

Running head: GOAL SPECIFICITY EFFECTS ON HYPOTHESIS TESTING

Goal specificity effects on hypothesis testing in problem solving

Bruce D. Burns

Michigan State University, USA

Regina Vollmeyer

Universität Potsdam, Germany

Address for correspondence:

Bruce Burns

Department of Psychology

Michigan State University

East Lansing, MI 48824-1117

USA

e-mail: burnsbr@msu.edu

Appeared in *The Quarterly Journal of Experimental Psychology*, 55A, 241-261, 2002

Abstract

Previous research has found that having a nonspecific goal (NSG) leads to better problem solving and transfer than a specific goal (SG). To distinguish between the various explanations of this effect requires direct evidence showing *how* a NSG affects a participant's behavior. Therefore we collected verbal protocols from participants learning to control a linear system consisting of 3 outputs by manipulating 3 inputs. This system was simpler than the one we had used previously, so in Exp. 1 we generalized our earlier goal specificity findings to this system. In Exp. 2 protocol analysis confirmed our prediction (based on dual-space theories of problem solving) that NSG participants focussed on hypothesis testing whereas SG participants focussed on the goal. However, this difference only emerged over time. We also replicated the goal specificity effect on performance and showed that giving participants a hypothesis to test improved performance.

Goal specificity effects on hypothesis testing in problem solving

Effects of goal specificity on problem solving have been found in a number of recent studies (Geddes & Stevenson, 1997; Miller, Lehman, & Koedinger, 1999; Sweller, 1988; Vollmeyer, Burns, & Holyoak, 1996). All these studies have shown that giving problem solvers a specific goal state to reach (e.g., bring a system to a certain value, or put a puck into a net) led to poorer learning of the task than if they were given a nonspecific goal, such as to explore and find the underlying rule. However, the question arises of why exactly does goal specificity affect learning from problem solving tasks? What is the process?

Sweller (1988) proposed that a specific goal places a greater cognitive load on the problem solver. He assumed that the goal evokes a means-ends strategy, which requires keeping subgoals and current states in working memory. This increased cognitive load produces poorer problem solving performance and less learning because it was assumed that both learning and subgoal management use cognitive resources from a limited pool.

Miller et al. (1999) tested Sweller's (1988) cognitive load theory using a learning task called Electric Field Hockey. Participants tried to push a puck between obstacles and into a net, but initially they were given either a *specific goal* (i.e., participants could see the net and obstacles) or a *nonspecific goal* (i.e., the net and the obstacles were removed) requiring participants to try to understand the game's properties. In addition, Miller et al. included a third condition: an *appropriate specific goal* which would increase cognitive load but which they predicted would help performance. This specific goal was still to put the puck into the net but to do so by following a specific trajectory path. Thus means-ends analysis could be employed by comparing the current trajectory path to the goal path, although according to Sweller (1988) this would increase cognitive load and thus lead to poorer performance. However, the specific-path condition led to as good performance as the nonspecific goal group and both were better than the

specific goal group not shown the possible trajectory. Miller et al. argued that it was not having a specific goal that inhibited performance in Sweller's tasks, but having an *inappropriate* specific goal. A similar argument has been made by Catrambone (1998) who found that giving the right math subgoals led to better math learning than simply providing the equation to be applied.

Miller et al.'s (1999) goal appropriateness theory predicted their result not despite the use of means-ends analysis, but because of it. They derived their predictions from a computer model using Soar (Newell, 1990), which is based on learning through means-ends analysis. Within this model, specific goals can lead to poor learning if they are inappropriate, whereas nonspecific goals leave open the possibility of the learners setting appropriate goals for themselves. This model explicitly does not incorporate hypothesis testing, so although Miller et al. (p. 310) commented that learners may use deliberate hypothesis testing, they regard whether a learner uses such a strategy as independent of the learner's goal.

Vollmeyer et al. (1996) found indirect evidence that goal specificity did change the problem solvers' strategies. In contrast to the focus of Sweller (1988) and Miller et al. (1999) on means-ends analysis, Vollmeyer et al. proposed that the better performance of learners given a nonspecific goal is due to increased use of hypothesis testing. Vollmeyer et al. gave solvers a complex system to explore and, when asked, to bring to a certain goal state. It was found that solvers who had been told the goal state at the beginning (i.e., they were given a specific goal) tended to change all the inputs at the same time, which is appropriate for reaching a goal immediately but not for learning the system's structure. In contrast, the problem solvers who were told to explore, but not told the goal state until when they actually had to reach it (i.e., given a nonspecific goal) tended to change one input variable at a time, which was the most effective strategy for inducing the rules governing the task. Vollmeyer et al. saw this result in terms of dual-space theories of search in problem solving (Klahr & Dunbar, 1988; Simon & Lea,

1974), which combines problem solving and induction by proposing that there is a coordinated search of hypothesis (or rule) space and instance space. (Below, dual space theories are further explained.) However, as these forms of evidence are indirect, we had to infer that these participants were using hypothesis testing.

Geddes and Stevenson (1997) also used the idea of dual-space search to predict the effects of goal specificity on a task usually used to examine implicit learning. This task, developed by Berry and Broadbent (1984), required a problem solver to learn to manipulate a computer-simulated person called *Clegg* into a state and keep him there. Previous results (Berry & Broadbent, 1984, 1987; Hayes & Broadbent, 1988) have found that participants learn to bring Clegg to a specific goal state without learning the rules that govern his behavior. Geddes and Stevenson drew on goal specificity results and dual-space theory to explain this result as due to the specific goal encouraging search of instance space, so they predicted that a nonspecific goal (learn to control Clegg) would evoke search of rule space. Such search would ultimately lead to better control of Clegg and better knowledge of the underlying rule. Geddes and Stevenson confirmed that participants performed better when given a nonspecific goal, and found evidence consistent with their explanation for this result. However, they noted that they could not be sure as to the actual strategies their participants used as they had no direct measure of this. Knowing that the nonspecific goal group learned better than the specific goal group but lacking direct evidence regarding how goals affected how they did the task, leaves results like these open to alternative explanations. For example, Buchner, Funke, and Berry (1995) proposed that participants may be learning a look-up table when they learn to control Clegg.

Geddes and Stevenson (1997) suggested that appropriate direct evidence regarding learners' strategies and behavior could be provided by verbal protocols. In this paper we report a verbal protocol study that investigated in detail what effect goal specificity had on the process of

problem solving. To derive our predictions we used the idea of dual-space search, therefore we will explain this theoretical framework in more detail.

Dual-space Theories

Our predictions in this paper used the assumption of Newell and Simon (1972) that problem solving can be seen as a form of search in a space consisting of all possible states of the problem. This space is defined by three components: the initial state, the goal state, and the operators, which transform any state. Simon and Lea (1974) wanted to incorporate induction with problem solving, so to do this they generalized Newell and Simon's definition of a problem space by proposing that search for the rules describing a task is conducted in a *rule space*. The states in rule space are all the possible rules, and the operators are processes for generating, modifying and testing rules. Testing rules, however, requires search of *instance space*, which consists of all possible states of the task, that is, instances. The operators in this instance space are processes allowed by the task for moving between instances. Thus the two problem spaces are conceptually distinct, but interact: test processes for rule space lead to the generation of instances, whereas information that results from such instances drives search in rule space.

