriptive in character and it can be evaluated with respect to veridicality (Anderson, 1982; Knowlton & Squire, 1993; Ohlsson, 1994; Squire, Knowlton, & Musen, 1993; Winograd, 1975). Common facts such as *the winter is cold in Minnesota* are prototypical examples. Declarative knowledge is task independent, for example, it is not encoded in memory in the context of particular goals or actions. When applied to a task, it has to be interpreted with respect to its consequences for action. The process of deriving the action consequences of declarative knowledge is sometimes called *proceduralization* or *knowledge compilation* (Anderson, 1983; Neves & Anderson, 1981; Ohlsson, 1996).

Procedural knowledge is knowledge about how to perform tasks and, more generally, about how to achieve particular types of results in certain types of situations (Anderson, 1982; Knowlton & Squire, 1993; Ohlsson, 1994; Squire et al., 1993; Winograd, 1975). Everyday skills such as driving and cooking are prototypical examples. Procedural knowledge is task specific and indexed in memory under the relevant goals, so its application is quick and efficient. It can be evaluated with respect to appropriateness and effectiveness.

A person's knowledge about a task can be either declarative, procedural, or some mixture of the two (Maxwell, Masters, & Eves, 2002). The two types of knowledge trade off applicability and efficiency in opposite ways (Anderson, 1978; Winograd, 1975): Declarative knowledge is widely applicable, but to derive its consequences for action is a slow and complex process. Procedural knowledge applies only in a narrow task context but the application is fast and efficient. This distinction is also consistent with the empirical results from the category learning literature showing that rule-based categorization applies generally but is more error prone and slower than specific exemplar-based categorization (Allen & Brooks, 1991; Palmeri & Nosofsky, 1995). For any learning scenario we can ask whether it generates declarative or procedural knowledge, or some mixture of the two.

How do the knowledge structures created by direct instruction and practice differ on the dimension of declarative versus procedural knowledge? The first expectation is that learning by being told generates declarative knowledge. Without any opportunity to practice, a learner is unlikely to acquire procedural knowledge. Hence, we would expect the application of what he

or she learned to a target task to be slow because he or she has to proceduralize the knowledge while performing the task.

The effect of worked-out example should be to provide the learner with additional *declarative* knowledge *about* the types of procedures or inferences needed to solve a task or problem. However, that knowledge still needs to be proceduralized to generate behavior. For example, studying a worked-out example of a geometry theorem provides a declarative representation of an inference pattern, but to use that knowledge as proof heuristics it needs to be proceduralized (Anderson, Greeno, Kline, & Neves, 1981).

A second expectation is that learning by practicing generates procedural knowledge (Anderson, 1983; Ohlsson, 1996). The application of this type of knowledge should be faster, because proceduralization has already been carried out during prior learning. However, the attempt to solve a task is likely to also generate some declarative knowledge.

1.2. *Abstract versus specific*

Abstraction is of central concern in studies of learning, because it might underpin people's ability to transfer what they learn in one context to other contexts (Gick & Holyoak, 1987; Goldstone & Sakamoto, 2003; Salomon & Perkins, 1989). Abstraction can be defined in various ways. The traditional definition is that the abstraction level of a concept is a function of the size of its reference set: "Mammal" is more abstract than "cow," because there are more mammals than cows. According to another definition, knowledge is abstract if it is decontextualized. A mathematical formula is abstract because it is separated from any particular context in which it might apply. Neither of these definitions capture the particular dimensions of abstraction that turn out to be important in this study. For the purposes of this article, we regard a mental representation as *abstract* to the extent that it encodes relational information but leaves out information about the identity and the specific attributes of objects and events (Ohlsson, 1993; Ohlsson & Lehtinen, 1997). The opposite of abstract is then *specific*.

We distinguish between object abstraction and relation abstraction (Ohlsson, 1993; Nokes & Ohlsson, 2003). We say a knowledge structure exhibits *object abstraction* when it leaves the attributes and features of objects and events unspecified, but encodes relations in a specific way. When the relations themselves are encoded with minimal semantic features, we say that the knowledge structure exhibits *relation abstraction*. The statement *the red block is on top of the green block* is a specific statement (in a context where the noun phrases select a unique object); *some block is on top of another block* exhibits object abstraction; and *some block is in contact with another block* exhibits relation abstraction because *being in contact with* is less specific than *being on top of*. The question of interest is whether learning scenarios differ in whether the knowledge structures they generate are specific, or exhibit either object or relation abstraction.

The distinction between levels of abstraction is orthogonal to the declarative–procedural distinction. Mathematical equations are examples of abstract, declarative knowledge, whereas statements of fact are examples of specific declarative knowledge. A weak method such as means–ends analysis (Newell & Simon, 1972) is an example of abstract procedural knowledge, whereas a cooking recipe is an example of specific procedural knowledge. Either type of knowledge can be expressed at different levels of abstraction.

In a previous study, we found evidence that our participants encoded declarative task knowledge at the object but not at the relation level of abstraction (Nokes & Ohlsson, 2003). They abstracted over particular objects, but we found no evidence that they spontaneously encoded relations at any level of abstraction higher than the one in which the task materials were presented. That result leads us to hypothesize that learning scenarios that produce more declarative than procedural knowledge—that is, direct instruction scenarios—will lead to knowledge structures that exhibit object abstraction. The implication is that those structures should transfer easily to a task that differs in specific details but shares the same relational structure, but less easily to a task that also differs in relational structure.

For practice scenarios, our expectations are different. In prior work on learning by doing (Ohlsson, 1996; Ohlsson, Ernst, & Rees, 1992; Ohlsson & Rees, 1991), we proposed a *specialization principle*: Cognitive skills are learned during practice by gradually constraining (specializing) an initial weak method until each component (e.g., production rule) becomes active only in those circumstances in which it leads to the correct action. The specialization principle predicts that practice should produce procedural knowledge that is specific rather than abstract. This expectation is moderated by the variability of the practice sequence. The more variable the sequence, the stronger the prompt toward an abstract representation (Salomon & Perkins, 1989).

1.3. Summary

We propose that different learning scenarios will lead to different knowledge representations, even when they are equally sufficient to produce mastery of a target task. We focus on differences with respect to the declarative versus procedural character of the knowledge and with respect to the level of abstraction. Prior work suggests that direct instruction should generate declarative knowledge that exhibits object abstraction, whereas practice should generate primarily procedural knowledge and that its abstraction level is a function of the variability of the practice sequence.

2. Task environment

The ideal target task shares key properties of ecological cognitive skills but enables experimental control and data capture. We used a version of the *letter-sequence extrapolation task* studied by Greeno and Simon (1974), Klahr and Wallace (1970), Restle (1970), Restle and Brown (1970), Simon (1972), Simon and Kotovsky (1963), and others. The learner is given a sequence of letters that exemplify a pattern, and he or she is asked to continue the sequence in such a way that the continuation also fits that pattern. As a simple example, to extrapolate the sequence ABMCDM to six places the participant would produce EFMGHM.

To solve such a problem, the participant must first identify the pattern in the given sequence and then use that pattern to generate the continuation of the sequence. The first part, *pattern detection*, consists of studying the given part of the sequence to identify the relations between the letters. The second part, *sequence extrapolation*, requires a series of inferences based on the identified pattern, one for each position extrapolated.

Based on this analysis, the solution process can be divided into two additive parts. The *deliberation time* is the time until a participant extrapolates their first letter, and the *extrapolation time* is the time it takes him or her to complete the problem. On the plausible assumption that the participants try to figure out the pattern before they attempt to extrapolate it, the deliberation time is an estimate of how long it takes a person to identify the pattern. The extrapolation time is an estimate of how long it took him or her to carry out the extrapolation inferences.

The letter-sequence extrapolation task has several properties that make it a suitable model of complex cognitive skills. *Prior knowledge*: Like ecological tasks, the process of mastering a letter-sequence extrapolation problem draws on the participant's prior knowledge, namely, the alphabet. *Conceptual content*: Understanding the pattern is a prerequisite for successful performance of the task. *Types of knowledge*: Knowledge of a pattern is an example of declarative knowledge, whereas the skill of extrapolation is an example of procedural knowledge. Both are equally important for successful performance. *Abstraction*: The pattern in a sequence can be encoded in terms of specific letters, in terms of relations between positions in the sequence (object abstraction), and in terms of which positions are related (relation abstraction; Nokes & Ohlsson, 2003; Ohlsson, 1993). *Generativity*: A sequence extrapolation problem cannot be solved by memorizing and recalling the given sequence, but requires that the participant generate a novel sequence of coordinated responses. These are interesting features that sequence extrapolation problems share with more complex tasks.

