COGNITIVE SCIENCE **20**, 75-100 (1996)

The Impact of Goal Specificity on Strategy Use and the Acquisition of Problem Structure

REGINA VOLLMEYER
University of Potsdam, Germany

BRUCE D. BURNS
KEITH J. HOLYOAK
University of California, Los Angeles

Theories of skill acquisition have made radically different predictions about the role of general problem-solving methods in acquiring rules that promote effective transfer to new problems. Under one view, methods that focus on reaching specific goals, such as means-ends analysis, are assumed to provide the basis for efficient knowledge compilation (Anderson, 1987), whereas under an alternative view such methods are believed to disrupt rule induction (Sweller, 1988). We suggest that the role of general methods in learning varies with both the specificity of the problem solver's goal and the systematicity of the strategies used for testing hypotheses about rules. In the absence of a specific goal people are more likely to use a rule-induction learning strategy, whereas provision of a specific goal fosters use of difference reduction, which tends to be a non-rule-induction strategy. We performed two experiments to investigate the impact of goal specificity and systematicity of rule-induction strategies in learning and transfer within a complex dynamic system. The results of Experiment 1 indicated that during free exploration of a problem space, greater learning occurred if participants adopted more systematic strategies for rule induction, and that participants come to favor such strategies. Experiment 2 revealed that participants who were provided with a specific goal performed well on the initial problem but were impaired on a transfer test using a similar problem with a different goal. Instruction on a systematic rule-induction strategy facilitated solution for both the initial and transfer problems, but participants' use of this strategy declined if they had a specific goal. Our results support Sweller's (1988) proposal that general problem-solving methods applied to a specific goal foster acquisition of knowledge about an isolated solution path but do not provide an effective way of learning the overall structure of a problem space. We interpret these results in terms of dual-space theories of search through problem space.

Preparation of this article was supported by NSF Grant SBR-9310614 to K.J.H. and by DFG Grant Vo 514/1 to R.V. We would also like to thank Michelene Chi, John Sweller, and an anonymous reviewer for comments on an earlier draft of this article. A preliminary report of Experiment 2 was presented at the sixteenth annual Conference of the Cognitive Science Society, 1994 (August).

Correspondence and requests for reprints should be sent to Regina Vollmeyer, Institut für Psychologie, Universität Potsdam, 14415 Potsdam, Germany. E-mail: <vollmeye@rz.uni-potsdam.de>.

INTRODUCTION

A central problem in knowledge acquisition is to identify the relationship between problem solving and learning. People can learn from solving problems, but it is unclear exactly how learning takes place or what is learned. People sometimes seem to learn little from a problem-solving episode except a specific solution to a particular problem; yet on other occasions, people acquire more general knowledge that can be applied to a wide range of related problems. What is the difference?

Impact of Goal Specificity on Learning

A particularly intriguing possibility is that some solution methods may be effective for finding solutions to specific problems, but relatively ineffective in promoting abstraction of knowledge of the structure of the problem that would support transfer to novel but related problems. A case of particular theoretical interest concerns the role of general problem-solving methods (often termed "weak methods") in learning. One such general method is means-ends analysis, which involves difference reduction (removing the largest difference between the current state and goal state), combined with subgoaling (recursively solving the subproblem of getting from the current state to that which satisfies the preconditions of required operators). Some theories of learning have claimed that means-ends analysis, while itself a weak problem-solving method used primarily by novices, is nonetheless a valuable stepping stone toward expertise. According to this view, solutions first generated by means-ends analysis are subsequently compiled into rules that allow more efficient solutions to be found for problems similar to the original one (e.g., Anderson, 1987; Larkin, 1981).

Other theorists, however, have argued that means-ends analysis and similar problem-solving methods can actually impede the acquisition of general rules (e.g., Mawer & Sweller, 1982; see Holyoak, 1991, for a brief review). Means-ends analysis can be applied to problems with well-defined operators and a specific goal; however, its immediate product is not knowledge of the rules that govern the problem, but simply a solution path that achieves the immediate goal. We will term a strategy that achieves a specific goal without necessarily yielding rules a *goal-oriented* strategy. In contrast, other learning strategies can operate on ill-defined problem situations that lack a specific goal. In the absence of a specific goal, free exploration of a problem space may yield rules about state transitions, which can later be used to achieve a relatively wide variety of goals, thus promoting transfer to a family of similar problems. Such exploratory strategies may be effective for rule acquisition, in contrast to goal-oriented strategies which may be most effective for achievement of a specific goal. We consider a goal to be *nonspecific* if the focus is other than towards reaching a solution state. In the absence of such a specific goal we assume that people will be more likely to focus on learning a concept or a rule that describes the structure of the problem.

Sweller and coworkers found evidence that people with a nonspecific goal gained more knowledge about a task than did people with a specific goal (e.g., Mawer & Sweller, 1982; Sweller, 1988). These investigators have examined goal specificity with a variety of different problems: the Tower of Hanoi problem (Sweller, 1983, Experiment 1), maze learning (Sweller, 1983, Experiment 2; Sweller & Levine, 1982), geometry (Sweller, 1988; Sweller, Mawer, & Ward, 1983, Experiments 4, 5, 6, 7), mathematics (Mawer & Sweller, 1982; Sweller, Mawer, & Howe, 1982), and kinematics (Sweller et al., 1983, Experiments 1, 2, 3). They have interpreted their results as evidence that people apply different learning strategies depending on the specificity of the stated goal. For example, participants in one experiment involving solving geometry problems were provided with partial information about the angles and sides of a triangle and were asked either to calculate all possible angles and sides (nonspecific goal) or to calculate a particular angle (specific goal); however, these problems were designed so that all angles had to be calculated in order to achieve either type of goal. Participants given a specific goal appeared to form subgoals (finding which angles and sides were necessary to calculate the unknown angle) and to solve the problem by means-ends analysis, a goal-oriented strategy. In contrast, participants given a nonspecific goal appeared to use a strategy suitable for inducing rules. Participants who received the nonspecific goal were subsequently more successful in solving problems.

By what mechanism could the specificity of a goal affect one's strategy? A number of theorists have argued that people can perform search in multiple problem spaces (e.g., Dunbar, 1993; Klahr & Dunbar, 1988; Simon & Lea, 1974). Klahr and Dunbar (1988) applied a dual-space framework to hypothesis testing. In their model, effective hypothesis testing consists of coordinated search through two different spaces: *hypothesis space* and *experiment space*. Search of hypothesis space consists of generating and modifying hypotheses about the structure of the system. Predictions derived from rules are than tested by experiments based on search of the experiment space. Klahr and Dunbar's model was an extension of that proposed by Simon and Lea (1974), who claimed that problem solving and induction can be contrasted as searches of different spaces. Simon and Lea defined two different types of problem spaces: *instance space* and *rule space*. Instance space is a problem space in which it is possible for the new state to be directly tested agains the goal state (e.g., checking if all disks in a Tower of Hanoi problem are on the destination peg). Klahr and Dunbar's experiment space is roughly equivalent to instance space. Rule space has the property that a state (i.e., a possible rule) cannot be tested against a goal in that same space; instead, rules must be tested by experiments that generate states in instance space. Rule space is thus roughly equivalent to Klahr and Dunbar's hypothesis space.