Dual-space search has been extended to scientific reasoning by Klahr and Dunbar (1988) in their Scientific Discovery as Dual Search model. They treat scientific discovery as a form of problem solving in which people search a *hypothesis space* (similar to rule space) and an *experiment space* (similar to instance space). Reasoners propose hypotheses, which are then tested by conducting experiments. Thus scientific discovery is an interactive search between these two spaces. Klahr and Dunbar found evidence from participants' verbal protocols that they searched both experiment space and hypothesis space when they had to discover the effect of a novel computer function. This finding was supported in subsequent studies (Dunbar, 1993; Klahr, Fay, & Dunbar, 1993). However, in their research they did not manipulate which space

was searched. (Note that in general, we will use the terms ‘rule’ and ‘instance’ space, rather than ‘hypothesis’ and ‘experiment’ space, simply because that terminology seems more general. Whereas rules usually are hypotheses, not all instances are experiments.)

In Vollmeyer et al. (1996) we proposed that goal specificity effects could be explained in terms of the dual-space theories of problem solving. Specific goals could be seen as encouraging search of an instance space. Such a space corresponds to what is usually meant when we refer to problem solving as search; that is, we set specific subgoals that are states in this space and reach the goal via those subgoals. A specific goal is a state in such a problem space, which is why specific goals encourage a focus on instance space. In contrast, a nonspecific goal could be seen as encouraging search of rule or hypothesis space because participants are not given a state in instance space to reach (i.e., a goal).

Vollmeyer and Burns (1996) tested an implication of explaining problem solving in terms of dual spaces, that directly placing participants into rule space should lead to better learning because it would encourage further hypothesis testing. So we asked participants to start by testing a hypothesis we gave them about a system. Vollmeyer and Burns found that even giving learners an *incorrect* hypothesis about a link in the system improved performance. A control condition (called *link-only*) provided the identical amount of *correct* information as was provided in the incorrect hypothesis condition, but participants were told that they could assume that the link was true rather than being asked to test this as a hypothesis. So simply encouraging hypothesis testing of even one hypothesis at the beginning of the task led to better performance.

The results of Vollmeyer et al. (1996) and Vollmeyer and Burns (1996) suggested that participants learned more when conditions (either a nonspecific goal or an incorrect hypothesis to test) seemed to encourage hypothesis testing. Although these studies did not show directly that the participants in the better performing conditions increased hypothesis testing, they were

consistent with an explanation in terms of dual-space theories. So it seemed reasonable to use dual-space theories to derive predictions for what we would expect to discover in verbal protocols taken from participants given a nonspecific or specific goal. In particular, we reasoned that if a nonspecific goal increased search of hypothesis/rule space, then we would expect the protocols of participants given a nonspecific goal to reveal more hypothesis testing than the protocols of participants given a specific goal. Verbal protocols appeared to be a suitable way of testing this hypothesis derived from dual-space theories, as verbal protocols were the main form of data Klahr and Dunbar (1988) and Dunbar (1993) used to support the claim that problem solvers performed a dual-space search. To show how we went about testing this prediction, we first will explain the problem solving task we used.

Water-tank: A Linear System

Various tasks have been used to examine goal specificity, as discussed above. In our own work (Vollmeyer & Burns, 1996; Vollmeyer et al., 1996) we had problem solvers learn to control a linear system (Funke, 1991) because in such systems goal specificity is easy to operationalize and we can measure performance easily. Such linear systems consist of a set of inputs with weighted links connecting them to outputs. Values entered for each input are multiplied by the weight on each link that connect an input to an output. These products are then added to the previous value of the outputs in order to determine each output's new state.

(Figure 1 about here)

In Vollmeyer et al. (1996) we used a system of this form called biology-lab, but this system included a decay factor. In the current experiments we used a simpler system called *water-tank* that lacked a decay factor (see Figure 1). Participants were never shown the links, but were told that they had to learn to control three water quality measures (oxygenation, chlorine [Cl] concentration, temperature) by varying three inputs (in Figure 1, salt, carbon, lime).

To learn about how the inputs affected the outputs, participants were told they would be given a series of trials. On each trial they entered inputs that modified the current state. The computer then displayed the new values for the outputs, and on the next trial participants could try to modify these new outputs by changing the inputs again. In an initial series of trials, learners explored the system (the *exploration phase*), but then they were told to bring the system to a goal (i.e., a particular set of outputs) as soon as possible (the *solution phase*). To implement the goal specificity manipulation, we varied when participants were informed of the exact output values that constituted the goal. Specific goal participants knew the values from the beginning of the exploration phase, whereas nonspecific goal participants were not informed of these values until after the exploration phase. In order to test transfer, both groups were then given a new goal (i.e., a new set of output values) to reach as soon as possible (the *transfer phase*).

The aspect of controlling linear systems that makes them particularly useful for our experiments is that during the exploration phase they can be approached in two different ways. The task can be treated as a pure problem solving task in which the participants try to bring the outputs to a particular goal state without worrying about the nature of the links. In terms of the dual-space theories, this is search of instance space. Alternatively, participants may treat water-tank as a hypothesis testing task in which they try to generate, test, and modify hypotheses to discover the rules governing the system's behavior. This approach requires an interactive search of both rule and instance space. Of course problem solvers may vary with regard to the extent to which they treat the task as induction or problem solving, and they may change during the task.

Plan of this Paper

The principle aim of this paper was to test the hypothesis that a nonspecific goal provokes hypothesis testing. In particular, we intended to test this by gathering direct evidence regarding what participants given specific or nonspecific goals do when given the water-tank

task. This evidence will be verbal protocols collected from problem solvers during the exploration phase. In detail, we predicted that participants given a nonspecific goal would tend to test hypotheses, whereas those given a specific goal would tend to be goal-oriented.

Verbal protocols should be shorter and easier to gather for a simpler system than biology-lab, which we knew to be difficult (Vollmeyer et al., 1996). So in Experiment 1 we first tried to validate the use of the simpler water-tank system by demonstrating that goal specificity effects were still present. Replicating goal specificity effects with this system was both an important preliminary step towards using it in a protocol study and a way of subsequently confirming that collecting protocols did not change the pattern of performance effects. This could also generalize goal specificity effects across levels of task difficulty. So we again predicted that our measures of learning – transfer performance and participants' ability to accurately describe the structure of the water-tank – would be higher when participants were given a nonspecific goal.

Experiment 2 was the protocol study designed to directly test our hypothesis. If our prediction of greater hypothesis testing by problem solvers given a nonspecific goal was confirmed then this could also be seen as support for the dual-space search theories from which our predictions were derived.

Experiment 1

The first experiment was designed to generalize goal specificity effects to a simpler system than that used in Vollmeyer et al. (1996), and hence to validate its use for the protocol study in Experiment 2. Perhaps hypothesis testing is unnecessary with simpler tasks, in which case a nonspecific goal may not increase learning.

A secondary aim of Experiment 1 was to try to throw some light on an issue that arises when working with linear systems: does prior knowledge of the task domain affect performance? Biology-lab was a completely arbitrary system, in that the links between the inputs and outputs

bore no relationship to any real phenomena in nature. However, prior knowledge may obviate the need for search of rule space. Beckmann (1994) suggested that prior knowledge affects learning in systems like biology-lab. If this is true, then the labeling of the inputs and outputs could be critical. Therefore it may be difficult to generalize between experiments using different cover stories or variables names.