Most important for our current purposes, the correct extrapolation of a sequence is facilitated if the person has learned the underlying pattern ahead of time. Prior declarative knowledge of the pattern ought to facilitate pattern detection. It should be easier to recognize a familiar pattern in a letter sequence than it is to identify an unfamiliar pattern. Similarly, practice in carrying out extrapolation inferences should generate procedural knowledge that makes future inferences easier. Both patterns and extrapolation skills should transfer to related problems if they are encoded at some level of abstraction. Sequence extrapolation tasks are thus well suited as instruments with which to study the learning outcomes produced by a variety of learning scenarios.

We report a laboratory experiment that recorded participants' performance in two qualitatively different training scenarios: direct instruction and practice. These two scenarios were implemented in two parametric variants each: Direct instruction was either short or long, and the practice sequence was of either low or high variability. We report accuracy and solution time results as well as participants' patterns of extrapolation. In a separate section, we provide an in-depth analysis of the structure of the solution times to arrive at a processing account of our findings.

3. Empirical study

The purpose of the study reported in this article was to compare what is learned when a letter-sequence extrapolation task is mastered via two contrasting learning scenarios. Specifically, we compare learning by being told and learning by doing with respect to their effects

on subsequent problem solving. We provided either instruction or practice before the participants encountered the target and transfer problems. Performance on the target and transfer problems thus served as tools for assessing the effects of the prior training. We refer to the target and transfer problems collectively as the *test problems*.

If the knowledge structures acquired during the two different training procedures are the same, the participants' behavior on the test problems should be similar. If the knowledge structures differ with respect to knowledge type (declarative vs. procedural) or level of abstraction (specific, object, or relation), then we expect to see differences in behavior.

We used variants of each learning scenario to be able to do both between- and within-scenario comparisons. Patterns in the data that recur across parametric variants of a learning scenario, but which are not in data from another learning scenario, have a high probability of being due to the basic features of that scenario. The within-scenario comparisons provide a handle on the robustness of the effects of a scenario. Robust effects should survive minor variations of a scenario. Also, the effects of the parametric variations provide fine-grained constraints on potential processing accounts.

3.1. Predictions

We expect that both training groups will show improved performance over a no-training control group on the test problems as measured by mean accuracy. However, we expect differences in how they apply their knowledge as measured by their solution times and pattern completion strategies.

Declarative knowledge requires proceduralization and hence should be slower in application than procedural knowledge. We therefore predict that direct instruction will produce longer solution times than practice, even for participants who perform at comparable levels of accuracy. Furthermore, because the application of procedural knowledge primarily takes place during the *extrapolation* component of the solution process, we predict longer times only for this portion of the solution times.

Declarative knowledge is also open to multiple different uses or interpretations. We therefore predict that participants given direct instruction will exhibit more variability in their patterns of completion (i.e., the order in which participants extrapolate the solution positions) as compared to the practice participants. Consistent with this hypothesis, we also expect that the instruction participants will be more likely to change their patterns of completion between target and transfer problems than the participants in the practice group who had more opportunities to proceduralize their knowledge.

In addition, the standard trade-off argument for the existence of two types of knowledge (Anderson, 1976, 1983; Winograd, 1975) implies that declarative knowledge, by virtue of not being indexed under particular goals or tasks, has a higher generality and transferability than task-specific procedural knowledge. Therefore, we expect a relative advantage for declarative over procedural knowledge, and hence for direct instruction over practice, on a transfer task. Because variable practice is widely believed to generate more abstract knowledge than uniform drill, the relative advantage for instruction over practice should be greater when compared to low-variability practice and lesser with respect to high-variability practice.

3.2. Methods

3.2.1. Participants

One hundred forty-nine undergraduate students from the University of Illinois at Chicago participated in return for course credit.

3.2.2. Materials

The target tasks were two letter-sequence extrapolation problems. Problem 1 had a periodicity of 6 items and Problem 2 a periodicity of 7 items (see Table 1). To enable the participants to detect the embedded pattern, the given segment was 12 items long for Problem 1 and 14 items long for Problem 2. That is, the given segments covered two complete periods of the patterns. These problems were created specifically for this experiment but are similar in character to the problems used by Kotovsky and Simon (1973).

Each target problem was associated with a transfer problem. The patterns embedded in the transfer problems were related, but not identical, to the patterns in the target problems. For Problem 1, the corresponding transfer problem was generated by quantitatively “stretching” particular relations of the pattern. For example, “forward 1 in the alphabet” was “stretched” to “forward 2 steps” (see Fig. 1 for a detailed example). The second transfer problem was generated in a similar way from target Problem 2. This method of generating transfer problems preserves the qualitative structure of the target pattern (i.e., which positions are related and the direction of each relation) but changes the quantitative aspect of the relations.

In addition, there were a total of three practice problems for each target problem. The three training problems followed the same pattern as the target problem (see Table 2). The training problems were constructed in such a way that they did not overlap (i.e., did not share surface features) with each other or the target problem. The low-variability practice group was trained on the first of the three training problems, and the high-variability practice group was trained on all three.

In addition, there were two sequence extrapolation tutorial booklets, one for short instruction (12 p.) and one for long instruction (14 p.). Both tutorials consisted of general instructions on how to find pattern sequences as well as detailed descriptions of the component relations of patterns (e.g., forward, backward, repeat, and identity relations). Complete versions of the instruction materials are available at <http://cognitivesciencesociety.org/supplements/>.

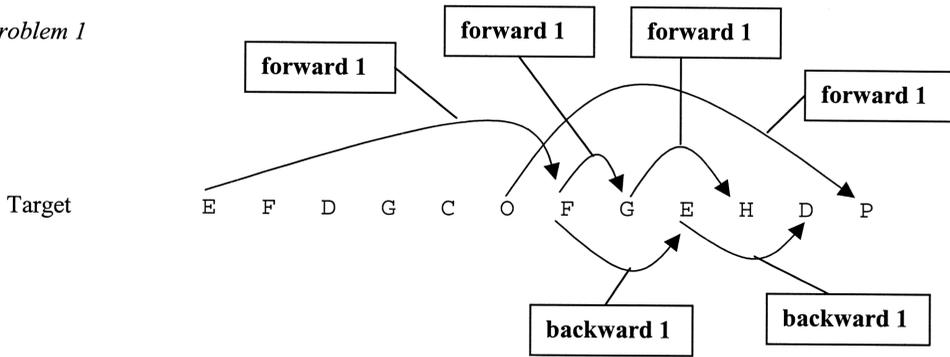
Table 1
Two sequence extrapolation problems and their associated transfer problems

Problem Type	Given Letter/Number Sequence	Correct 8-Step Extrapolation
Problem 1		
Target	EFDGCOFGEHDP	GHFIEQHI
Transfer	EGDICOGIFKEP	IKHMGQKM
Problem 2		
Target	ACZDBYYDFXGEWW	GIVJHUUJ
Transfer	AEGZCXXGKVMITT	MQRSOPPS

Problem Type

Identified Relations

Problem 1



Transfer

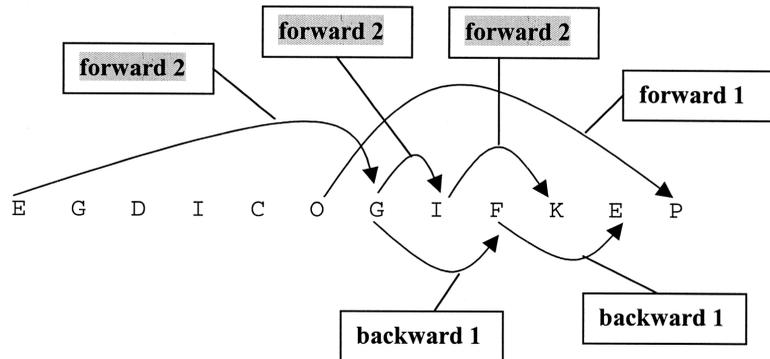


Fig. 1. Example target and transfer problem with the relations identified.