Simon and Lea (1974) defined a problem-solving task as one that only requires search of instance space, whereas an induction task requires an interaction between search of both instance space and rule space. The nature of the task, thus, dictates the type of strategy that will be used: A strategy effective for searching the instance space will be used with problem-solving tasks; strategies suitable for searching rule space will be necessary for induction tasks. The dual-space framework also provides a way of understanding the distinction between specific and nonspecific goals. A specific goal is a state in the instance space. A nonspecific goal is the absence of a goal in instance space. Because they lack direction as to how to search instance space, learners who do not have a specific goal may use exploration of rule space to direct their search of instance space. If the specificity of goals affects whether a problem is seen primarily as search of one space or the other, then this factor may affect what types of strategies are used.

By Simon and Lea's (1974) definition, in a pure problem-solving task there is no need to generate hypotheses; rather, all that is required is movement through instance space by the application of operators. Insofar as a task is amenable to well-defined, goal-oriented strategies that are effective for searching the instance space, such as means-ends analysis or difference reduction, it will be possible to achieve goals defined in instance space. But the more a task requires the generation and evaluation of rules (i.e., movement through rule space), the less beneficial will be a strategy that only searches instance space, and the more useful will be a strategy that searches rule space. Means-ends analysis, like any goal-oriented strategy, compares a current state to a goal state in the same space, and thus it is only applicable to search of instance space. Means-ends analysis may support hypothesis testing by providing an efficient strategy for testing rules, but as standardly interpreted, it is not a strategy for generating rules. One might claim that means-ends analysis could be used to generate rules if we allow nonspecific goals such as "find rule" to be considered specific solution states within rule space; however, this usage is quite different from the standard form of means-ends analysis as it is applied in instance space.

Simon and Lea (1974) pointed out that the same task can often be treated as either an induction task or a problem-solving task, depending on the space people search. For example, the widely studied problem-solving task, the Tower of Hanoi, can be treated as a hypothesis-testing task in which a general procedure for accomplishing the movement of disks is induced. Usually, however, participants only search instance space in their efforts to solve the problem. Most problem-solving tasks are probably in fact a mixture of induction and problem solving; however, the degree to which participants will search both spaces, rather than instance space alone, may vary with the goals of the participant. In order to investigate the impact of goal specificity on learning, this study examined the performance of participants on the same basic task, which could be treated as either a hypothesis-testing or a problem-solving task. Participants with a nonspecific goal should be more likely to treat it as a hypothesis-testing task and use an appropriate strategy which will lead them to gain more knowledge about the structure of the task.

Impact of Systematicity on Effectiveness of Strategies

People differ in the degree of systematicity with which they formulate and test their hypotheses, and those who formulate hypotheses in a task-appropriate and testable way generally gain more knowledge (e.g., Dunbar, 1993; Klahr & Dunbar, 1988; Klahr, Fay, & Dunbar, 1993). Tschirgi (1980) identified different strategies for testing hypotheses about the influence of multiple factors on one or more dependent variables. One highly systematic strategy is the VOTAT (Vary One Thing At a Time) strategy, in which one factor is varied while the others are held constant. VOTAT contrasts with less systematic strategies such as changing all factors haphazardly (CA, for Change All). (Similar strategy classifications have been suggested by Branke, 1991, and Putz-Osterloh, 1993.) As Tschirgi pointed out, the VOTAT strategy allows the logical disconfirmation of alternative hypotheses, and thus it is central to experimental design in science. Tschirgi found that VOTAT was the most common strategy employed by adults when they encountered negative outcomes. In a study of hypothesis generation and evaluation, Klahr et al. (1993) found that participants who changed only one aspect of the task at a time (i.e., used VOTAT) were more successful at identifying correct rules than were those who varied multiple aspects at once.

Although previous work implicates both goal specificity and systematicity of strategies as important determinants of what is learned during problem solving, prior research has not investigated the relationship between these factors. This study seeks to measure and manipulate both factors in order to investigate the influence of goal specificity and systematicity of learning strategies on the acquisition and transfer of knowledge about a complex dynamic system. We suggest that although these are distinct factors, they will have similar effects on a complex problem-solving task because each alters the extent to which the task is treated as hypothesis testing, thus encouraging use of a strategy appropriate to searching rule space. Accordingly, each factor should have an impact on problem solving and transfer performance through improvements in learners' knowledge. The learning domain, biology-lab (description follows), was chosen because it is especially suitable for investigating the interrelationships between problem solving and hypothesis testing, and because it makes it possible to measure the quality of learners' knowledge in multiple ways.

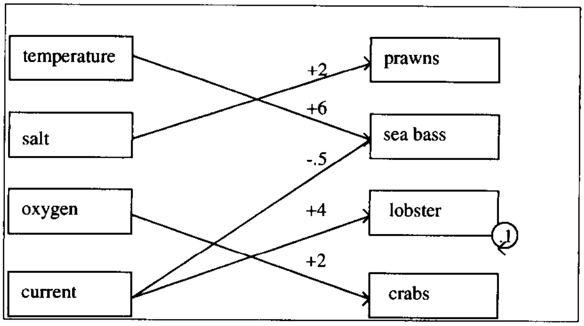

Figure 1. Structure of the system for biology-lab.

Biology-Lab: A Dynamic Problem Environment

Since the early 1980s, researchers have used computer-simulated scenarios to study complex problem solving (for a review, see Funke, 1991). These tasks are relatively complex, as multiple variables have to be manipulated in order to achieve multiple goals simultaneously. In this study, we used a computer-driven dynamic problem environment we termed *biology-lab,* constructed using the shell DYNAMIS (Funke, 1991). In our cover story, participants were told that they were in a biology lab in which there is a tank with four species of sea animal (crabs, prawns, lobsters, sea bass). These species are affected by four input variables (temperature, salt, oxygen, current). The structure of the environment, illustrated in Figure 1 (which was never shown to our participants), was such that two of the outputs (prawns and crabs) are relatively simple to manipulate because each is influenced by only one input. The other two outputs are more complex, because each is influenced by two factors. One output (sea bass) is affected by two inputs, and the other (lobster) is affected by a decay factor (marked as a circle connected to the output) in addition to a single input variable. The decay factor was implemented by multiplying the output by a constant factor (less than 1) on each trial. Decay is a dynamic aspect of the system, because it yields state changes even if there is no input (i.e., all inputs are set to zero). The system is thus complex in that it involves multiple input variables that must be manipulated to control multiple output variables, and dynamic in that the state of the system changes as a joint function of external inputs and internal decay.

To calculate the new value for each output value i on trial t, the following formula was applied:

$$\text{output}_{i,t} = (1-\lambda_i)\text{output}_{i,t-1} + \Sigma\, w_{ij}\, \text{input}_{j,t}$$

where λ_i is the decay factor, $input_j$ is the number entered for input unit j, and w_{ij} is the weight on the link between input unit j and output unit i. Note that an input variable represents change, so that setting an input to zero means that the variable is not responsible for any change in the system.

The biology-lab task can be approached in two different ways. It can be treated as a problem-solving task in which the participant tries to bring the system to a goal state. Reaching a specific goal could be accomplished by manipulating the pattern of inputs in order to reduce the difference between the current output state and the goal state, thus searching the instance space. Alternatively, the biology-lab can be treated as a hypothesis-testing task in which a participant tries to discover the rules that govern the behavior of the system, which is a nonspecific goal, and to test these rules by setting inputs to see if the predicted outputs are generated. Such an approach involves search of both rule space and instance space. The biology-lab allows us to assess the amount of knowledge that participants induce both by such indirect measures as transfer performance, as previous studies have done, and by more direct measures of their knowledge of specific regularities.