To test effects of prior knowledge, we designed two versions of the water-tank system. The system for the *no prior knowledge* condition is shown in Figure 1 and it was intended that the variable labels would not suggest any links. The *prior knowledge* version of the system was identical except that we gave the inputs labels that were designed to make the links as obvious as we could imagine, short of using the identical words for inputs and outputs. So in the system shown in Figure 1 we substituted the input labels *chlorine*, *hot water*, and *oxygen* for *salt*, *carbon*, and *lime*, respectively. The links were now: *chlorine* increases *chlorine concentration*, *hot water* increases *temperature*, and *oxygen* increases *oxygenation*, plus *hot water* decreases *chlorine concentration* (this last link was not intended to be obvious). Thus, if participants were using prior knowledge to try to guess the nature of links, then this would be a successful strategy in the prior knowledge condition. Therefore we predicted better performance on all our measures for the prior knowledge condition as compared to the no prior knowledge condition. However, as long as this effect did not interact with goal specificity, it would be difficult to claim that prior knowledge could form part of an alternative explanation for goal specificity effects.

Method

Participants

Sixty-three undergraduate (38 female, 25 male) students at the University of California, Los Angeles, participated for course credit.

Design and Procedure

The experiment used a 2x2 between-subjects design, with the two levels of goal specificity (nonspecific goal [NSG] vs. specific goal [SG]) and two levels of prior knowledge (prior vs. no prior).

The procedure was similar to Vollmeyer et al. (1996). Participants first read instructions specifying that the experiment examined knowledge acquisition while solving a problem. They were asked to imagine that they were in a laboratory, and that their research required them to figure out how to control the effects of various inputs on the water quality in a tank. So they had to learn about the relationships between inputs and the water quality factors. Before starting, the experimenter explained how the computer interface worked, and showed how to vary inputs and interpret the changes to outputs. Participants were told that eventually they would be given a goal to reach; but they were also told that they did not have to reach it until they were asked to.

On the computer screen the participants could see the values for each input and each output since the beginning of a round. On each trial, participants decided how many of the inputs they wanted to change and by how much. These changes could be any real number, positive or negative, within the range of -100 to +100. After making a change the computer displayed the new values for each output, then on the next trial the inputs modified the current state (i.e., the output values produced by the previous trial). Six trials made up a *round* and each round started with the same output values (namely, oxygenation = 100; chlorine concentration = 500; temperature = 1000). Only at the beginning of each round were the outputs reset to these values, so problem solvers could not simply repeatedly practice using the same inputs to jump from the initial to the goal state, even if they discovered such a set of inputs.

Before starting, all participants were instructed that the best strategy for exploring the task was to vary only one input variable at a time. Vollmeyer et al. (1996) found that giving

participants this strategy led to much better and less variable performance because the possibility of using this strategy appeared to never occur to some participants if it was not suggested to them. However, participants were not prohibited from using another strategy, so they could choose to do something different.

The first two rounds were the *exploration phase*. At the end of each of these two rounds, participants were presented with a diagram similar to Figure 1 (but with all links removed) in order to assess their knowledge of the structure of the system. They were asked to draw links between inputs and outputs they thought were linked, and place directions and weights on these links if they thought they knew them. From these diagrams a *structure score* could be calculated for each of the first two rounds.

The *solution phase* followed the completion of the exploration phase. In this phase participants were given one round to bring the system to a goal state (i.e., a certain set of outputs, in this experiment: 50 oxygenation, 700 Cl concentration, and 900 temperature) and to maintain the system in that state. How close they got to the goal was expressed as the *solution error*.

The *transfer phase* followed the solution phase. Participants were given one round during which they tried to reach and maintain the transfer goal state. How close they got to this new goal (250 oxygenation, 350 Cl concentration, and 1100 temperature) was the *transfer error*.

The goal specificity manipulation was implemented by varying when participants were informed of the goal to reach in the solution phase: SG participants were told the goal – the set of output values – at the start of the exploration phase, whereas NSG participants were not told these values until the start of the solution phase. Figure 2 gives an overview of the procedure.

The task was controlled and presented by a PC computer and the entire experiment took under an hour to complete.

(Figure 2 about here)

Results

Dependent Variables

Three dependent variables were analyzed which measured both knowledge and accuracy in reaching the goal states in the same way as in Vollmeyer et al. (1996).

(1) *Structure score*. Participants completed the structure diagram after each of the two rounds of the learning phase. From this diagram we derived a structure score that reflected a participant's degree of knowledge about the underlying structure of the system. This score was computed by finding the proportion of correct specifications for each of the three components of the structure: links, directions, and weights. As in Vollmeyer et al. (1996), a correction for guessing was used, and the maximum value was 3.0.

(2) *Solution error*. Solution error in reaching the specific goal state during the solution phase was computed as the sum of the absolute differences between each goal value and the value produced for each of the three output variables. Thus low scores indicate better performance. As in Vollmeyer et al. (1996), we report the mean solution error over all six trials averaged across all three output variables, but with a log transform applied due to the skewness of the distribution.

(3) *Transfer error*. Transfer error for the transfer phase was calculated in the identical way as the solution error, except that performance was compared to the goal state given in the transfer phase. Transfer error was a measure of how well participants could apply their knowledge to reaching a new goal; hence it is a measure of how much participants know about the system's operation, as opposed to what they know about reaching a specific goal.

An alpha level of .05 was used for all statistical tests.

Relationship between Knowledge and Performance

We predicted that participants having more knowledge about the system's structure

would come closer to the goal states than participants having less knowledge. Confirming this, we found significant correlations between structure score at the end of the solution phase with both solution error, $r(63) = -.28, p = .026$, and transfer error, $r(63) = -.40, p < .001$.

Effects of Prior Knowledge and Goal Specificity on Knowledge Acquisition

Knowledge was measured by calculating structure scores at the end of Round 1 and Round 2 of the exploration phase (see means in Table 1). A 2x2x2 repeated-measures ANOVA with factors of round, prior knowledge, and goal specificity showed that the NSG group learnt more about the structure of the task, $F(1, 59) = 8.07, p = .003$, and that there was a marginal significant effect of round, $F(1, 59) = 3.08, p = .084$, but no effect of prior knowledge, $F(1, 59) = .02$. None of the interactions approached significance (all p 's $> .20$).

(Table 1 about here)

Effects of Prior Knowledge and Goal Specificity on Performance

Performance was measured with the solution error in the solution phase and the transfer error in the transfer phase. These two error measures were correlated, $r(63) = .74, p < .001$. Although participants in the NSG condition achieved greater overall knowledge of the system, as shown by structure scores, those in the SG condition had two additional rounds during which they could have attempted to attain the goal set for all participants in the solution phase. Vollmeyer et al. (1996) found no effect of goal specificity on solution error, but a significant interaction between error type and goal specificity as NSG participants out-performed SG participants on transfer error. So we again expected, after controlling for solution error, that the SG group would do worse than the NSG group on transfer error. Therefore the interaction between solution and transfer errors should be significant. We ran a repeated-measures 2x2x2 ANOVA with prior knowledge and goal specificity as between-subjects factors and error (solution vs. transfer) as a within-subject factor. The results showed the predicted interaction

between goal specificity and error, $F(1, 59) = 7.41, p = .008$. Table 1 shows that both goal specificity groups had almost identical performance on solution errors, but in the transfer phase the NSG group improved their performance whereas the SG group stayed the same. However, the difference due to goal specificity on transfer error alone was only marginally significant, $F(1, 59) = 3.22, p = .078$. Again prior knowledge had no effect at all, either directly on transfer error, $F(1, 59) < .01$, in interaction with goal specificity, $F(1, 59) = 1.15, p = .29$, or in interaction with both goal specificity and error type, $F(1, 59) = 0.76$.