The long instruction tutorial had two additional pages of general instruction describing how to extrapolate patterns. Two example problems were worked through in detail, extrapolating one letter at a time. Participants were told that to extrapolate a pattern, they must first find the relations that make up the pattern and then use those relations to continue it. For example, to extrapolate the first letter of the pattern below participants were instructed as follows:

A B M C D M ... → E F M

“So first we need to decide on the 1st letter of the third period. The 1st letter in the second period, C, is one forward from the 2nd letter of the first period. So the letter we are looking for should be the one forward from the 2nd letter of the second period, which is D. So the letter we are looking for is E.”

Table 2
Three practice problems for each problem type

Problem Type	Given Letter/Number Sequence	Correct 8-Step Extrapolation
Problem 1		
1	I J H K G S J K I L H T	K L J M I U L M
2	R S Q T P B S T R U Q C	T U S V R D U V
3	M N L O K W N O M P L X	O P N Q M Y P Q
Problem 2		
1	G I T J H S S J L R M K Q Q	M O P P N O O P
2	R T I U S H H U W G X V F F	X Z E A Y D D A
3	N P M Q O L L Q S K T R J J	T V I W U H H W

The rest of the extrapolation was described in a similar way. Participants were then given step-by-step extrapolation instructions for a more difficult problem. These instructions were intended to focus the learner on the procedure of extrapolating a pattern.

Both instruction groups also received a general tutorial test that consisted of four recall questions and one comprehension question for the short instruction group and two comprehension questions for the long instruction group. An example of a recall question was to write a brief description of the *repeat letter* relation. An example of a comprehension question was to describe what a period is and give an example of a periodic sequence. In addition, there were two diagrammatic illustrations of the underlying pattern relations for each of the target problems as well as two blank diagrammatic recall sheets (see Fig. 2 for an example of a diagrammatic pattern illustration). There were also two distracter tasks that consisted of three multiplication problems each.

The test and training problems were presented on a Macintosh computer with a 17-in. color monitor, standard keyboard, and mouse. Problem stimuli were presented in black 30-point font in the center of the screen (see Fig. 3) for screen layout. The computer materials were designed and presented using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). Direct instruction training materials were presented in booklet form.

Problem 1 and Problem 2 and associated training stimuli were counterbalanced across all conditions.

3.2.3. Design

The participants were randomly assigned to one of four groups: *low-variability practice* ($n = 30$), *high-variability practice* ($n = 31$), *short instruction* ($n = 28$), and *long instruction* ($n = 30$). In addition there was a *no-training control group* ($n = 30$) to be used as a comparison group of baseline performance.

3.2.4. Procedure

Participants were tested individually. The procedure for the training groups consisted of two cycles, each encompassing a training phase and a test phase. Participants in the no-training condition were just given the test phase but no training phase.

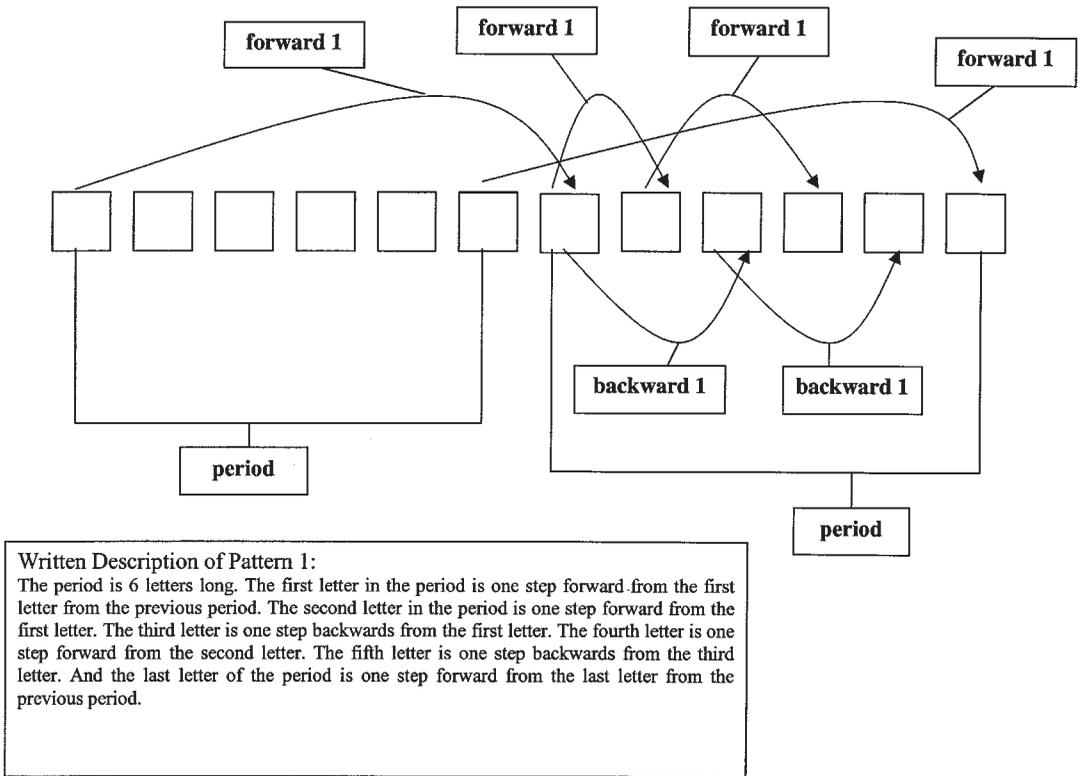


Fig. 2. Diagrammatic pattern illustration.

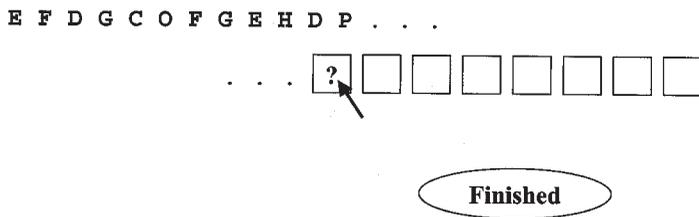


Fig. 3. Example problem as it was presented on the computer.

3.2.4.1. *Procedure for practice groups.* Participants were first given general instructions on how to solve sequence extrapolation problems. They were then given the first sequence extrapolation training problem. Participants were instructed to extrapolate each of the eight positions by clicking the mouse on any given position and typing in the answer (see Fig. 3 for an example). They were given 6 min to solve each problem. After participants finished solving a problem or 6 min time elapsed, they continued to the next problem by pressing the space bar. They were not given any feedback as to the accuracy of their solutions.

After participants solved all three training problems they were given the target problem instructions. Target problem instructions were the same as the training instructions except that they included the hint that if participants noticed a pattern on any of the prior problems it would help them solve the next problem. Participants were then given 6 min to solve the target problem. Finally, they were given the transfer problem and were instructed to solve it in the same manner as the target problem.

The second cycle proceeded in the same way. The entire procedure took 60 to 80 min.

3.2.4.2. Procedure for direct instruction groups. First participants were given the general tutorial text to read after which they were given the tutorial test. Participants were then given 3 min to memorize the first diagrammatic pattern illustration. Next, participants were presented with the diagrammatic blank recall sheet and instructed to recall and write down the relations of the pattern. Participants were then given the distracter task to prevent memory rehearsal strategies of the rules. Next, participants were presented with the general instructions for the test problems. They were then given 6 min to solve the target problem. Finally, they were given the transfer problem and were instructed to solve it in the same manner as the target problem. The second cycle proceeded in the same way. The entire procedure took 70 to 90 min.

3.2.4.3. Procedure for the no-training group. First participants were given general instructions. They were then given 6 min to solve the target problem. Next, they were given the transfer problem and were instructed to solve it in the same manner as the target problem. The second cycle proceeded in the same way. The entire procedure took 20 to 30 min.

3.3. Results

The two primary questions are whether participants acquired knowledge of the target pattern from the training procedures and whether that knowledge facilitated performance on the subsequent test problems (target and transfer). The question of the nature of the differences in what was learned is discussed in more detail in the processing account section.

Alpha was set to .05 for all main effects and interactions. Modified Bonferroni tests were conducted for all planned comparisons (Keppel, 1991). The greatest number of comparisons for any one statistical test was 10 so the corrected alpha level was set to .02 for all planned comparisons. Effect sizes (eta squared) were calculated for main effects, interactions, and main comparisons. Cohen (1988; see also Olejnik & Algina, 2000) suggested that effects be regarded as small when $\eta^2 < .06$, as medium when $.06 < \eta^2 < .14$, and as large when $\eta^2 > .14$.