The biology-lab task contrasts with the tasks used by Sweller and coworkers (Mawer & Sweller, 1982; Sweller, 1983, 1988; Sweller & Levine, 1982; Sweller et al., 1982; Sweller et al., 1983) in that the tasks used by them were not dynamic and were less complex. Their problem-solving tasks involved a small number of well-defined operators, and the tasks required discovering how to use the operators. Thus, their tasks were simple in that their behavior was governed by a small number of known rules. In biology-lab, by contrast, the rules are unknown and the major aim is to discover what the rules are. This rule space is very large as participants must discover which variables are linked, the nature of these links (linear or multiplicative), and the magnitudes of the weights on these links, in order to discover how the whole system works. The number of operators is large, as any real number can be entered as the value for each input variable; and the instance space is also large as it consists of all possible combinations of values for the output variables. In complex tasks in which discovering the rules is very difficult, using strategies that explore only instance space by generating new states may help learners move closer to the goal, even if they do not discover the rules governing the system's behavior. This possibility is illustrated by the work of Berry and Broadbent (1987), who found no relationship between a participants' performance in a complex dynamic-system task and his or her explicit knowledge of that system. In complex tasks, participants may show similar performance in reaching goals but do so via different methods, either problem solving or hypothesis testing. It follows that in complex dynamic tasks participants may achieve performance of equivalent quality in terms of reaching a goal specified from the beginning, but those taking a hypothesis-testing approach should show better transfer to a task involving a new goal. The biology-lab task, which has the requisite complexity and dynamic quality, may allow us to distinguish when learning is due to one or the other of these approaches.

We used the biology-lab task to test the influence of goal specificity on participant's knowledge of the system, their accuracy in solving a specific initial goal, and their ability to transfer their knowledge to similar problems with different goals. We predicted that participants given a specific goal would learn enough to achieve the given goal but have a poorer knowledge of the structure of the task than participants given a nonspecific goal. The more complete representation of the system attained by the latter group was expected to lead to more effective transfer to other biology-lab problems with altered goals.

We also manipulated systematicity of participants' hypothesis-testing strategies. Our hypothesis was that a systematic strategy, VOTAT, would be more effective than less systematic strategies for acquiring general structural knowledge about the domain, resulting in more knowledge and more effective problem-solving performance. In Experiment 1, we examined participants' spontaneous use of strategies when given a nonspecific goal, in attempting to validate our assumption that VOTAT would be an effective strategy for learning about the system. In Experiment 2, both goal specificity and systematicity of learning strategy were experimentally manipulated in order to investigate their individual and joint effects.

EXPERIMENT 1

Experiment 1 was designed to explore the strategies participants would spontaneously use when presented with a nonspecific goal in the learning phase of the biology-lab task. The biology-lab system requires relatively complex induction as rule space is quite large, in that there are four input and four output variables with unknown connections. At the simplest level of hypotheses about what is linked to what, every possible connection between an input and an output variable (16 in total) and four possible decays constitute possible hypotheses. Moreover, it is necessary not only to learn the basic connections between inputs and outputs (e.g., oxygen has an influence on crabs), but also to test hypotheses about the nature of the link (linear, multiplicative, or some type of interaction) and to discover the magnitude of the influence of each connection (e.g., the input of oxygen is multiplied by 2 and added to the total number of crabs). Only after observing that a specific input variable influences a specific output variable can people begin to formulate a hypothesis and induce a rule.

The task consisted of two parts, a learning phase and a specific problem-solving phase. During the learning phase, participants were expected to construct a representation of the dynamic system. Then in the problem-solving phase, they were expected to use their mental representation of the problem space to bring the system to a specified goal state. Depending on the strategy adopted during the learning phase, we expected different representations to be formed. Given that the goal was nonspecific, we expected most participants to treat the task as hypothesis testing and, thus, to employ a systematic strategy. Our analyses, therefore, focused on the systematicity of the strategies that emerged. Participants who use a systematic strategy such as VOTAT should have more knowledge of the rule space than those who adopt less systematic strategies such as CA, because an appropriate systematic strategy should be most effective for rule induction. More complete knowledge of the system should help participants to reach the specific goal in the subsequent problem-solving phase. Unlike previous studies of the impact of different strategies on rule induction, Experiment 1 included relatively direct measures of the amount of knowledge that participants acquired during the initial learning phase. One of the aims of Experiment 1 was to validate our measures of knowledge.

Method

Participants

Thirty-six undergraduate (16 female, 20 male) students at the University of California, Los Angeles, participated for course credit.

Task and Procedure

The biology-lab problem required participants on each trial to set the levels of the four input variables and observe the resulting values of the output variables (numbers of each of four species of sea animals). The underlying structure of the system was as depicted in Figure 1. Presentation of the problem and collection of participants' responses was controlled by a microcomputer. Participants were first informed that their basic task was to discover how various water quality factors influence the reproduction of sea animals. The experimenter explained the interface and demonstrated how to manipulate the input variables and to observe how the output variables changed after every trial. A trial consisted of participants assigning inputs for each water quality factor. These inputs could be any number, positive or negative, including zero. Participants were informed that the number of one of the species was reduced on each trial by a decay factor. They were not told which species was affected by decay (it was lobster) or that another species was affected by two input factors (sea bass; see Figure 1). On each trial, participants could manipulate as many input variables as they wanted by entering any real number for each input variable. Each series of six trials was defined as a *round*. Each round (both learning and problem solving) started with the system initialized at the same number of each species (namely, 100 crabs, 200 prawns, 1,000 lobsters, 500 sea bass). Participants received four initial learning rounds followed by a fifth round in which they were asked to produce a specific goal state (namely, 50 crabs, 400 prawns, 900 lobsters, and 700 sea bass). This final round was called the *solution round.*

After each round of the learning phase (Rounds 1-4), participants completed a *structure diagram*, in which they indicated how they believed the input variables affect the output variables. They were provided with a diagram showing the inputs and outputs as in Figure 1, but with all links omitted. Their task was to draw links between variables that they believed to be dependent and also to assign weights indicating how strong they felt was the influence of the input variable. The concept of a weight between an input and an output variable was explained using the data participants generated to illustrate a possible weight from the input *oxygen* to the output *crabs*. The special symbol for decay (a loop from an output variable to itself) was also explained.

In the fifth round, the participants were presented with a specific goal state, which they had to reach in as few trials as possible and then try to maintain. After concluding the fifth round, participants answered 10 prediction questions such as, "If you have 100 crabs in your system and you set oxygen = 20, how many crabs do you have afterwards?" Success in answering the prediction questions provided another index of participants' knowledge of the overall problem space. The entire experiment took about 1 hour.

Results and Discussion

Classification of Hypothesis-Testing Strategies

For each of the four learning rounds, participants were classified into three groups on the basis of the hypothesis-testing strategies they spontaneously adopted. The three categories were (a) varying only one input variable at a time while setting the others to zero (VOTAT), (b) changing all variables in a haphazard way (CA), and (c) the heterogeneous collection of all other strategies (HT). A round was coded as reflecting the use of the VOTAT or CA category if four of the six trials of that round exhibited the relevant pattern. For example, if on four trials of a round, the participant varied all four input variables, then the strategy for that round was classified as CA. Any strategy that could not be classified as VOTAT or CA was coded as HT. These heterogeneous strategies exhibited varying degrees of systematicity. In some cases, participants tended to hold one variable constant while varying the others; in other cases, they varied two variables at a time. Use of a HT strategy was also recorded if all four variables were changed but the change was systematic so that rule induction was possible (e.g., one variable was set negative and three positive, or one variable was set to a large value and three to small values). The frequencies of these HT strategies were individually too low to warrant more fine-grained analyses; however, it is reasonable to view this category as comprising strategies of systematicity intermediate between that of VOTAT and CA.