Discussion

Experiment 1 showed that a simpler system was suitable for producing goal specificity effects. Participants with a nonspecific goal obtained more knowledge about the system and were better at reaching a transfer goal, replicating the result found with a more difficult system in Vollmeyer et al. (1996). Thus this simpler system appeared suitable for use in Experiment 2.

The results showed no significant effects of prior knowledge, either as a main effect or in interaction with other factors. Like any null result, this finding is open to interpretation and questions concerning statistical power. However, we still found this a surprising result given the apparent strength of the prior knowledge manipulation. In the no prior knowledge condition none of the links seemed to be implied by the labels (though some participants may have inferred salt increasing Cl concentration), but in the prior knowledge condition three of the four links were: chlorine increases Cl concentration, hot water increases temperature, and oxygen increases oxygenation. These links were intended to be obvious so we found our results somewhat surprising, but also a relief. If such strong connections did not affect performance, it seems less plausible that participants are strongly affected by subtle hints about links that might be suggested by the particular labels chosen in a particular experiment. This makes it easier to generalize across experiments using different labels for variables. The results point towards the

possibility that participants pay little attention to the semantic content of the system to be learnt.

Experiment 2

In order to provide direct evidence regarding how goal specificity affects the behavior of participants, in Experiment 2 participants were asked to think aloud while they performed the water-tank task under either specific or nonspecific goal conditions.

A secondary aim of this experiment was to follow up the finding by Vollmeyer and Burns (1996) that giving a single hypothesis to test at the beginning ultimately led to better learning. We had assumed that this occurred because testing one hypothesis encouraged further hypothesis testing, rather than that it resulted from participants simply learning about the link tested via the initial hypothesis. Verbal protocols allowed us to directly assess if participants given a hypothesis to test would continue to do more hypothesis testing compared to those not given a hypothesis to test.

Detailed predictions for protocols

We expected that protocols from specific goal participants would reveal that these participants would spend more time trying to reach the goal they were given at the beginning of the task, despite the fact that they were told that they did not need to reach the goal until the solution phase. We based this prediction on the dual-space search theory and the indirect evidence from Vollmeyer et al. (1996) and Geddes and Stevenson (1997) that specific goal learners pushed the system they were interacting with towards the goal they were given. If the verbal protocols of specific goal learners were not to reveal substantial amounts of goal-directed behavior, we would be surprised and question our goal specificity manipulation.

The critical question we sought to answer through the protocol analysis was: What do nonspecific goal learners spend their time doing? Answering this question could distinguish between different explanations of the goal specificity effects on problem solving. We predicted

that nonspecific goal learners would devote more trials to testing hypotheses than would specific goal learners. Geddes and Stevenson (1997) would appear to agree with this prediction. However, Miller et al. (1999) assumed that the amount of hypothesis testing would be independent of goal, and instead suggested that nonspecific goal learners may still be goal-oriented but set themselves appropriate goals.

Sweller (1988) suggested that nonspecific goal learners may simply explore the problem space in order to see what moves are possible, but his main focus was on what he regards as the increased cognitive load goal-oriented behavior places on specific goal learners. Therefore we tried to reason through the cognitive load implication of our task. Sweller claimed that the requirement to hold in mind the current state and goal state creates a high cognitive load when performing means-ends analysis, but hypothesis testing would also appear to require keeping in mind simultaneously the current state and a projected state. So if both goal-oriented behavior and hypothesis testing increase cognitive load, perhaps cognitive load could be minimized through behavior we will term *nonpredictive testing*, which would involve simply trying different possibilities without anticipating an outcome. This may fit to Sweller's concept of simple exploration, as participants could learn all they need by entering inputs that address questions of the type "I wonder what happens when I change the salt." Answering such questions would appear to minimize cognitive load because they require no calculation, no planning, and no need to keep in mind more than one state. Both hypothesis testing and nonpredictive testing would be consistent with the "vary one thing at a time" strategy that Vollmeyer et al. (1996) observed in the pattern of inputs produced by nonspecific goal learners.

Given the lack of direct evidence regarding the strategies of specific goal participants, it is hard to rule out any possibility that might produce better learning performance for nonspecific goal learners than for specific goal learners. For example, nonspecific learners may simply be

producing a better look-up table as Buchner et al. (1995) suggested, a table that could be produced in many ways without using hypothesis testing. It is even possible that there is nothing systematic about the behavior of nonspecific goal learners; they might simply perform better because they do not suffer from the disadvantage the goal represents for specific goal learners.

In this experiment we tested our hypothesis that a nonspecific goal would increase hypothesis testing, however we did not expect goal specificity to have an all-or-none effect. We propose that goal specificity affects the strategies learners use, but someone given a specific goal could choose to ignore that goal and instead use hypothesis testing to learn the rules of the system. Similarly, not being informed of a specific goal does not prevent participants from generating their own specific goal. Further, strategies could be mixed. So rather than a dichotomy, we expected the effects of goal specificity to be a continuum.

Method

Participants

Eighty introductory psychology students (34 female, 46 male) at the University of California, Los Angeles, participated for course credit.

Design and Procedure

Experiment 2 utilized a 2x2 between-subjects design. As in Experiment 1, we manipulated goal specificity (NSG vs. SG), but we also varied hypothesis-instruction (incorrect hypothesis vs. link-only information). The version of the water-tank system was the *no prior knowledge* version of the task used in Experiment 1 (see Figure 1). The instructions were identical to those for Experiment 1, except that at the end of the instructions the hypothesis manipulation was given.

Hypothesis-instruction was manipulated in the same way as in Vollmeyer and Burns (1996, Exp. 2), except that their correct hypothesis condition was excluded. Participants in the

incorrect hypothesis condition were told that an unnamed researcher had a hypothesis about how the water quality was affected. The researcher proposed that in this tank environment the water quality factor oxygenation is affected by changes in lime, and that the weight on the link between lime and oxygenation is -5. Participants in this condition were told that they should begin the task by testing this hypothesis. The weight and direction suggested by this hypothesis is false, but the link is correct. This link was chosen to be the focus of the hypothesis because it was the simplest, as reflected by the finding that most of the participants in Experiment 1 (83%) discovered this link and its weight. Thus little information was provided by this hypothesis that participants were not likely to discover by themselves. The control condition was the *link-only* condition as in Vollmeyer and Burns. In the link-only condition participants were told that an unnamed researcher said that in this tank environment the water quality factor oxygenation is affected by changes in lime. Thus the link-only condition provided the identical *correct* information as the incorrect hypothesis condition, but the link-only condition did not suggest an incorrect weight or instruct participants to start by testing a hypothesis.

The task was performed in the same way as Experiment 1, so participants had an exploration, a solution, and then a transfer phase (see Figure 2). As in Experiment 1, specific goal participants were told the goal state at the start of the exploration phase, whereas nonspecific goal participants were not told the goal state until the end of the exploration phase. The goal states for the solution and transfer phases were identical to those used in Experiment 1. The major addition to the procedure in Experiment 2 was that talk-aloud protocols were collected. After reading the instructions and having the task explained to them, participants were told that they would be asked to talk-aloud while they did all aspects of the task. We used the instructions presented by Ericsson and Simon (1993, p. 376).

Results and Discussion

The dependent variables for learning performance – *structure score*, *solution error*, and *transfer error* – were calculated in the identical way to Experiment 1.

Performance Measures

Relationship Between Knowledge and Performance

As in Experiment 1, we expected that better knowledge about the system's structure would lead to better performance. To test this hypothesis, we correlated the structure score at the end of the solution phase with the two performance measures (solution error and transfer error). Participants with better knowledge had lower solution error, $r(80) = -.49$, $p < .001$, and lower transfer error, $r(80) = -.60$, $p < .001$, than participants who knew less about the structure. Performance in the solution and transfer phases were again closely related, $r(80) = .77$, $p < .001$.