3.3.1. Training performance

The purpose of the analysis in this section is to document that the participants acquired the target knowledge during training.

3.3.1.1. Training time. The average time spent on training was calculated for each training group. The low-variability practice group spent 1,117 sec (~19 min; $SD = 412$) on solving the practice problems, whereas the high-variability practice group spent 1,163 sec ($SD = 352$). In

contrast, the short instruction group spent 1,572 sec (~ 26 min; $SD = 197$), and the long instruction group 1,624 sec ($SD = 150$) on training (i.e., tutorial, test, memorization and recall of the abstract pattern). These data show that the instruction groups spent considerably more time than the practice groups with the training materials. If time on task is a major determinant of the learning outcomes in this scenario, we should therefore expect the instruction groups to perform better than the practice groups.

3.3.1.2. Practice. The first question is whether problem-solving performance increased across practice trials. If participants extracted knowledge of the pattern from the initial problem, it should facilitate performance on subsequent training problems. The low-variability group should show a positive linear increase across training trials, because they solved the exact same problem on each occasion. The high-variability group should show the same pattern of results if the knowledge gained from the initial problem was abstract and accessible for transfer across training problems.

The *practice training score* was determined as the number of correct extrapolations for each problem-solving task. Because participants were asked to extrapolate each problem to eight positions, their problem-solving scores varied from 0 to 8. An initial one-way analysis of variance (ANOVA) was conducted to investigate the effect of problem (Problem 1 vs. Problem 2) on problem-solving performance. The ANOVA revealed no effect of problem, $F(1, 60) = 2.83$, mean square error (MSE) = 3.41, *ns*, indicating that the type of pattern had no significant effect on problem-solving performance.

Fig. 4 shows the mean problem-solving scores for the low- and high-variability groups on the three training problems collapsed across problem. A 2 (group: low variability vs. high vari-

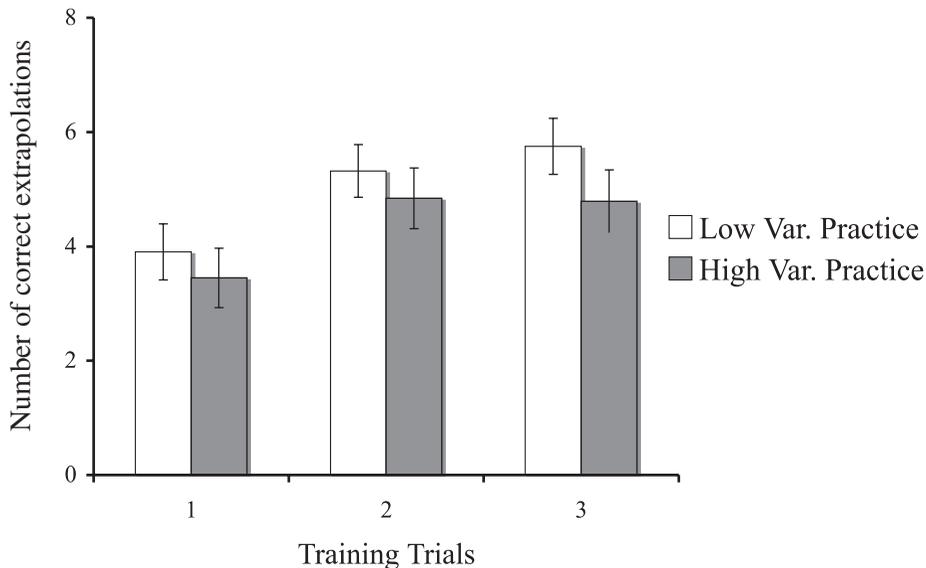


Fig. 4. Mean problem-solving scores for the low- and high-variability practice groups on the three training problems.

ability) \times 3 (trial: 1 vs. 2 vs. 3) mixed ANOVA was conducted to examine the effect of type of training across problem-solving trials. The ANOVA revealed a large main effect of trial, $F(2, 118) = 31.83$, $MSE = 1.45$, $p < .05$, $\eta^2 = .35$, indicating that performance increased significantly with training. There was no effect of group, $F(1, 59) = .87$, $MSE = 20.85$, *ns*, indicating that participants in the low-variability condition performed at the same level as participants in the high-variability condition. There was also no interaction of Group \times Trial, $F(1, 118) = .87$, $MSE = 1.45$, *ns*, indicating that performance did not significantly differ across training trials as a function of group.

To follow up the effect of trial a linear ANOVA was conducted on training Trials 1 to 3. The ANOVA revealed a significant positive linear trend across training trials, $F(1, 59) = 42.69$, $MSE = 1.82$, $p < .05$, $\eta^2 = .42$, indicating performance improvements across all three trials collapsed across training groups.

3.3.1.3. Direct instruction. Two training measures were taken: participant ratings of how well they understood the general tutorial and pattern recall. Participants were asked to rate how well they understood each page of the general tutorial. Ratings were made on a 1 to 5 Likert-type scale; from 1 (*I don't understand at all*) to 5 (*I understand completely*). Mean self-rating scores were 4.68 ($SD = .48$) for the short instruction group and 4.79 ($SD = .43$) for the long instruction group.

A one-way ANOVA comparing short instruction ratings to long instruction ratings was not significant, $F(1, 57) = .75$, $MSE = .20$, *ns*, indicating instruction groups were equally likely to report that they understood the general tutorial.

The second measure was the number of relations correctly recalled from the diagrammatic pattern descriptions. The *pattern recall score* was based on the number of relations embedded in a given pattern; hence, scores varied from 0 to 7 for Pattern 1 and 0 to 8 for Pattern 2. Mean recall scores for the short instruction group were 6.25 ($SD = 1.32$) on Pattern 1 and 4.64 ($SD = 2.63$) on Pattern 2. Mean recall scores for the long instruction group were 5.67 ($SD = 1.77$) on Pattern 1 and 5.13 ($SD = 2.58$) on Pattern 2.

A 2 (group: short instruction vs. long instruction) \times 2 (pattern: Pattern 1 vs. Pattern 2) mixed ANOVA revealed a large main effect for pattern, $F(1, 56) = 39.05$, $MSE = .004$, $p < .05$, $\eta^2 = .41$, indicating that participants recalled significantly more relations for Pattern 1 than for Pattern 2. There was no effect of group, $F(1, 56) = .03$, $MSE = .11$, *ns*, indicating that the two direct instruction groups did not differ in recall performance. However, there was a marginally significant interaction of Group \times Pattern, $F(1, 56) = 3.54$, $MSE = .004$, $p = .064$, $\eta^2 = .06$, indicating that group recall performance differed as a function of pattern. Inspection of the means show that the short instruction group recalled slightly more relations than the long instruction group on Pattern 1 (6.25 vs. 5.67), and the opposite trend was observed on Pattern 2 (4.64 vs. 5.13). We have no explanation for this unexpected finding. The effect was small, $\eta^2 = .03$ for Pattern 1 and $\eta^2 = .009$ for Pattern 2.

In sum, the two direct instruction groups scored high on the comprehension and recall tasks, indicating that they learned something about the relevant pattern. In addition, the two groups performed almost identically on both measures. Thus, both the instruction and the practice groups acquired knowledge of the relevant patterns during the training phase.

3.3.2. Test performance

The purpose of the analysis in this section is to investigate the effect of the knowledge acquired during training on the participants' performance on the test problems. We report three types of evidence: accuracy, solution time, and completion patterns. To assess test performance we used only those participants who received the top two thirds of the scores on the training measures (low-variability practice = 20, high-variability practice = 21, short instruction = 19, and long instruction = 20). We used this selection criterion because the participants differed in their ability to learn from the various training procedures, presumably due to differences in cognitive ability, differences in motivation, and perhaps other factors. For the purpose of this study we are concerned with the effects of training on subsequent problem solving when training is successful.

We employ this approach because we are most interested in testing whether the training knowledge generalizes to subsequent transfer tasks. The participants in the upper two thirds of learning performance mark those who are closest to acquiring mastery of the training knowledge and therefore are the best participants to test what that knowledge can be used for. The participants who did not perform well and hence did not acquire the knowledge provide no information about this issue and hence can only weaken the test by diluting the results.

The effect of restricting our attention to the upper two thirds of the participants (based on the training measures, not test performance) accentuates those effects that are due to differences between the types of training.¹ We used the two-thirds criterion as opposed to median splits because it leaves a larger number of participants in each condition to perform the subsequent statistical analyses.