Dependent Variables

Three dependent variables were analyzed to provide evidence of learning and transfer.

(1) ***Structure score.*** The structure diagram completed by all participants after each of the four learning rounds was used to derive a score reflecting degree of knowledge of the underlying structure of the system. This structure score was computed by finding the proportion of correct specifications for each of the three elements of the structure—links, directions, and weights—adjusted for guessing by subtracting the proportion of all possible incorrect elements given (see Woodworth & Schlosberg, 1954, p. 700). The structure score is the sum of these three adjusted proportions. Because the structure score after Round 4 was most relevant to participants' knowledge at the end of the initial learning phase, this score was used in all analyses reported here.

(2) ***Solution error.*** Solution error in reaching the specific goal state during the solution round (Round 5) was computed for each of the four output variables as the absolute difference between the target value and the obtained value. As this measure produced a skewed distribution, the variance was corrected by applying a logarithmic transformation. Some participants reached the goal states exactly, but as zero cannot be transformed logarithmically, one was added to the error scores of every participant. Solution errors were computed for each of the six trials that comprised Round 5, in order to determine how quickly participants were able to approach the target goal. In most analyses, a single solution error score was used, obtained by summing the error for the four output variables and taking the mean over the six trials of the solution round. For analyses of variance, however, the individual error scores for each trial and output variable were included as factors.

(3) ***Prediction score.*** The third dependent variable was the number of correct predictions (out of 10) regarding the influence of specified inputs on specified outputs.

Development of Strategies

Figure 2 depicts the percentage of participants classified as using each of the three types of strategies for each of the four learning rounds. In Round 1, most of the participants (67%) used the unsystematic strategy of changing all four input variables at once (CA), whereas just 19% began by using VOTAT. By Round 2, 47% of the participants used CA, and another 47% used VOTAT. By Round 4, only 22% still used CA, whereas 56% tested their hypotheses with VOTAT. The percentage of participants using HT strategies remained relatively low across the four rounds, never exceeding

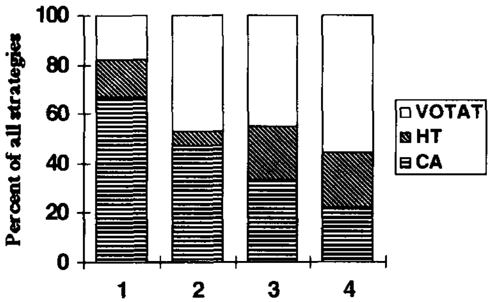

Figure 2. Percentage of participants using each of the three types of strategies over the four learning rounds of Experiment 1.

22%. The trend towards use of VOTAT is very clear. Between Rounds 1 and 4, 15 participants (42% of total sample) changed between VOTAT and one of the other strategies, and all except 1 of these participants changed to VOTAT rather than away from it.

As a measure of the extent to which participants used systematic strategies over the course of the learning phase, a *strategy systematicity score* was calculated by assigning to each round a score of 0 for use of the CA strategy for the round, a score of 1 for using a HT strategy, and a score of 2 for using VOTAT. The means of these scores increased monotonically over the four rounds (Round 1: 0.53 [$SD = 0.81$]; Round 2: 1.00 [$SD = 0.98$]; Round 3: 1.28 [$SD = 0.85$]; Round 4: 1.33 [$SD = 0.82$]), showing a significant linear trend, $F(1, 35) = 23.00$, $p < .001$, and a quadratic trend, $F(1, 35) = 4.49$, $p < .05$. An alpha level of .05 was used for all statistic tests reported in this article.

The observed pattern of strategy development thus indicates that participants spontaneously progressed to more systematic strategies over the course of the learning phase. Only 1 or 2 participants per round regressed to a less systematic strategy, whereas many participants eventually shifted from using CA to using VOTAT.

Analyses of Structure Scores

Participants' structure scores increased monotonically across the four learning rounds: $M = 0.88$, $SD = 0.87$; $M = 1.40$, $SD = 0.98$; $M = 1.78$, $SD = 0.82$; and, $M = 1.79$, $SD = 0.81$; respectively, for Rounds 1-4. Across the four rounds there was a significant linear trend in scores, $F(1, 35) = 40.79$, $p < .001$, but also a significant quadratic trend, $F(1, 35) = 4.60$, $p < .05$. Scores appeared to improve at a diminishing rate across rounds, with minimal improvement between the final two learning rounds. It thus appears that three rounds were sufficient for participants to acquire most of what they were capable of learning about the system.

We analyzed the relationship between strategy in Round 4 and the structure score for Round 4, because participants' knowledge at the end of the learning phase would be expected to have the closest relationship to subsequent problem-solving performance. Final structure scores differed across the three strategy groups, $F(2, 33) = 19.03$, $p < .001$. Pairwise Newman-Keuls tests indicated that participants who used VOTAT ($M = 2.25$, $SD = 0.47$) had higher structure scores than those who used HT ($M = 1.64$, $SD = 0.76$) strategies, who had higher structure scores than did those who used CA ($M = 0.78$, $SD = 0.59$). Another measure of this relationship is provided by the high correlation between the summed strategy systematicity scores (summed over the four rounds) and the structure score, $r = .76$, $p < .001$, indicating that the greater the extent to which participants used a systematic strategy, the better their knowledge of structure. Thus, greater systematicity of strategies was associated with acquisition of more accurate knowledge of the problem space.

Analyses of Solution Error

Solution error tended to be higher for participants who used CA ($M = 4.80$, $SD = 0.42$) than for those who used HT strategies ($M = 4.11$, $SD = 1.34$) or VOTAT ($M = 3.87$, $SD = 1.47$), although not significantly, $F(2, 33) = 1.49$, $p > .10$. However, the overall extent to which participants' used systematic strategies was associated with problem-solving success, as the strategy systematicity score correlated negatively with solution error, $r = -.32$, $p < .05$.

Collapsing across the strategy variation, participants' solution errors were not correlated with structure scores, $r = -.05$, $p > .25$. However, examination of the scatter-plot revealed 3 clear outliers who had the three highest solution errors, but reasonable structure scores. When data from these 3 participants were eliminated, a significant negative correlation of solution error with final structure scores was obtained, $r = -.54$, $p < .01$.

Analyses of Prediction Scores

Strategy variations in Round 4 had a pronounced relationship with performance on the prediction task that followed the solution round. The mean number of correct predictions (out of 10) was 4.40 ($SD = 3.00$) for participants using VOTAT, $M = 2.75$, ($SD = 2.92$) for those using HT strategies, and none at all for those using CA, $F(2, 33) = 8.00$, $p < .01$. Pairwise Newman-Keuls tests indicated that participants using CA had significantly fewer correct predictions than those using VOTAT or HT strategies. Consistent with this finding, strategy systematicity scores were positively correlated with prediction scores, $r = .43$, $p < .01$. In addition, participants' prediction scores were positively correlated with their final structure scores, $r = .82$, $p < .001$.

In summary, the results of Experiment 1 revealed that the systematicity of participants' spontaneous hypothesis-testing strategies predicted their success in learning the structure of the biology-lab problem space. Those who used the least systematic strategy, CA, were significantly impaired in providing links and weights in the diagram task (structure score) and in making predictions about outputs for new input levels (prediction score). Overall, use of more systematic strategies tended to be associated with lower solution errors in the solution round, in which participants were asked to generate a specific goal state for the output variables. The results of Experiment 1 served to validate the diagram task as a measure of participants' knowledge of the problem space, and provided preliminary evidence that learning is influenced by the systematicity of problem-solving strategies. Although the strategy classification was somewhat imprecise, most participants eventually adopted the most systematic strategy, VOTAT, and the more they used it, the better they tended to perform. However, even those using the VOTAT strategy generally did not solve the problem completely, as the mean number of correctly identified weights (uncorrected for guessing) for such participants was only 2.0 out of 6.0. Participants clearly find the biology-lab task to be difficult.