Effects of Hypothesis-instruction and Goal Specificity on Knowledge Acquisition

Knowledge was measured by calculating structure scores at the ends of Round 1 and Round 2 (see means in Table 2). A 2x2x2 repeated-measures ANOVA with factors of round, hypothesis-instruction (incorrect vs. link-only) and goal specificity (SG vs. NSG) showed again that the NSG group learnt more about the system's structure, $F(1, 76) = 5.27$, $p = .026$. Consistent with previous findings, we found no interactions between goal specificity and round, or goal specificity and hypothesis-instruction, all $F(1, 76) < 1.0$.

(Table 2 about here)

There was also no main effect of hypothesis-instruction, $F(1, 76) = 1.71$, $p = .20$, but there was an interaction between hypothesis-instruction and round, $F(1, 76) = 4.58$, $p = .036$ (the three-way interaction was not significant, $F(1, 76) = 1.03$, $p = .31$). Link-only participants start with an advantage in terms of knowing the structure because they are given a link, but by the end of the second round the advantage of testing a hypothesis appears to have manifested itself.

Effects of Hypothesis-instruction and Goal Specificity on Performance

For solution and transfer errors we performed a 2x2x2 repeated-measures ANOVA with goal specificity and hypothesis-instruction as between-subjects factors, and error (solution vs. transfer) as a within-subject factor (for means, see Table 2). Again we found the predicted goal specificity by error interaction, $F(1, 76) = 4.20, p = .044$, but there was also a goal specificity by hypothesis-instruction interaction, $F(1, 76) = 4.02, p = .048$. No other main effects or interactions were significant (all p 's $> .20$).

To investigate the interaction between goal specificity and hypothesis-instruction we focussed on the critical transfer error. The hypothesis-instruction by goal specificity interaction for transfer error alone was significant, $F(1, 76) = 4.19, p = .044$, and a post-hoc LSD-test showed that the link-only/SG group was the worst group in reaching the transfer goal state and differed significantly from each of the other three. However these three groups did not differ significantly from each other. This result suggests that the SG participants were poorer at learning how to control the system, but that being given the incorrect hypothesis inoculated them against this disadvantage.

Summary of performance findings

Experiment 2 replicated the goal specificity and hypothesis-instruction effects found in earlier experiments, despite participants having to produce verbal protocols during the task. Again, learning with a nonspecific goal increased learning of the system's structure and improved transfer performance. Testing an incorrect hypothesis was also found to improve performance but only for those learners disadvantaged by being given a specific goal to reach. Perhaps this is not surprising given that this group received neither of the manipulations that were intended to encourage hypothesis testing. However, the group that received both beneficial manipulations (nonspecific goal and incorrect hypothesis), was not better than the groups

receiving only one. This finding may be an indicator that the effects of these manipulations are not additive. In terms of dual-space theories, perhaps once one is focussed on rule space, further encouragement to do so may have little effect.

Protocol Analysis for Exploration Phase

Although all participants produced protocols, the tapes for only the first 20 participants have been transcribed, five in each of the four conditions. The pattern of results appeared to be very strong, thus the analysis of more protocols may not be productive.

Initial examination of the protocols indicated that the most useful unit of analysis was a trial. Most trials could be classified on two dimensions: whether the participant was trying to reach a goal; and, the way they tried to learn about the task. When trying to learn about the task there was a clear distinction between testing specific hypotheses (hypothesis testing), and just investigating what an input does without making predictions (nonpredictive testing). While sometimes participants did both on one trial, usually the protocols indicated that they did one or the other. Thus each trial could be classified on each of the following approaches:

1) *Nonpredictive testing*: Trials on which participants tried to discover the function of an input, but had no expectation regarding the effect that input would have. For example, "I wonder what happens if I change salt."

2) *Hypothesis testing*: Trials on which participants manipulated an input, and had an expectation as to what should happen to the outputs. For example, "So now I'm going to see if reducing lime by a negative number will decrease the oxygenation." Such a hypothesis concerned the existence of a link, the direction on a link, or the weight on a link.

3) *Goal-oriented*: Trials on which the protocols indicate that the participant was trying to bring at least one output closer to its goal state. For example, "I'll change the oxygen, it's at 100, I'm going to change it to the desired 50."

It is possible that a participant could be trying to bring an output to its goal state, while simultaneously trying to test a hypothesis about the link to that input. In this case the trial was scored as both hypothesis testing and goal-oriented. Because participants may change more than one input at a time, it was also possible for a participant to test a hypothesis for one input and simply investigate the effect of another input. In this case we classified the trial on the basis of the more sophisticated approach, that is, hypothesis testing. To test the reliability of these three categories an independent rater also coded the verbal protocols of 10 participants resulting in a rater agreement of 87%.

For each round we calculated the mean proportion of trials using each approach. Figure 3 (upper histogram) presents these means for the SG group in each round. Figure 3 (lower histogram) gives the mean proportions for the NSG group in each round.

(Figure 3 about here)

Goal specificity effects

To test the effects of our manipulations on protocols, we ran a 2x2x2 ANOVA (goal specificity, by hypothesis testing, by round) for each category of protocol (goal-oriented, nonpredictive testing, and hypothesis testing). For brevity we only report the F-tests and contrasts of interest.

The proportion of hypothesis testing trials did not differ between goal specificity conditions in the first round, $F(1, 16) = 0.80$, but in the second round goal specificity had a huge effect, $F(1, 16) = 40.27$, $p < .001$. Although, the proportion of goal-oriented trials was significantly greater for the SG group even in Round 1, $F(1, 16) = 5.92$, $p = .027$, this effect increased greatly in Round 2, $F(1, 16) = 65.79$, $p < .001$.

Overall, there was greater use of nonpredictive testing by the NSG group than the SG group, $F(1, 16) = 17.29$, $p = .001$, but across both groups there was a decline in use of this

approach in the second round, $F(1, 16) = 16.86$, $p = .001$. However, as the previous analysis suggested, the two groups diverged with regard to what they substituted for nonpredictive testing. For SG participants (upper histogram in Figure 3), the most common approach in the first round was hypothesis testing, but use of hypothesis testing virtually disappeared in the second round for the SG group. This round effect was significant, $F(1, 8) = 14.11$, $p = .006$. Instead SG participants had a huge increase in the proportion of goal-oriented trials from Round 1 to Round 2, $F(1, 8) = 21.30$, $p = .002$. In contrast, the pattern for the NSG group was quite different, as Figure 3 (lower histogram) shows. Use of the nonpredictive approach also declined from Round 1 to Round 2, $F(1, 8) = 15.33$, $p < .004$, but there was a large increase in use of hypothesis testing, $F(1, 8) = 9.30$, $p = .016$.

Hypothesis-instruction effects

The effect of hypothesis-instruction was clear. Across both rounds, participants with link-only information ($M = .43$, $SD = .27$) made greater use of nonpredictive testing than did those with the incorrect hypothesis ($M = .21$, $SD = .14$), $F(1, 16) = 11.57$, $p = .004$. Hypothesis-instruction had no effect on the proportion of trials that were goal-oriented, $F(1, 16) = .10$ (Link-only: $M = .26$, $SD = .35$; incorrect hypothesis: $M = .28$, $SD = .31$), but use of hypothesis testing was greater with an incorrect hypothesis ($M = .52$, $SD = .26$) than with link-only information ($M = .26$, $SD = .16$), $F(1, 16) = 12.40$, $p = .003$. There were no interactions between hypothesis-instruction and round for any of the approaches. In addition, no significant interactions between goal specificity and hypothesis-instruction were found.