3.3.2.1. Accuracy. The *problem-solving score* was determined by the number of correct extrapolations for each problem-solving task. Because participants were asked to extrapolate each problem to eight positions, their problem-solving scores varied from 0 to 8.

A one-way repeated measures ANOVA was conducted to investigate effect of problem (1 vs. 2) on problem-solving performance. The ANOVA revealed no effect of problem, $F(1, 109) = .32$, $MSE = 1.71$, *ns*, indicating that the type of pattern had no significant effects on problem-solving performance. Fig. 5 shows the mean problem-solving scores and standard errors for the practice, direct instruction, and no-training groups on target and transfer problems collapsed across problem type.

A 5 (training: low-variability practice vs. high-variability practice vs. short instruction vs. long instruction vs. no training) \times 2 (test problem: target vs. transfer) mixed ANOVA revealed a large main effect of training, $F(1, 105) = 11.34$, $MSE = 10.68$, $p < .05$, $\eta^2 = .30$, indicating that problem-solving performance differed across training groups. There was also a medium-sized effect of test problem, $F(1, 105) = 12.85$, $MSE = 1.40$, $p < .05$, $\eta^2 = .11$, showing that participants performed better on the target problem than on the transfer problem. There was no interaction of Training \times Test problem, $F(4, 105) = .90$, $MSE = 1.40$, *ns*, indicating that the difference in problem-solving performance between target and transfer problems did not vary as a function of training. Planned comparisons for the effect of training are summarized in Table 3.

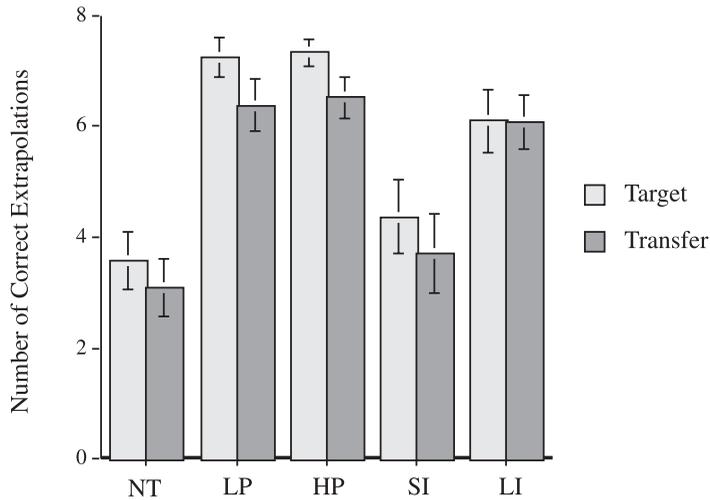


Fig. 5. Mean problem-solving accuracy scores and standard errors for the no-training (NT), low-variability practice (LP), high-variability practice (HP), short instruction (SI), and long instruction (LI) groups on the target and transfer problems.

In addition, we were interested in whether or not problem-solving performance differed across experiment cycles (Cycle 1 vs. Cycle 2). It is possible that participants became aware of the connection between the training and the test phase after finishing Cycle 1 and deliberately changed their approach to the training materials on Cycle 2, thus leading to performance differences across cycles as a function of training type. A 4 (group: low-variability practice vs. high-variability practice vs. short instruction vs. long instruction) × 2 (Cycle: 1 vs. Cycle 2) mixed ANOVA revealed no effect for cycle, $F(1, 76) = .914, MSE = 1.51, ns$. In addition, there was no interaction of Training × Cycle, indicating that performance across cycles did not change as a function of training, $F(3, 76) = 1.40, MSE = .151, ns$.

Table 3
Follow-up comparisons for the effect of training on test performance as measured by accuracy

Training	No Training	Practice Low	Practice High	Short Instruction	Long Instruction
No training	—	$F = 22.59^*, \eta^2 = .18$	$F = 29.76^*, \eta^2 = .22$	$F = 1.06$	$F = 16.94^*, \eta^2 = .16$
Practice low	—	—	$F = .33$	$F = 11.16^*, \eta^2 = .11$	$F = .34$
Practice high	—	—	—	$F = 15.59^*, \eta^2 = .13$	$F = 1.36$
Short instruction	—	—	—	—	$F = 7.65^*, \eta^2 = .07$
Long instruction	—	—	—	—	—

*Significant planned comparisons, $p < .02$.

In sum, the accuracy results show that participants in the two practice groups and the long instruction group performed better than the participants in the short instruction and no-training groups on both target and transfer problems. The long direct instruction and the practice groups did not significantly differ from each other, thus constituting alternative paths to the same high-level performance.

3.3.2.2. Time measures. We defined *solution time* as the total amount of time (in seconds) it took the participants to solve the problem. Following a long-standing tradition in cognitive psychology, we interpret a solution time as an estimate of the amount of cognitive processing the participant engaged in to produce his or her problem solution. The purpose of the analysis reported here is to determine whether those participants who had higher accuracy scores engaged in more processing. Fig. 6 shows the mean solution times and standard errors for the practice, direct instruction, and no-training groups on target and transfer problems. A detailed analysis and interpretation of the deliberation versus extrapolation time subcomponents is presented in the processing account section.

A 5 (training: low-variability practice vs. high-variability practice vs. short instruction vs. long instruction vs. no-training), \times 2 (test problem: target vs. transfer) mixed ANOVA revealed a large main effect of training, $F(4, 105) = 19.81$, $MSE = 2.03$, $p < .05$, $\eta^2 = .43$, indicating that mean solution times differed across training groups. There was also a large effect of test problem, $F(1, 105) = 17.77$, $MSE = .62$, $p < .05$, $\eta^2 = .15$, and a large interaction of Training \times Test problem indicating that solution times differed across test problems as a function of group, $F(4, 105) = 3.94$, $MSE = .62$, $p < .05$, $\eta^2 = .13$.

Follow-up comparisons for the main effect of training revealed that the practice groups solved the test problems much faster than the instruction and no-training groups, $F(1, 105) = 73.58$, $MSE = 2.03$, $p < .02$, $\eta^2 = .40$. Follow-up comparisons for the interaction of Training \times

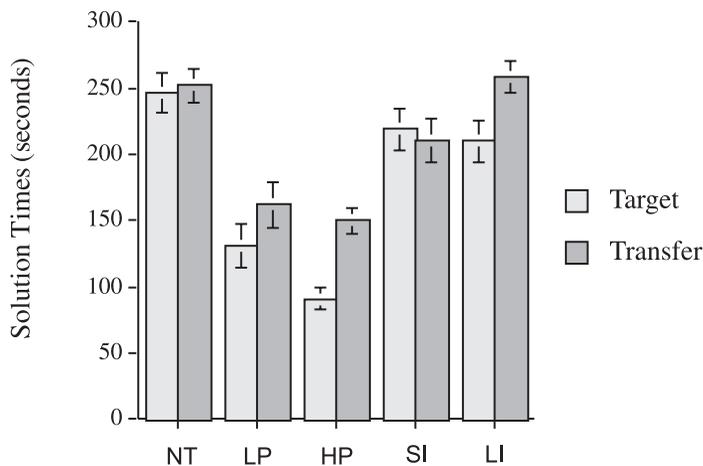


Fig. 6. Mean solution times (seconds) and standard errors for the no-training (NT), low-variability practice (LP), high-variability practice (HP), short instruction (SI), and long instruction (LI) groups on the target and transfer problems.

Test problem revealed that the practice and long instruction groups solved the target problems faster than the transfer problems, $F(1, 60) = 47.00$, $MSE = .39$, $p < .02$, $\eta^2 = .44$, whereas the short instruction and no-training groups solved both target and transfer problems equally fast, $F(1, 48) = .001$, $MSE = .89$, *ns*.

In sum, the results show that the practice groups solved the target and transfer problems faster than both of the direct instruction and no-training groups. Although the long instruction group achieved the same high-level performance as the practice groups as measured by mean accuracy, they did not show an advantage over the short instruction and no-training groups with respect to solution time. In addition, both the practice and long instruction groups showed a significant slow down on the transfer problem.

3.3.2.3. Patterns of extrapolation. In addition to accuracy and solution time measures we also examined the order in which the participants extrapolated—or filled in—the problem solutions. Analyzing participants' *completion patterns* provides data to assess the types of strategies and procedures they used to solve the problems. If the training participants' completion patterns are different from those of the control group this would provide further evidence that the knowledge acquired from training affected how they solved the test problems.