Experiment 1 provided evidence that VOTAT is an effective strategy for learning about the biology-lab system, and that participants come to favor this strategy. This result provided the foundation for Experiment 2, in which we sought to more precisely test the hypothesized linkage between strategy and learning by directly manipulating participants' strategies. Experiment 1 was limited in that the learning phase always involved a nonspecific goal, which would be expected to maximize use of systematic strategies. Experiment 2 was performed to assess learning and transfer while manipulating both systematicity of strategies and goal specificity.

EXPERIMENT 2

Experiment 1 established that the biology-lab task is suitable for investigating the relationship between strategies for hypothesis testing and subsequent problem-solving performance. Whereas Experiment 1 examined spontaneous strategy use, Experiment 2 manipulated the strategies that participants used both by direct instruction and by varying the specificity of their goals during the learning phase. If Sweller (1988) was correct in his analysis of the impact of goal specificity, we would expect a specific goal to encourage use of a goal-oriented strategy. Such a strategy can be expected to yield good performance in solving a problem based on that particular goal, but relatively poor knowledge of the overall structure of the problem space. In contrast, a nonspecific goal will encourage use of systematic strategies such as VOTAT, which was shown in Experiment 1 to be an effective strategy for discovering the structure of a task. Therefore, a nonspecific goal should lead to greater knowledge of the rule space and hence superior transfer performance. Our test of transfer was a task with the same structure, but different goal. Although this may appear to be a trivial transfer task, it will only be trivial if the participants' knowledge is of the structure of the task. If, as a consequence of trying to reach a specific goal, what participants learn is specific to that goal, then transfer performance to a new goal should be degraded. In addition, if participants often fail to use VOTAT spontaneously (as was the case in Experiment 1), then instruction in its use may be necessary to achieve superior learning from a free exploration phase.

Method

Participants

Sixty undergraduate students (22 female, 38 male) at the University of California, Los Angeles, participated for course credit. Data were discarded from an additional 5 participants (distributed across all four conditions) who told the experimenter that they had no idea how to learn anything about the system.

Design and Procedure

The experiment included four between-subject conditions, defined by the factorial combination of two levels of goal specificity (specific vs. nonspecific) and two levels of strategy instruction (instruction to use VOTAT vs. no instruction). Fifteen participants served in each condition. As in Experiment 1, participants received an initial learning phase (three rounds, rather than four as in Experiment 1) followed by a solution round (Round 4) in which they were asked to produce a specific goal state (namely, 50 crabs, 400 prawns, 900 lobsters, and 700 sea bass). A fifth round was then provided in which all participants were asked to solve an additional transfer problem (reaching the goal state of 250 crabs, 200 prawns, 1,000 lobsters, and 350 sea bass).

Before starting to manipulate the system on the computer, all participants received general instructions about the task. All participants were told that they should explore the system so as to learn as much as possible, but the four groups of participants were distinguished by instructional manipulations of how they approached the initial learning phase. Participants in the nonspecific goal groups were not given any specific goal until Round 4. In Rounds 1 through 3, these participants were simply asked to set inputs and observe outputs in order to figure out how the system works, just as was the case for participants in Experiment 1. In particular, participants in the specific-goal groups were informed of the goal for the solution round from the outset of Round 1; thus, they were exposed to the goal for four rounds, although they too were told that they should explore the system so as to learn as much as possible.

Participants in the strategy-instructed groups were given written instructions explaining that the optimal strategy (VOTAT) was to vary just one variable at a time, setting the remaining variables to zero. The VOTAT strategy was first explained using an example from science: If scientists wish to test different medicines that may cure a disease, they do not give all the medicines to each patient; rather, they test one medicine at a time to determine the separate influence of each. The strategy was then illustrated with an example from the biology-lab task: To learn about the effect of temperature, it would be necessary to manipulate this input variable in isolation. Strategy-instructed participants were directed to use the strategy described to them in exploring the system. In contrast, the strategy-uninstructed groups received no advice on how to explore the system, although they were told that during the first three rounds their aim was to explore the system.

The identical biology-lab task as had been used in Experiment 1 was also used in Experiment 2 (see Figure 1). After each round of the learning phase (Rounds 1-3), participants completed a structure diagram in the same manner and with the same instructions as in Experiment 1. In the solution round (Round 4), all participants were presented with a specific goal state, which was the same as the specific-goal groups had had throughout the learning phase. In the transfer round (Round 5), all participants were asked to achieve a different goal state, one which was new to participants in all conditions. Performance on this new goal provided a measure of the degree to which learning over Rounds 1 through 4 yielded transfer to a novel problem drawn from the same problem space. Time to complete each of the five rounds was recorded. However, response times of Round 1 are not meaningful as they include the time used to explain the task before participants started to manipulate the system. Finally, a prediction task similar to that used in Experiment 1 was administered to all participants. As in Experiment 1, 10 questions were used, such as, "If you have 100 crabs in your system and you set oxygen = 20, how many crabs do you have afterwards?" However, rather than asking for an open-ended answer as in Experiment 1, a four-alternative multiple-choice format was used. For this example, the choices were 40, 60, 140, and 180. The entire experiment took 1 hour to complete.

Results and Discussion

Classification of Learning Strategies

We examined participants' patterns of settings for the four inputs during the initial learning phase in order to determine their strategy on each round. Participants were coded as using the strategies of VOTAT, CA, and HT, using the same criteria as in Experiment 1. In addition, participants given a specific goal could also be coded as using a difference-reduction (DR) strategy, a goal-oriented strategy that focused directly on attaining the specific goal state. Participants were classified as using the DR strategy when two criteria were met during a round: (a) at least one of the four output states for the specific goal was reached; and (b) for at least one such input, on four out of the six trials, participants generated incrementally closer approximations to the goal state. The DR category did not arise in Experiment 1, because that study did not introduce a specific goal during the learning phase.

Dependent Variables

Five dependent variables were analyzed to provide evidence of learning and transfer.

1. *Structure score.* The structure score was computed as in Experiment 1. Only the score for the final learning round (Round 3) was used in the reported analyses.
2. *Solution error.* Solution error in round 4 was computed as in Experiment 1 (i.e., mean over six trials for the summed log-transformed absolute errors for the four output variables). Also as in Experiment 1, trial and output variables were included as factors in analyses of variance.
3. *Transfer error.* Transfer error was computed for the novel problem introduced in Round 5 in the same manner as solution error in Round 4.
4. *Prediction score.* For each of the 10 multiple-choice questions, participants received 2 points for selecting the correct answer and 1 point for selecting an alternative consistent with the correct sign of the relevant weight. The sum out of 20 points constituted the prediction score.
5. *Response time.* The time required to complete each round was analyzed separately for the three phases (learning, solution, transfer).

Analyses of Learning Strategies

Analyses were performed to assess whether our manipulation of learning strategy by instructions was successful. Figure 3A shows how the goal conditions influenced the percentage of strategy-instructed participants using each strategy on each round. Eighty percent of all participants in the strategy-instructed conditions followed the VOTAT strategy in the first round. Most strategy-instructed participants in the nonspecific-goal condition continued with the VOTAT strategy through Round 3. However, those strategy-instructed participants who were in the specific-goal condition exhibited a strong tendency to switch from the VOTAT strategy to a DR strategy which focuses directly on reaching the stated goal.