Testing inappropriate rules

While our results have consistently suggested that testing hypotheses was beneficial for performance and learning, we sometimes observed that participants tested impossible hypotheses. Such hypotheses were not just incorrect, but impossible in the sense that no minor

modification could make them correct. In particular, the protocols revealed that some participants tested hypotheses about interactions or combinations of variables. For examples:

"So I think we need to find an equilibrium between salt and carbon as they both affect the same element."

"So actually the mix has something to do with the chlorine, neither one alone does. "

"Salt, carbon ... I guess we can see if they interact."

Ten of the 20 participants whose protocols were analyzed mentioned such hypotheses two in each condition except for four in the link-only/NSG condition. Testing a hypothesis like this could have just been a way of exploring once there was nothing else to learn, so it does not necessarily indicate a confused participant. However, this behavior suggests that there are unfruitful hypotheses for linear systems that participants may have a predisposition to test.

Summary of protocol findings

The pattern of protocol results appeared to be strong. The goal specificity manipulation changed the extent to which participants tried to reach the goal, but only substantially in the second round. In the first round there was little effect of goal on how much learners engaged in nonpredictive testing or hypothesis testing. In the second round the two groups diverged as NSG participants tested hypotheses more, whereas SG participants became more goal-oriented. Vollmeyer et al.'s (1996) analysis of strategies also seemed to indicate that SG participants might not be as goal-oriented early in the exploration phase compared to later. However this measure was indirect (participants had to get closer to the goal on three consecutive trials) as failure to reach closer to the goal would also be categorized as not trying to reach the goal. We now have direct evidence that it is not simply that SG participants try to reach the goal and fail, but that they focus on exploration early in the task. However they switched prematurely to a strategy of trying to reach the goal rather than testing hypotheses.

The incorrect hypothesis encouraged hypothesis testing in both rounds. In contrast, the link-only group had a greater tendency to continue to search instance space through nonpredictive testing. Thus we confirmed our prediction that the effect of the hypothesis-instruction manipulation was to increase hypothesis testing.

General Discussion

That goal specificity affects learning has been shown a number of times and we once again showed that a nonspecific goal led to better learning. However, protocol analysis provided direct evidence that a nonspecific goal affected the process problem solvers used by increasing their use of hypothesis testing. All theoretical accounts of goal specificity would expect giving a specific goal to increase the amount of goal-oriented behavior, and we now have direct evidence that this occurs. However, the theoretical accounts varied with regard to how they explained the empirical results showing that a nonspecific goal improved learning. The direct evidence from protocols that a nonspecific goal increases hypothesis testing also offers support to the dual-space search account of goal specificity proposed by Vollmeyer et al. (1996) and Geddes and Stevenson (1997) because we derived our predictions from this theoretical framework.

In our protocol analysis we have not reported participants' use of means-ends analysis which was the focus of both Sweller's (1988) and Miller et al.'s (1999) explanations of goal specificity effects. Our reasons for doing this were two-fold. First, we saw no direct evidence in the protocols of means-ends analysis, at least of the form: "I will do X in order to create the conditions for doing Y." Second, it was hard to know what to define as means-ends analysis in this task. Anderson (1993, p. 2) pointed out that statements of the form "people solve problems by means-ends analysis" are not falsifiable because what constitutes means-ends analysis depends on one's definition. For example, in our protocols we see that participants say that they will first test a hypothesis about one variable before moving onto another. This is a form of

subgoaling and could be regarded as means-ends analysis. So Miller et al. might be able to incorporate into their model a form of means-ends analysis that would test hypotheses, but this would require dropping their assumption that hypothesis testing is unaffected by goal specificity. We think they are correct that the best goal is the appropriate goal, but that our evidence regarding hypothesis testing requires an expanded concept of what can be an appropriate goal.

Similarly, Sweller's (1988) cognitive load approach could be supported if it could be claimed that hypothesis testing requires a low mental load. We did not directly measure cognitive load, so we cannot rule out this possibility. Our finding of initial use of nonpredictive testing by all participants regardless of goal specificity appears consistent with Sweller's approach, if such testing requires a low cognitive load. However, it is not clear how to explain in terms of cognitive load why our participants voluntarily tend to give up such an approach in favor of higher load hypothesis testing or goal-oriented behavior.

Although all theoretical accounts would predict that a specific goal would lead to more goal-oriented behavior, the pattern of this group's behavior is interesting. They did not simply try to reach the goal from the beginning, or switch back-and-forth between goal-oriented and explorative behavior; instead specific goal participants essentially did the same thing in the first round of the exploration phase as the nonspecific goal participants, but then became largely goal-oriented in the second round. The specific goal participants were neither completely following their instructions that they do not need to reach the goal until they are asked to, nor were they misinterpreting their instructions and trying to reach the goal from the beginning. Instead they were first seeking information then switching to trying to reach the goal. The event that appears to have produced this switch is the break between rounds, during which participants had to fill out the structure diagram. It may be that participants decided that reaching the goal immediately was not realistic, but filling in the diagram led them to judge that they knew enough to try to

reach the goal. This finding could be seen as related to Bjork's (1994) argument that a number of studies show that learners can have poor metacognition of their state of knowledge, and that this can lead to behavior that produces poorer learning than may have resulted from a more accurate assessment. Finding that even the strategies of the specific goal participants are more complex than expected, suggests that the role of metacognition in goal specificity experiments is worthy of further study. We appear to have found that goal specificity did not just affect strategies, but that it affected how participants changed strategies.

Goal Specificity and Dual-Space Search

Our experiments were not intended as a definitive test of dual-space search theories of problem solving. What would be a definitive test is unclear. However, our results add plausibility to this theoretical approach. First, we replicated Klahr and Dunbar's (1988) finding that protocols show that learners both test hypothesis and focus on doing experiments intended to gather information. Second, in Experiment 2 the hypothesis-instruction effect found by Vollmeyer and Burns (1996) was replicated and it was shown that giving problem solvers a hypothesis to test increased search of rule space, even if the hypothesis was incorrect. This supported dual-space theories by showing that directly placing problem solvers into rule space increased the amount they searched rule space and how much they learnt about the task. This was true even if the hypothesis contained incorrect information and it was contrasted with a condition in which participants were simply told the correct link.

Most importantly though we used the dual-space theoretical framework to derive predictions about the effects of goal specificity and we confirmed these predictions. The success of dual-space search theories in predicting our results suggests this theoretical framework could provide a definition of goal specificity. What exactly constitutes a nonspecific goal has been somewhat unclear because "nonspecific goal" does not mean "no goal". However, if we define a

nonspecific goal as one in hypothesis/rule space, and a specific goal as one in instance space, then this provides a general definition. Such a definition also allows other phenomena to be related to goal specificity, such as the explicit/implicit learning distinction (as proposed by Geddes & Stevenson, 1997), and the self-explanation effects observed by Chi, Bassok, Lewis, Reimann, and Glaser (1989) and Renkl (1997).

Do Nonspecific Goals always Help Problem Solving?

The two experiments added to the generality of the goal specificity effects found earlier by Vollmeyer et al. (1996). The effects appeared robust across different levels of task difficulty, cover story, and prior knowledge.