In addition to assessing the overall effect of training on pattern extrapolation, we also can further examine the hypothesis that instruction and practice generate different types of knowledge structures. Specifically, if participants given direct instruction acquired primarily declarative knowledge, we should expect them to exhibit *more variability* in their completion patterns than the participants given practice who had three opportunities to proceduralize their knowledge. Consistent with this prediction we should also expect more participants from the instruction condition to change their completion patterns between the target and transfer problems as compared to the participants in the practice conditions.

Because participants had to extrapolate each problem to eight positions, the total number of possible completion patterns for a given problem was $8!$ or 40,320. We therefore had an adequately large problem space to detect differing amounts of strategy variability. First we examined the completion patterns of the no-training group to provide a baseline measure of extrapolation strategies. The completion pattern most frequently used by these participants was a straightforward 1-2-3-4-5-6-7-8, which is a left-to-right strategy. This was the predominant strategy used by the no-training group regardless of the test problem (target vs. transfer) or problem type (Problem 1 vs. Problem 2).

Second, to assess the differences in strategy variability between the training groups, we classified each participant's completion patterns into one of three categories. If the participant used the left-to-right strategy his or her solution was classified as "basic." If the participant used a completion pattern other than the basic pattern, and another participant from his or her training group used an identical pattern for that problem, the participant's solution was classified as "common." Finally, if the participant used a completion pattern that was not the basic pattern and was not used by anyone else from his or her training group for that problem, the participant's solution was classified as "unique."

In Table 4, columns 2 through 4, we present the proportion of completion patterns for each training group classified as basic, common, or unique (collapsed across test problem and prob-

Table 4

The proportion of completion patterns classified as basic, common, or unique and the average number of strategies used by each training group

Training (<i>n</i>)	Basic (%)	Common (%)	Unique (%)	Number of Strategies
No training (30)	75	11	14	6.75
Practice low (20)	30	42.5	27.5	8.75
Practice high (21)	37	37	26	9.00
Instruction short (19)	58	8	34	8.25
Instruction long (20)	34	16	50	12.5

lem type). In the fifth column we present the average number of *different completion patterns* used by each group (also collapsed across test problem and problem type).

The results show that the training participants' use of the three types of completion strategies significantly differed from that of control. Specifically, the participants in the control group used more basic pattern completions than the other training groups, $\chi^2(4, N = 439) = 58.36, p < .05$. This result indicates that the knowledge acquired from training facilitated other equally or more effective pattern completion strategies.

Analysis of the completion patterns also showed that participants in the long direct instruction group used more unique pattern completions than the participants in the practice groups, $\chi^2(2, N = 243) = 13.40, p < .05$. In addition, the long instruction group also used a higher number of different completion patterns than the practice groups ($M = 12.44$ vs. $M = 8.88$). In sum, these results indicate that the long instruction group showed significantly more variability in its completion strategies than the practice groups.

3.3.2.3.1. Change of strategy. To assess whether participants changed strategies between the test problems, each participant's target and transfer completion patterns were compared for each problem type. If the completion patterns were identical, the participant was classified as having used the "same strategy," if they differed, he or she was classified as having "changed strategies." Table 5 presents the proportion of subjects to change or use the same strategies for the test problems collapsed across problem type.

Table 5

Proportion of participants to change strategies across target and transfer problems

Training (<i>n</i>)	Used Same Strategy (%)	Changed Strategy (%)
No training (30)	72	28
Practice low (20)	68	32
Practice high (21)	64	36
Instruction short (19)	47	53
Instruction long (20)	25	75

The results show that the participants in the long direct instruction group were more likely to change strategies between the target and transfer problems than the participants in the practice groups, $\chi^2(2, N = 61) = 10.18, p < .05$ and $\chi^2(2, N = 61) = 8.04, p < .05$ for Problem types 1 and 2, respectively.

3.3.2.3.2. Strategy overlap. One final source of evidence to evaluate whether participants given practice versus instruction acquired different knowledge structures is to examine the amount of strategy overlap between the two groups. If the two groups acquired different knowledge structures, and that knowledge facilitated different types of extrapolation strategies, then the strategy overlap should be low. Given this logic we should also expect the number of shared strategies within each training group to be greater than the strategy overlap between groups. That is, the participants within a given training group should exhibit more similarity in extrapolation strategies to one another than to participants from a qualitatively different type of training.

We calculated one between-group measure of strategy overlap (practice vs. instruction) and two within-group measures, one for practice (low and high variability) and one for instruction (short and long). The percentage of between-group overlap was determined by dividing the number of shared strategies—that is, completion patterns that were used by at least one participant from each training condition—by the total number of strategies used by those groups. The percentage of within-group overlap was determined by dividing the number of shared strategies—that is, completion patterns that were used by at least two participants—by the total number of different strategies used by that group. The within-group measures revealed 33% overlap for the practice groups and 16% overlap for the instruction groups, and there was only 9% overlap between the instruction and practice groups.

In summary, both of the training groups showed higher within-group strategy overlap than between-group overlap. The practice groups had the highest amount of overlap, suggesting that repeated practice of finding and generating a particular pattern led to the discovery of some of the same strategies for extrapolating that pattern. The instruction group had less within-group strategy overlap than the practice group. This result is consistent with the notion that the interpretation of a declarative structure results in greater variance for articulating that knowledge, resulting in many strategies and less overlap. Finally, the overlap between the two groups was extremely low, only 9%, showing that these groups used very few of the same strategies. This result supports the hypothesis that the two training groups acquired different knowledge and that knowledge led to different strategies for solving the extrapolation problems.

3.4. Discussion

Direct instruction and practice can both facilitate performance on sequence extrapolation problems, as shown by the fact that performance on the target problem was more accurate for both the long instruction and the two practice groups than for the no-training group. This result was not self-evident, because not every training procedure produces higher accuracy in this domain: The short direct instruction group did not perform significantly better than the

no-training group. The higher accuracy of the better groups was not due to time on task, because time spent on training was longer for instruction than for practice, and the mean solution time on the target task was shorter for the groups with the higher accuracy. We infer that a person can master letter-sequence extrapolation problems via either instruction or practice, but what is learned differs in the two scenarios.

3.4.1. Type of knowledge

The fact that the two instruction groups took longer to complete the problems than the two practice groups is consistent with the expectation that the instruction scenario prompts the construction of more declarative than procedural knowledge, whereas the opposite is true for the practice scenario. Declarative knowledge needs to be proceduralized or compiled to guide the solving of an unfamiliar problem, a cognitive process that is complex and likely to take time. Practice, on the other hand, generates procedural knowledge and its application to a new problem is fast; hence the shorter solution times. In the next section we show that this difference in time holds at each level of accuracy, but only for the extrapolation time component of the solution process.

Additional support for this hypothesis comes from the expectation that declarative knowledge is subject to multiple different uses or articulations, whereas procedural knowledge is not. Our results are consistent with this hypothesis in that participants in the instruction groups showed more variability in their completion patterns than did the practice groups. In particular, the participants in the long instruction condition exhibited more pattern completion variability than all other training groups. They were also more likely than the other training participants to change strategies between the target and transfer problems.

There was a clear difference between the short and long instruction groups in favor of the latter. The primary difference in training between those two groups was the inclusion of two extra pages describing the type of inferences one needs to carry out to extrapolate a letter sequence. Because this information was procedural in nature, one might expect it to facilitate the generation of procedural knowledge. However, the lack of effect on the solution time is inconsistent with this prediction. Nevertheless, the difference in accuracy between the two groups demonstrates that the long instruction group benefited from the extra instruction in some way. It is possible that the two extra pages resulted in a *declarative* representation of the required inference type. We return to this issue in a later section.

3.4.2. Abstraction

The fact that the training facilitates performance on the target problem shows that what is learned during training is not completely specific. Although the target problem used the same pattern as the participants encountered during training, that pattern was instantiated in different letters. What was learned must therefore have been of at least object abstraction to apply to the target problem. The fact that long instruction and practice groups took longer to solve the transfer problem shows that the abstraction level nevertheless was limited. Additional cognitive processing was needed to apply what had been learned to the transfer problem.

In the next section, we move toward an explanation of these findings by considering in detail how much processing was required in each group.