Figure 3B shows the comparable strategy analysis for strategy-uninstructed participants. In the absence of strategy instructions, most participants did not spontaneously use the VOTAT strategy. If the specific-goal group and the nonspecific-goal group in Figure 3B are compared, it is apparent that giving participants a goal to reach did have an effect on strategies. Many

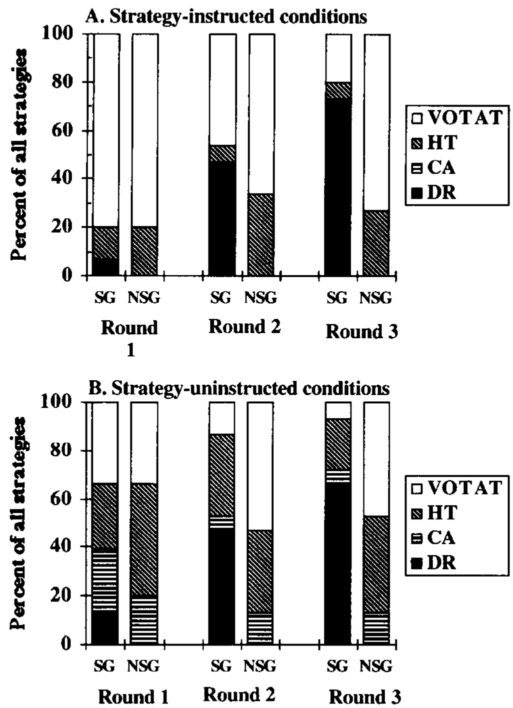

Figure 3. Percentage of strategy-instructed (Panel A) and strategy-uninstructed (Panel B) particiants using alternative strategies over the three initial rounds in Experiment 2. (SG = specific goal, NSG = nonspecific goal).

participants with a specific goal either never used VOTAT or switched to DR even if instructed on VOTAT.

For participants with the nonspecific-goal it was again possible to calculate strategy systematicity scores in the same way as in Experiment 1. Strategy-instructed participants ($M = 1.73$, $SD = 0.34$) had higher scores than strategy-uninstructed participants ($M = 1.31$, $SD = 0.66$), $F(1, 28) = 4.86$, $p < .05$. There was no significant effect of round, $F(2, 56) = 0.07$, nor any interaction between round and strategy instruction, $F(2, 56) = 1.33$, $p > .25$.

The strategy systematicity score could not be used for goal participants as the option of using a DR strategy had been introduced. (By definition, nonspecific-goal participants could not use DR, and it is extremely unlikely that the criteria for DR could be met randomly.) Although DR is a systematic strategy, it is of a completely different type than VOTAT. Therefore, to analyze specific-goal participants we gave them a score of 1 for each round if they used VOTAT or 0 if they did not. Mean scores over the three rounds were then calculated. While VOTAT was used by strategy-instructed participants ($M = .49$, $SD = 0.33$) more than by strategy-uninstructed participants

($M = 0.20$, $SD = 0.28$), $F(1, 28) = 6.76$, $p < .025$, these strategy-instructed participants had a strong linear trend away from use of VOTAT, $F(1, 14) = 21.00$, $p < .001$. The strategy-uninstructed participants, who were much less likely to start by using VOTAT, also showed a linear trend away from VOTAT, but this was only marginally significant, $F(1, 14) = 3.50$, $p < .10$. By scoring DR as 1 (used) or 0 (not used), both specific-goal groups displayed strong linear trends towards increased use of DR: strategy-instructed, $F(1, 14) = 28.00$, $p < .001$; strategy-uninstructed, $F(1, 14) = 7.98$, $p < .025$.

It thus appears that our strategy instruction was indeed effective in promoting use of the VOTAT strategy, but that providing a specific goal created a strong pressure to employ a DR strategy, even for those told to use VOTAT.

Relationship Between Structure Scores and Performance

Analyses were performed to determine whether the measure of participants' representations based on structure scores for the diagram-completion task predicted success on the various problem-solving and prediction tasks. If the structure score derived from their structure diagrams after the final round of initial learning (Round 3) provides a valid assessment of what they had learned about the system, then the structure score would be expected to correlate inversely with solution error measured on Round 4 and transfer error on Round 5. This was indeed the case. Participants with higher structure scores produced lower solution error when they had to reach the goal state in Round 4, $r = -.57$, $p < .001$, as well as lower transfer error in Round 5, $r = -.65$, $p < .001$. In addition, structure scores were correlated with degree of success on the prediction task, $r = .62$, $p < .001$. These results confirm the comparable findings from Experiment 1, again validating our structure score as a measure of the quality of participants' knowledge of the rule space.

Influence of Goal Specificity and Strategy on Learning and Transfer

Participants achieved higher structure scores when given a nonspecific-goal ($M = 2.35$, $SD = 0.63$) rather than a specific goal ($M = 1.69$, $SD = 0.71$), $F(1, 56) = 15.50$, $p < .001$, and when given instruction in the VOTAT strategy ($M = 2.22$, $SD = 0.70$) rather than no strategy instruction ($M = 1.82$, $SD = 0.76$), $F(1, 56) = 5.87$, $p < .025$. The interaction between goal specificity and strategy instruction as determinants of structure score was not significant. For solution error scores a $2 \times 2 \times 4 \times 6$ repeated measures ANOVA was run with goal specificity and strategy instruction as between-subjects factors, and output variable and trial as within-subject factors. As summarized in Table 1, participants instructed in the use of the VOTAT strategy achieved marginally lower solution error than uninstructed participants ($M = 2.28$, $SD = 1.40$; and $M = 2.81$, $SD = 1.02$; respectively) in the solution round, $F(1, 56) = 3.23$, $p < .10$. Solution error did not differ as a function of goal specificity, $F(1, 56) = 0.12$, and there were no significant interactions. Although participants in

TABLE 1
Means and Standard Deviations (in Parentheses)
for Each Learning Condition on Each Dependent Variable in Experiment 2

	Structure Score	Solution Error	Transfer Error	Predictions
Uninstructed/Specific Goal	1.53 (0.73)	2.93 (0.96)	3.15 (1.28)	13.00 (3.53)
Uninstructed/Nonspecific Goal	2.10 (0.70)	2.70 (1.09)	2.47 (1.13)	16.13 (2.56)
Instructed/Specific Goal	1.84 (0.70)	2.26 (1.14)	2.46 (1.23)	15.07 (3.33)
Instructed/Nonspecific Goal	2.61 (0.46)	2.29 (1.40)	1.83 (1.17)	15.87 (4.47)

the nonspecific-goal condition achieved greater overall knowledge of the system structure, those in the specific-goal condition had three additional rounds of practive in attaining the goal set for all participants in Round 4. Both possible approaches to this task, learning the rules governing how the system works or concentrating on finding a path to the specific goal, should help participants reach the goal. Which is the most effective may depend on the specific system. Therefore, it is conceivable that even if participants were using quite different approaches, both approaches could lead to equivalent performance in reaching the goal.