However, claiming that specific goals can never help problem solving would seem to be too extreme a claim. Experiment 2 hinted that there may be conditions under which a nonspecific goal may lead to poorer learning. Some participants tested hypotheses that were not just wrong, but inappropriate in that no modification could make them work. Testing interaction hypotheses for this system was not just fruitless, but potentially could be confusing if persisted with. If the participant's understanding of the task is such that interaction hypotheses are part of the rule space, then the rule space will be hard or even impossible to search effectively. In such a case, encouragement via a nonspecific goal to search rule space may lead to poorer performance than that shown by problem solvers given a specific goal who learn to push towards the goal.

This line of reasoning raises the question of what determines the problem solver's rule space? Klahr and Dunbar (1988) suggested that learners have a *frame* of hypotheses and having the right frame is critical, however they do not address where that frame comes from. Better understanding of what determines the rule space may push dual-space theories in interesting directions and is already driving researchers towards tentative proposals of multispace theories (Burns & Vollmeyer, 2000; Schunn & Klahr, 2000).

References

- Anderson, J. R. (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum
- Beckmann, J. F. (1994). *Lernen und komplexes Problemlösen: Ein Beitrag zur Konstruktvalidierung von Lerntests* [Learning and complex problem solving: A contribution to the construct validity of test of learning potential]. Bonn, Germany: Holos Verlag.
- 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.
- Bjork, R. A. (1994). Memory and metamemory consideration in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: knowing about knowing* (pp. 185-205). Cambridge, MA: The MIT Press.
- Buchner, A., Funke, J., & Berry, D. C. (1995). Negative correlations between control performance and verbalizable knowledge: Indicators for implicit learning in process control tasks? *The Quarterly Journal of Experimental Psychology*, 48A, 166-187.
- Burns, B. D., & Vollmeyer, R. (2000). Problem solving: Phenomena in search of a thesis. In L. Gleitman & A. K. Joshi (Ed.), *Proceedings of the Twenty-second Annual Meeting of the Cognitive Science Society* (pp. 627-632). Hillsdale, NJ: Lawrence Erlbaum.
- Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, 127, 355-376.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 18, 439-477.

- Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science*, 17, 397-434.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis*. Cambridge, MA: MIT Press.
- 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* (pp. 185-222). Hillsdale, NJ: Erlbaum.
- Geddes, B. W., & Stevenson, R. J. (1997). Explicit learning of a dynamic system with a non-salient pattern. *The Quarterly Journal of Experimental Psychology*, 50A, 742-765.
- Hayes, N. E., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition*, 28, 249-276.
- 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.
- Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science*, 23, 305-336.
- Newell, A., & Simon, H. A., (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, 21, 1-29.
- Schunn, C. D., & Klahr, D. (2000). Multiple-space search in a more complex discovery microworld. In D. Klahr (Ed.), *Exploring science* (pp. 161-199). Cambridge, MA: The MIT Press.
- Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W.

Gregg (Ed.), *Knowledge and cognition* (pp. 105-127). Hillsdale, NJ: Erlbaum.

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

Vollmeyer, R., & Burns, B. D. (1996). Hypotheseninstruktion und Zielspezifität: Bedingungen, die das Erlernen und Kontrollieren eines komplexen Systems beeinflussen. [Hypothesis instruction and goal-specificity: Factors that influence the learning and controlling of a complex system]. *Zeitschrift für Experimentelle Psychologie*, *43*, 657-683.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity and systematicity of strategies on the acquisition of problem structure. *Cognitive Science*, *20*, 75-100.

Author Note

Bruce D. Burns, Department of Psychology, Michigan State University. Regina Vollmeyer, Institut für Psychologie, Universität Potsdam.

This research was supported by a research stipend to the first author from the Alexander von Humboldt Foundation and a grant from the German American Academic Council to both authors. We would like to thank Ruth Byrne, Geoff Ward, Mareike Wieth, and an anonymous reviewer for helpful comments on an earlier draft of this manuscript. Correspondence concerning this article should be addressed to Bruce D. Burns, Department of Psychology, Michigan State University, East Lansing, MI 48824-1117, USA, E-mail: burnsbr@msu.edu.

Table 1

Means for dependent measures in Experiment 1 for each condition: goal specificity (NSG = nonspecific goal, SG = specific goal) and prior knowledge (prior vs. no prior). (standard deviations are in parentheses)

	<i>n</i>	structure score (Round 1)	structure score (Round 2)	solution error	transfer error
NSG (32)		2.52 (0.75)	2.72 (0.48)	2.59 (1.37)	1.96 (1.65)
prior knowledge	16	2.64 (0.67)	2.80 (0.27)	2.53 (1.38)	1.70 (1.44)
no prior	16	2.41 (0.83)	2.62 (0.62)	2.64 (1.40)	2.21 (1.85)
SG (31)		2.04 (0.98)	2.10 (0.97)	2.56 (2.01)	2.76 (1.88)
prior knowledge	16	1.93 (1.00)	1.94 (1.04)	2.72 (1.97)	2.99 (1.86)
no prior	15	2.15 (0.98)	2.24 (0.89)	2.40 (2.10)	2.58 (1.93)

Table 2

Means for dependent measures in Experiment 2 for each condition: goal specificity (NSG = nonspecific goal, SG = specific goal) and hypothesis-instruction (link-only vs. incorrect hypothesis). (standard deviations are in parentheses)

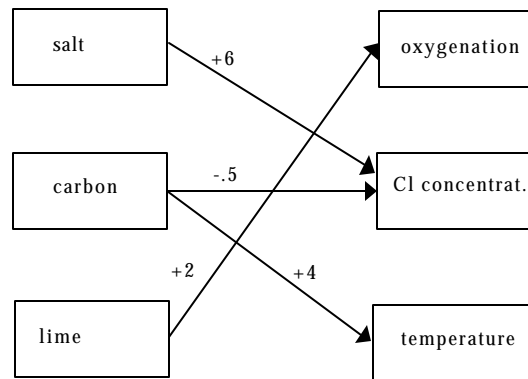
	<i>n</i>	structure score (Round 1)	structure score (Round 2)	solution error	transfer error
NSG (40)		2.69 (0.84)	2.55 (0.95)	1.82 (1.60)	1.47 (1.49)
incorrect	20	2.60 (0.84)	2.77 (0.55)	1.94 (1.68)	1.65 (1.51)
link-only	20	2.78 (0.32)	2.33 (1.20)	1.70 (1.54)	1.29 (1.48)
SG (40)		2.26 (0.80)	2.25 (0.96)	1.98 (1.60)	2.12 (1.68)
incorrect	20	2.34 (0.85)	2.44 (0.78)	1.49 (1.49)	1.59 (1.46)
link-only	20	2.17 (0.75)	2.05 (1.09)	2.47 (1.60)	2.66 (1.75)