Table 6
Mean deliberation and extrapolation times and standard deviations for the training and control groups

Training Group	Deliberation Time		Extrapolation Time	
	Target (<i>SD</i>)	Transfer (<i>SD</i>)	Target (<i>SD</i>)	Transfer (<i>SD</i>)
No training	122 (61)	126 (73)	125 (63)	126 (57)
Short instruction	60 (35)	50 (33)	159 (58)	161 (57)
Long instruction	54 (35)	45 (28)	156 (55)	214 (56)
Practice low	56 (47)	66 (53)	75 (44)	96 (43)
Practice high	32 (18)	48 (22)	59 (27)	102 (38)

4. A processing account

The purpose of this section is to formulate an account of what happened in our experiment in terms of what was learned—which type of knowledge and which level of abstraction—and in terms of the cognitive processing that the knowledge required in each learning scenario.

Closer scrutiny of the solution times show that the three scenarios we studied differed in ways that cut across similarities and differences in mean accuracy. We examine the deliberation and extrapolation time components of the solution time. Recall that the deliberation time is an estimate of how long it took the participant to identify the pattern in the given sequence, and the extrapolation time is an estimate of how long it took him or her to carry out the extrapolation inferences. It is possible that some participants may have engaged in a more dynamic and integrated deliberation and extrapolation process than that assumed here. We acknowledge this possibility and assert that these measures are only approximations of pattern detection and extrapolation times.

In the following analysis, we focus on the long instruction group and the high-variability practice group and use the results from the parametric variants of these scenarios as auxiliary evidence. Modified Bonferroni tests were conducted for all planned comparisons and the alpha level was set to .02 (Keppel, 1991). Table 6 shows the deliberation and extrapolation times for the training and control groups.

4.1. Interpretation: Direct instruction

During training, the long instruction participants learned to recall the relations in the pattern to a high degree of accuracy, so they must have acquired some representation of the target pattern. These participants had no opportunity to practice, so it is plausible that their knowledge consisted of a schema or some other type of declarative representation. This representation was applicable to a novel letter sequence, because the deliberation time for the target problem was shorter for the instruction group than for the control group, $t(48) = 4.48, p < .02$. Hence, it must have been of at least object abstraction. Presumably, the deliberation time for this group was spent matching their schema to the given sequence in the target problem.

The data suggest that their schema was also applicable to the transfer problem with little further cognitive work. The time to the first extrapolation was no longer for the transfer problem than for the target problem; instead, it was 9-sec shorter, a clear case of declarative transfer.

Presumably, the facilitation occurred because the new pattern had the same structure, that is, the important relations held between the same positions in the target and transfer patterns, and because the particular relations in the transfer pattern were generated from the relations in the target pattern via a systematic transformation (i.e., stretching; see Methods section). In sum, the data suggest that the deliberate study of the target pattern generated a relation abstraction, an easily transferable representation of the target pattern.

The extrapolation time presents a contrasting picture. Because they had had no prior opportunity to practice extrapolation inferences during the training phase, the direct instruction participants—especially those who performed at a high level of accuracy—must have constructed the necessary procedural knowledge in the course of performing the target problem. Consistent with this, the time for extrapolation on the target problem was considerably longer in the long instruction group than in the practice groups, $t(39) = -7.23, p < .02$ and $t(38) = -5.15, p < .02$, for high and low variability.

One might expect the procedural knowledge generated from the declarative schema to be at the same level of abstraction as the schema itself and hence easily transferable, but the data suggest that this was not the case. The extrapolation time for the target problem was 156 sec, and for the transfer problem 214 sec; see Table 6, $t(19) = -3.80, p < .02$. This 37% increase contrasts with the slight decrease in the corresponding deliberation times. This result should be considered in the context of the short transfer distance from the target to the transfer problem. It appears that the amount of cognitive work required for extrapolation on the transfer problem (at which point the participants had some relevant procedural knowledge, constructed in the course of solving the target problem) was greater than the work needed for extrapolation on the target problem (for which the participants had no prior procedural knowledge). The participants either invested cognitive work into generalizing the rules created in the course of solving the target problem, or else they started over and created brand new rules for the transfer problem. Under either interpretation, we infer that the procedural knowledge constructed in the course of solving the target problem was not transferable.

The short instruction group exhibited the same pattern of results as the long instruction group with respect to the deliberation and extrapolation times. The deliberation time on the target problem was shorter than that of the control group, $t(47) = 4.02, p < .02$, and comparable to that of the long instruction group. The extrapolation time was longer than for the high- and low-practice groups, $t(38) = -7.14, p < .02$ and $t(37) = -5.19, p < .02$, and again comparable to that of the long instruction group. Finally, the short instruction group showed the same pattern of a small decrease in deliberation time on the transfer problem, coupled with an extrapolation time that was as long as that for the target problem. Although the short instruction group did not perform as well as the long instruction group as measured by mean accuracy, the replication of the structure of deliberation and extrapolation times indicate that the processing was similar and that the knowledge this group acquired was of the same type and level of abstraction.

4.2. Interpretation: Practice

The high-variability practice group improved its performance across the three practice problems. Because the practice participants were not shown the underlying pattern, but had to

work it out for themselves, this improvement most likely signals the acquisition of a mixture of declarative knowledge about the target pattern and procedural knowledge about how to detect patterns. This representation was more useful than the one learned by the instruction groups, because the deliberation time for the high-variability practice group on the target problem was shorter than the corresponding times for either the long or the short instruction group; see Table 6, $t(39) = -2.56, p < .02$ and $t(38) = -3.25, p < .02$.

What was the level of abstraction attained by this group? The practice group exhibited an increase in time to first extrapolation from 32 sec on the target problem to 48 sec on the transfer problem, in contrast to the small decrease in deliberation time exhibited by both instruction groups. Hence, what the high-variability group learned about the pattern cannot have been so abstract that it applied to the transfer problem without modification. As one would expect, their knowledge was characterized by object abstraction.

Over the course of attempting three different extrapolation problems during the training phase of the experiment, the high-variability practice group must have created the procedural knowledge needed to carry out the relevant extrapolation inferences. Because these participants saw three different letter sequences that all followed the same pattern, they had an incentive and opportunity to create abstract rules to carry out the extrapolation inferences. The data bear this out. The practice group exhibited a shorter extrapolation time on the target problem than the instruction group (59 vs. 156 sec), indicating that they could apply the procedural knowledge they had already constructed rather than starting from scratch. They could achieve this in the span of only four prior problems (three practice, one target) because the distance between the target and the transfer patterns was short.

If those rules were of object abstraction, they needed to be revised and adjusted to apply to the transfer problem, but the adjustment would be minor. Again, the data bear out this expectation. For the high-variability practice group, the extrapolation time for the transfer problem was 43 sec longer than the time for the target problem, $t(20) = -8.10, p < .02$.

There was reason to expect a different pattern to hold in the low-variability practice group. Because they saw exactly the same problem (the same given letter sequence) three times before encountering the target problem, they had an opportunity to specialize their procedural knowledge to that particular sequence. If this had happened, we would expect to see evidence that they had to invest more cognitive work into adapting their pattern knowledge to the target problem (which for them exhibited a novel letter sequence) than the high-variability group. This is what the data show. Time to the first extrapolation on the target problem was longer for the low-variability than for the high-variability group, 56 sec versus 32 sec, a marginally significant effect, $t(39) = 2.23, p = .03$. The same argument would predict that the time to compute the extrapolation inferences would also be longer. Although not statistically significant, the data were in the predicted direction, 75 sec for the low-variability group versus 59 sec for the high-variability group. Once they abstracted their knowledge to fit the target problem, the low-variability group should have been at a smaller disadvantage on the transfer problem. This was indeed the case; the deliberation time on the transfer problem was still somewhat longer (66 sec vs. 48), but the extrapolation time was virtually the same as for the high-variability group (96 sec compared to 102).

In short, although the low-variability group incurred an initial cognitive cost of the kind one would expect—they had to put more cognitive work into adapting their knowledge to the unfam-

miliar tasks than the high-variability group—this initial disadvantage was overcome as soon as they had had an opportunity to abstract their procedural knowledge.

4.3. Results separated by accuracy

The previously suggested interpretation could be critiqued on the grounds that the deliberation and extrapolation times are likely to be imprecise estimates of the time needed to identify the pattern and to carry out the extrapolation inferences. In particular, participants might have oscillated between pattern detection and pattern extrapolation to a greater extent than our analysis presupposes. Furthermore, one could argue that the tendency to oscillate in this manner is a function of how well the person understands the relevant pattern, so that what looks like differences between the groups with respect to deliberation time and extrapolation time is in actuality nothing but a side effect of differences in performance level as measured by mean accuracy.