The most crucial results concern transfer performance on Round 5, when a goal that was novel to all participants was introduced. These results are shown in Table 1. A similar ANOVA to that used for solution error was used for transfer error and revealed the same pattern as that observed for the structure score: Strategy-instructed participants could better solve the transfer round than uninstructed participants, $F(1, 56) = 4.67$, $p < .05$, and nonspecific-goal participants could better reach the goal in the transfer round than those with a specific goal, $F(1, 56) = 4.43$, $p < .05$. None of the interactions was significant. To confirm the differences between the effects of the manipulations on the solution round and the transfer round, we ran a similar repeated measures analysis to that run for solution and transfer rounds individually, but now with round (solution vs.transfer) as an additional within-subject factor. This analysis revealed an interaction between goal specificity and round, $F(1, 56) = 11.00$, $p < .005$, whereas there was no interaction between strategy instruction and round, $F(1, 56) = 0.65$. From the means (Table 1), it is evident that participants with a goal during the learning rounds performed well in the solution round but deteriorated in the transfer round, whereas participants without a specific goal improved slightly in the transfer round. Thus, although participants given a specific goal (who predominantly used a DR strategy) were able to effectively achieve that specific goal, they were relatively poor in transferring their knowledge to a similar problem with a new goal.

We also examined the effects of goal specificity and strategies on response time for the initial learning phase (Rounds 2 and 3 combined; response times from Round 1 could not be analyzed because they included the time used to explain the task before the start of the round) and the solution phase (Round 4). The strategy manipulation did not significantly influence response time in either phase, either as a main effect or as an interaction. However, goal specificity influenced response time in different ways in each phase, $F(1, 56) = 20.54$, $p < .001$. During the learning phase, participants given a specific goal took more time to manipulate the input variables ($M = 421$s/round, $SD = 131$) than did those with a nonspecific goal ($M = 341$ s/round, $SD = 148$), $F(1, 56) = 4.98$, $p < .05$. The longer times associated with having a specific goal are consistent with Sweller's (1988) claim that goal-oriented strategies are more cognitively demanding than are strategies used for inducing rules. This response-time pattern also demonstrates that the superior transfer performance observed for participants in the nonspecific-goal condition who received VOTAT instructions cannot be attributed to greater initial study time.

In the solution phase, on the other hand, solution time was significantly faster for participants in the specific-goal condition ($M = 396$ s, $SD = 171$) than for those in the nonspecific-goal condition ($M = 598$ s, $SD = 238$), $F(1, 56) = 14.48$, $p < .001$. This difference is consistent with the fact that the former group was solving the same problem as they had received in Rounds 1 through 3, whereas the latter group was receiving it for the first time. The two goal-specificity conditions did not differ significantly in response time on the transfer problem in Round 5, $F(1, 56) = 1.86$, $p > .10$, which was novel for all participants ($M = 394$ s, $SD = 154$, for the specific-goal condition, $M = 445$ s, $SD = 135$, for the nonspecific-goal condition).

Participants with a nonspecific goal obtained higher scores on the prediction task than did those who had a specific goal (see Table 1), $F(1, 56) = 4.64$, $p < .05$. However, neither the main effect of strategy instruction nor the interaction between strategy and goal specificity approached significance. A possible explanation is that the multiple-choice task was too easy, making it relatively insensitive to differences between groups in knowledge.

GENERAL DISCUSSION

The aim of this study was to test alternative theories of the relationship between problem solving and acquisition of rules. The biology-lab, a complex dynamic system involving multiple input variables that must be manipulated to control multiple output variables, provided a rich environment in which to explore the influence of goal specificity and hypothesis-testing strategies on learning and transfer. The results of Experiment 1 indicate that people who spontaneously adopt more systematic strategies for rule induction during free exploration acquire more complete knowledge of the rule space and are more successful at subsequent tasks that tap this knowledge.

Experiment 2 manipulated both learning strategies and goal specificity during the learning phase. We found that providing participants with a specific goal from the outset of learning produced a strong tendency to use a goal-oriented strategy. The predominant strategy for such participants was a difference-reduction strategy that focused on incrementally reducing the difference between obtained outputs and the specific goal. This strategy was adequate for eventually solving the particular goal but was suboptimal as a vehicle for discovering the overall structure of the system. As a result, provision of a specific goal impaired eventual transfer to a new problem drawn from the same problem space but involving a different goal state.

Acquisition of the structure of the system was fostered both by using a nonspecific goal and by providing explicit instruction in a systematic strategy, VOTAT, which involves varying a single factor while holding other factors constant at zero. However, participants who were given a specific goal tended to abandon the VOTAT strategy over the course of the learning session, shifting to a goal-oriented strategy. Participants who were not taught the VOTAT strategy tended to use either a goal-oriented strategy (if a specific goal was provided) or some other suboptimal strategy (if no specific goal was provided). Thus, optimal transfer performance required a combination of a nonspecific goal coupled with instruction in use of a systematic strategy.

We have assumed that participants given a nonspecific goal tried to induce the rules governing the behavior of the biology-lab. Our results tend to support this assumption, as participants with nonspecific goals learned more about the structure of the system; however, we cannot be sure that they actually set themselves goals such as learning the rules. Future studies using protocol analysis might provide more direct evidence regarding people's goals, and whether they explicitly explore both instance and rule space.

Our results run counter to theories of skill acquisition that stress the importance of learning from weak problem-solving methods as a means of inducing general rules (e.g., Anderson, 1987; Larkin, 1981). It is certainly possible that people sometimes learn general rules in the aftermath of solving problems by variants of means-ends analysis or other goal-oriented strategies; however, at least in the absence of prior knowledge of the domain, this approach does not appear to provide an optimal path toward either general knowledge of the stucture of a complex system or successful transfer to problems with an altered goal. Rather, acquisition of system structure is fostered to a greater extent by free exploration of the problem space.

It should be noted that the strategy used by many of our specific-goal participants, although goal directed, did not meet the technical definition of means-ends analysis (i.e., removing the largest difference between the current state and goal state, in the process recursively solving the subproblem of getting from the current state to that which satisfied the preconditions of required operators). Thus, our results do not directly show that the full means-ends strategy would fail to promote learning of overall problem structure. Nonetheless, the present strategy did involve difference reduction (i.e., search in which each step progresses closer to the specified goal), which is a major component of means-ends analysis. Our theoretical analysis suggests that the key factor limiting acquisition of overall structure is focus on a specific goal, which tends to encourage search in the space of instances rather than the spaces of both rules and instances. It follows that as long as a specific goal is given, even full means-ends analysis should prove relatively ineffective in promoting learning. However, further research will be required to test this possibility.

Another caveat concerning these findings relates to the fact that our study used a problem domain in which our participants were complete novices. A different pattern of results might emerge in a problem domain for which participants have a prior theory of the domain. In a more knowledge-rich domain, mechanisms of explanation-based learning (e.g., DeJong, 1986; Mitchell, Keller, & Kedar-Cabelli, 1986) might allow people to form generalizations of solutions initially obtained by weak methods, such as means-ends analysis. One direction for future work would involve manipulating domain knowledge together with participants' learning strategies and examining transfer performance in the aftermath of initial problem solving.

These results are broadly in agreement with the findings of Sweller and his colleagues (Sweller, 1988; Mawer & Sweller, 1982), who also found that reduced goal specificity yields better performance. Our findings do not directly address Sweller's (1988) explanation for these results, which is that participants perform more poorly with specific goals because having to monitor goals increases the cognitive load of the task, thus reducing the capacity available for deriving rules. The timing data from Experiment 2 is consistent with Sweller's claim, in that a specific goal increased response times during the learning phase. However, other aspects of the present findings are less consistent with a capacity explanation. Participants in Experiment 2 who were given a specific goal and were instructed to use VOTAT tended to initially use VOTAT, but later to change to DR. According to Sweller's assumptions, such a change implies that participants were voluntarily choosing a task with a higher load, which seems implausible. Sweller's model appears to predict that if the overall processing load on participants is increased (e.g., by increasing complexity of the system to be learned, or by adding a stressor), they should be more likely to use goal-oriented strategies, which are presumed to require less capacity. In contrast, a dual-space model appears to make the opposite prediction, assuming that representing two spaces should tend to place a higher load on the individual than representing just the instance space.