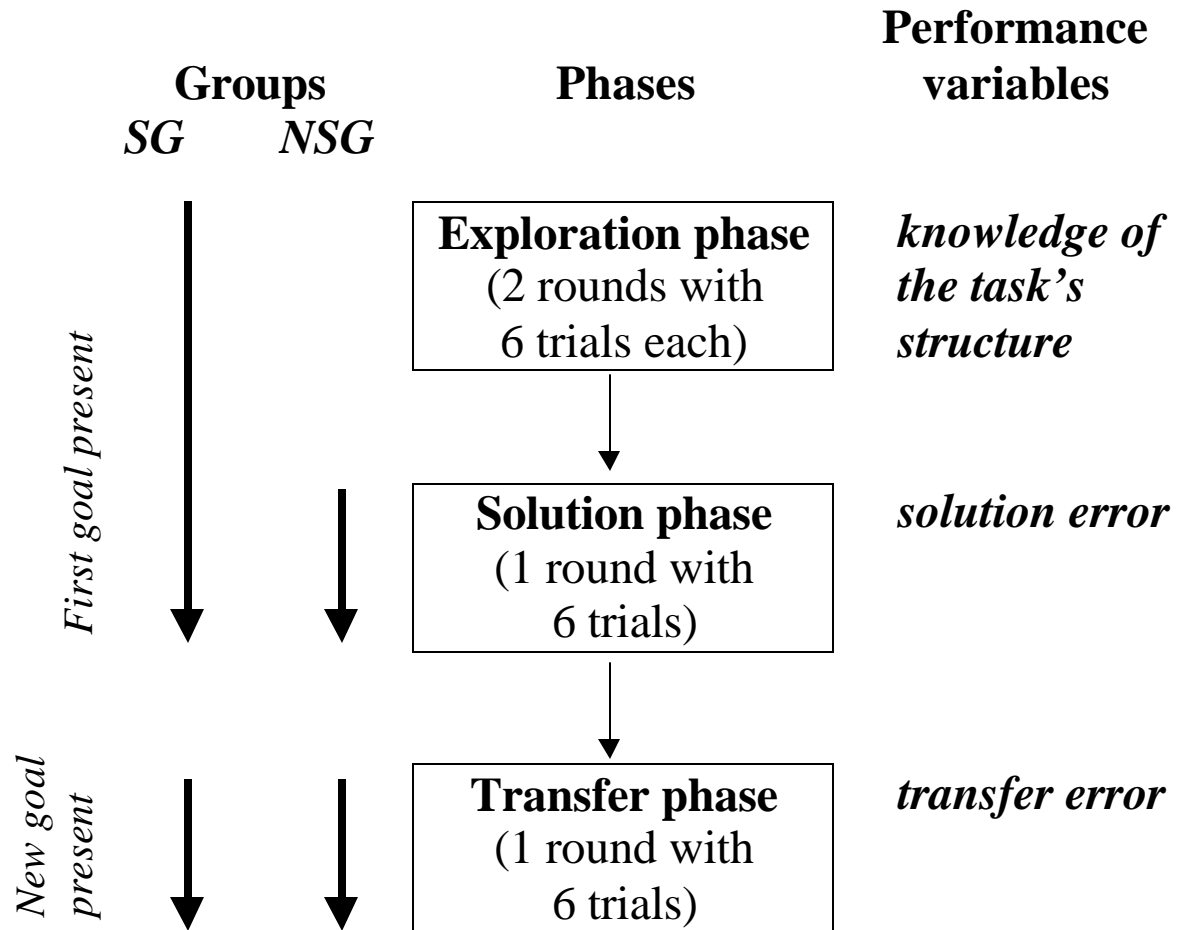
Figure Captions

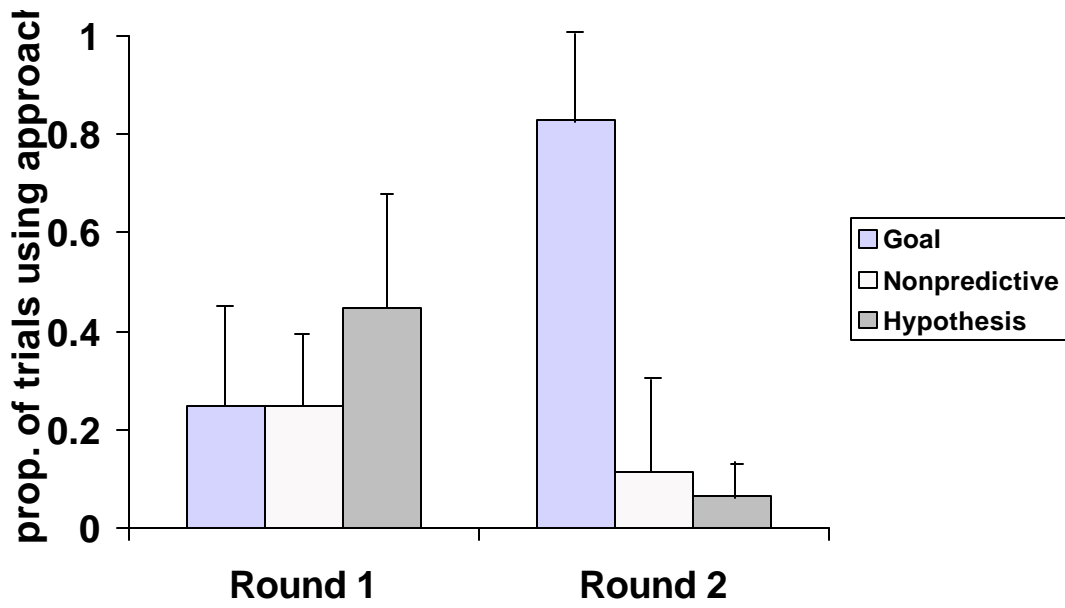
Figure 1. The water-tank system (note that *Cl concentrat.* is the chlorine concentration output).

Figure 2. Diagram for the procedure of the two experiments illustrating when the goal states are present for each group (*NSG* = *nonspecific goal*, *SG* = *specific goal*), and indicating which performance measures were collected in each phase.

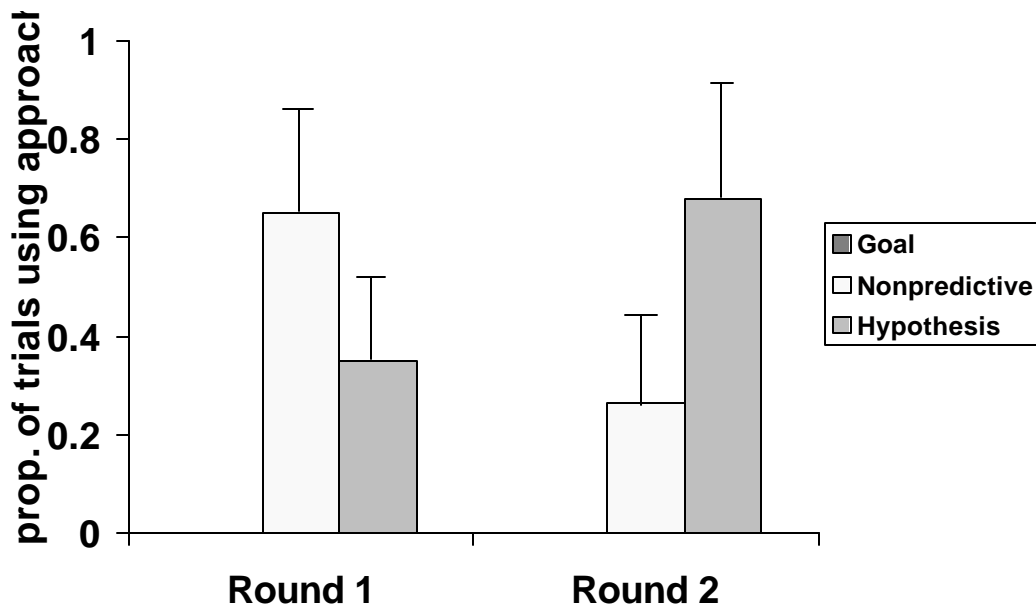
Figure 3. Mean proportions of trials used for each approach (goal-oriented, nonpredictive testing, and hypothesis testing) in each round for specific goal (upper histogram) and nonspecific goal groups (lower histogram). Bars indicate 95% confidence intervals.







Specific goal group



Nonspecific goal group