This argument is contradicted by Table 7. This table shows that the pattern of findings described previously recurs at each level of accuracy. For this analysis, we divided the participants into three levels based on their accuracy on the target problem. The top level included participants who performed perfectly with 8 correct extrapolations, the second level included those with 5 to 7 correct extrapolations, and the bottom level included those with 0 to 4 correct. To facilitate the comparison between accuracy levels, we limit Table 7 to the long instruction and high-variability practice groups, the two groups for which our claim of alternative paths to mastery is strongest.

As Table 7 shows, those features of the data on which we based our interpretation recur at each level of accuracy. The deliberation times for the practice groups are comparable (with a slight advantage for practice in the top and bottom levels); the deliberation times for transfer are longer than target for practice, but not for the instruction in the top and bottom levels; the

Table 7

Mean solution times (seconds) and standard deviations for the high variability practice and long direct instruction groups on target and transfer problems

Training Group	<i>n</i>	Deliberation Time		Extrapolation Time	
		Target (<i>SD</i>)	Transfer (<i>SD</i>)	Target (<i>SD</i>)	Transfer (<i>SD</i>)
High-level accuracy (8 correct extrapolations)					
Long instruction	9	43 (28)	42 (24)	128 (34)	211 (59)
Practice high	14	34 (29)	53 (34)	46 (32)	94 (33)
Midlevel accuracy (5–7 correct extrapolations)					
Long instruction	8	51 (22)	60 (34)	174 (79)	227 (52)
Practice high	8	50 (24)	64 (39)	71 (28)	117 (42)
Low-level accuracy (0–4 correct extrapolations)					
Long instruction	13	82 (12)	62 (38)	178 (66)	191 (67)
Practice high	9	52 (34)	61 (63)	85 (45)	73 (63)

extrapolation times for instruction are more than twice as long for practice on the target problem; and the instruction group does not catch up with the practice group on the target problem. At each of the three levels of accuracy, the extrapolation time for instruction is longer on the transfer than on the target problem. More important, at each level of accuracy, the extrapolation time is more than 100 sec longer for instruction than for practice.

The recurrence of this pattern at each level of accuracy shows that the pattern is not a side effect of the overall differences in accuracy. In particular, those instruction participants who solved the target problem perfectly obviously did not have any serious deficiency in their knowledge and understanding of the relevant patterns or any lack of relevant cognitive ability, but they still exhibited radically longer extrapolation times on both transfer and target than did the practice groups. This indicates, first, that they had to invest cognitive work into compiling the required procedural knowledge and, second, that the resulting procedural knowledge was not abstract.

5. Conclusion

Our results call into question the received view of the relation between declarative and procedural knowledge (Anderson, 1976; Winograd, 1975). The supposed advantages of declarative knowledge as abstract and context independent are real, but only as long as we are talking about declarative transfer. A person can transfer a pattern readily enough and, hence, see a new situation in the light of a previously learned pattern. However, this ability is of limited value and may or may not be expressed in overt behavior, because it does not bring with it an equally powerful and flexible ability to decide what to do in the new situation. The obstacle to efficient action resides in the process of articulating the action implications of the declarative knowledge, not in the declarative knowledge itself. In contrast, procedural knowledge generated via practice, supposedly context dependent, was here shown to be easier to transfer. Expected context dependence emerged for low variability but not for high-variability practice. Consequently, we have to call into question the usual trade-off explanation for why people have these two types of knowledge (Anderson, 1983; Winograd, 1975).

That people can learn about a cognitive task in different ways, that they can reach similar performance levels from qualitatively different types of training, that behavior can differ in its articulation and chronological structure within similar performance levels, and that such a structure can be replicated within distinct performance levels—these observations constitute a set of phenomena that a general learning theory ought to be able to account for. That is, a learning theory should be able to predict, for a given path to mastery, what knowledge is acquired along that path and how that knowledge is encoded with respect to type, abstraction level, and other properties, and to predict the pattern of differences in a set of outcome measures applied to several paths. It is possible that our particular findings will not replicate in a different task domain, but for every task domain there is some pattern of such differences, and that pattern is a phenomenon against which we can test any learning theory that claims generality.

Notes

1. The main accuracy and solution time results reported in this section were also significant in the entire sample and are available at <http://cognitivesciencesociety.org/supplements/>.

Acknowledgments

The work reported here was supported by an Abraham Lincoln Fellowship from the University of Illinois at Chicago to the first author and Grant No. N00014-99-1-0929 from the Cognitive Science Program of the Office of Naval Research (ONR) to the second author. No endorsement should be inferred. We thank Brandy Jones and Chris Stanley for help in collecting and analyzing the data. We are also grateful to James Dixon, Robert Goldstone, and two anonymous reviewers for their many helpful comments and suggestions on the paper.

References

- Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3–19.
- Anderson, J. R. (1976). *Language, memory, and thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R., Greeno, J. G., Kline, P. J., & Neves, D. M. (1981). Acquisition of problem-solving skill. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 191–230). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review*, *86*, 124–140.
- Carroll, J. B. (1968). On learning from being told. *Educational Psychologist*, *5*, 4–10.
- Cohen, J. (1988). *Statistical power analysis of the behavioral sciences* (2nd ed.). New York: Academic.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271.
- Gick, M. L., & Holyoak, K. J. (1987). The cognitive basis of knowledge transfer. In S. M. Cormier & J. D. Hagman (Eds.), *Transfer of learning* (pp. 9–46). San Diego: Academic.
- Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*, 414–466.
- Greeno, J. G., & Simon, H. A. (1974). Processes for sequence production. *Psychological Review*, *81*, 187–198.
- Keppel, G. (1991). *Design and analysis: A researcher's guide*. Upper Saddle River, NJ: Prentice Hall.
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*, 661–667.
- Klahr, D., & Wallace, J. G. (1970). The development of serial completion strategies: An information-processing analysis. *British Journal of Psychology*, *61*, 243–257.
- Knowlton, B. J., & Squire, L. R. (1993, December 10). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.
- Kotovsky, K., & Simon, H. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology*, *4*, 399–424.
- Markman, A. B. (1999). *Knowledge representations*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

- Maxwell, J. P., Masters, R. S. W., & Eves, F. F. (2003). The role of working memory in motor learning and performance. *Consciousness and Cognition*, 12, 376–402.
- Neves, D. M., & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 57–84). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.
- Nokes, T. J., & Ohlsson, S. (2003). Declarative transfer from a memory task to a problem solving task. *Cognitive Science Quarterly*, 3, 259–296.
- Ohlsson, S. (1993). Abstract schemas. *Educational Psychologist*, 28, 51–66.
- Ohlsson, S. (1994). Declarative and procedural knowledge. In T. Husen & T. Neville-Postlethwaite (Eds.), *The international encyclopedia of education* (Vol. 3, 2nd ed., pp. 1432–1434). London, England: Pergamon.
- Ohlsson, S. (1996). Learning from performance errors. *Psychological Review*, 103, 241–262.
- Ohlsson, S., Ernst, A., & Rees, E. (1992) The cognitive complexity of doing and learning arithmetic. *Journal for Research in Mathematics Education*, 23, 441–467.
- Ohlsson, S., & Lehtinen, E. (1997). Abstraction and the acquisition of complex ideas. *International Journal of Educational Research*, 27, 37–48.
- Ohlsson, S., & Rees, E. (1991) The function of conceptual understanding in the learning of arithmetic procedures. *Cognition & Instruction*, 8, 103–179.
- Olejnik, S., & Algina, J. (2000). Measures for effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, 25, 241–286.
- Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 548–568.
- Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, 77, 481–495.
- Restle, F., & Brown, E. (1970). Organization of serial pattern learning. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 4, pp. 249–331). New York: Academic.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomena. *Educational Psychologist*, 24, 113–142.
- Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.
- Simon, H. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 79, 369–382.
- Simon, H. A., & Kotovsky K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, 70, 534–546.
- Squire, L. R., Knowlton, B., & Musen, G. (1993). The structure and organization of memory. *Annual Review of Psychology*, 44, 453–495.
- Thurstone, L., & Thurstone, T. (1941). *Factorial studies in intelligence*. Chicago: University of Chicago Press.
- Winograd, T. (1975). Frame representations and the declarative-procedural controversy. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 185–210). New York: Academic.