The present study increases the generality of Sweller's (1988) goal-specificity results by demonstrating similar phenomena in the domain of a complex dynamic system, as opposed to the static mathematical domains primarily used in earlier studies. Whereas Sweller and colleagues have demonstrated the effect of goal specificity on knowledge using relatively indirect performance measures (e.g., response times), we were able to obtain a more direct measure of participants' knowledge (structure diagrams). With this method it was possible to provide evidence that the influence of the goal and strategy manipulations on performance was mediated by variations in the quality of participants' knowledge of the system. In addition, this study goes beyond previous work in identifying the relationship between hypothesis-testing strategy and the impact of reduced goal specificity. In a complex task environment such as the biology-lab, college students are not generally prepared to spontaneously make full use of an effective rule-induction strategy, even when they are given a nonspecific goal. It is therefore important to provide instruction in the use of such a strategy in order to allow maximum benefit from free exploration of the problem space. It is not enough to simply "wander" through a haphazard series of input-output relations; rather, effective learning depends on systematic investigation of controlled variations in the inputs. Our results, thus, have important educational implications for designing effective techniques for encouraging problem-based learning in complex domains.

Finally, our results suggest explanations for other anomalous findings concerning learning and reasoning, which may be viewed as resulting from changes to the goals of the participants. For example, Berry and Broadbent (1984, 1987) examined complex problem solving and found only a weak or even negative relationship between knowledge and performance. They interpreted this independence of explicit knowledge and performance as demonstrating that implicit (unconscious) knowledge was gained. However, in these studies, all participants had specific goals from the beginning. In experiments in which participants had no specific goal during the training phase (e.g., Funke & Müller, 1988), the relationship between knowledge and performance was positive. Thus, goal specificity could be the factor that is responsible for the inconsistent results.

Experiments using Wason's (1966) 2-4-6 task have found that participants tend to generate instances that would confirm the rule they have in mind (Klayman & Ha, 1987). However, Tweney et al. (1980) found that performance can be improved if instead of telling participants that a number triple conforms to the rule, an instance is labeled "DAX" if it fits the rule and "MED" if it does not. In our terms, this manipulation can be interpreted as involving a change of the participants' goals. Participants told that an instance was "right" or "wrong" must first generate a rule; they then appear to set themselves the goal of confirming their rule, and thus spend most of their time searching instance space. However, DAX/MED participants may accept a less specific goal, that of finding a rule that distinguishes the two labels, and therefore focus on searching rule space. This increased focus on searching rule space may increase their success in finding the DAX rule. More generally, goal specificity and learning strategies may have important effects on many tasks involving reasoning, problem solving, and learning.

A Dual-Space Interpretation

Our results are consistent with our theoretical framework that problem solving and hypothesis testing require search of a dual problem space: the space of rules, and that of instances. When participants were not given a specific goal, they appeared to be more likely to search rule space; if they had a specific goal, they appeared to be more likely to focus on search of instance space. Specific goals thus appeared to encourage use of non-rule-induction strategies, whereas nonspecific goals encouraged rule induction strategies that are efficient for searching rule space.

Systematic rule-induction strategies may have an impact beyond allowing more efficient search of rule space: They may alter what is learned, not just how much is learned. Much like provision of a nonspecific goal, a systematic strategy may be important in encouraging the learner to venture into rule space; moreover, by facilitating successful induction, such a strategy may lead learners to continue to search rule space. This interpretation is consistent with our finding that providing participants with a systematic strategy had very similar effects on performance as did specificity of goals. Both manipulations fostered acquisition of more knowledge of a task's structure, which could explain the superior performance on transfer tasks by strategy-instructed and nonspecific-goal groups. In addition, a systematic strategy appeared to help learners whose primary focus was on reaching a specific goal, perhaps by facilitating whatever search of rule space they may have performed. Even when learners have a specific goal, it is likely that they sometimes try to discover rules, especially when they first encounter an unfamiliar problem for which they initially lack adequate knowledge of how operators affect states. However, specific-goal participants showed a strong tendency towards use of non-rule-induction strategies, as the desire to reach a specific goal may have kept enticing such learners into instance space, which is where their specific goal was located. In contrast, participants in Experiment 1 with a nonspecific goal who were allowed to freely choose a strategy, spontaneously came to adopt the use of VOTAT, which appears to be a strategy effective for rule induction. The effects of goal specificity on problem solving may therefore be mediated by the differential strategy use encouraged by different types of goals.

REFERENCES

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

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

Berry, D., & Broadbent, D.E. (1987). The combination of implicit and explicit knowledge in task control. *Psychological Research, 49,* 7–15.

Branke, S. (1991). *Die Interaktion von Personen-und Aufgabenmerkmalen beim komplexen Problemlösen* [The interaction of individual differences and task in complex problem solving] (Report No. 19). Mainz, Germany: Department of Psychology, Johannes Gutenberg-University.

DeJong, G.F. (1986). An approach to learning from observation. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (Vol. 2). Los Altos, CA: Morgan Kaufmann.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science. 17,* 397–434.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R.J. Sternberg & P.A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms.* Hillsdale, NJ: Erlbaum.

Funke, J., & Müller, H. (1988). Eingreifen and Prognostizieren als Determinanten von Systemidentifikation und Systemsteuerung [Intervention and prediction as determinants of systems identification and system control]. *Sprache & Kognition. 7,* 176–186.

Holyoak, K.J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits.* Cambridge: Cambridge University Press.

Klahr, D., & Dunbar, K. (1988) Dual space search during scientific reasoning. *Cognitive Science. 12,* 1–55.

Klahr, D., Fay, A.L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25,* 111–146.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychology Review, 94,* 211–228.

Larkin, J.H. (1981). Enriching formal knowledge: A model for learning to solve textbook physics problems. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Erlbaum.

Mawer, R.F., & Sweller, J. (1982). Effects of subgoal density and location on learning during problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 252–259.

Mitchell, T.M., Kelle, T., & Kedar-Cabelli, S. (1986). Explanation-based generalization: A unifying view. *Machine Learning, 1,* 47–80.

Putz-Osterloh, W. (1993). Strategies for knowledge acquisition and transfer of knowledge in dynamic tasks. In G. Strube & K.F. Wender (Eds), *The cognitive psychology of knowledge.* Amsterdam: Elsevier.

Simon, H.A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L.W. Gregg (Ed.), *Knowledge and cognition.* Hillsdale, NJ: Erlbaum.

Sweller, J. (1983). Control mechanisms in problem solving. *Memory & Cognition, 11,* 32–40.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257–285.

Sweller, J., & Levine, M. (1982). Effects of goal-specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 463–474.

Sweller, J., Mawer, R.F., & Howe, W. (1982). Consequences of history-cued and means-end strategies in problem solving. *American Journal of Psychology, 95,* 455–483.

Sweller, J., Mawer, R.F., & Ward, M.R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112,* 463–474.

Tschirgi, J.E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51,* 1–10.

Tweney, R.D., Doherty, M.E., Worner, W.J., Pliske, D.B., Mynatt, C.R., Gross, K.A., & Arrkelin, D.L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology, 32,* 109–123.

Wason, P.C. (1966). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140.

Woodworth, R.S., & Schlosberg, H. (1954). *Experimental psychology,* New York: Holt, Rinehart, and Winston.