

Evidence for abstract, schematic knowledge of three spatial diagram representations

LAURA R. NOVICK, SEAN M. HURLEY, and MELISSA FRANCIS
Vanderbilt University, Nashville, Tennessee

Spatial diagram representations such as hierarchies, matrices, and networks are important tools for thinking. Our data suggest that college students possess abstract schemas for these representations that include at least rudimentary information about their applicability conditions. In Experiment 1, subjects were better able to select the appropriate spatial diagram representation for a problem when cued to use general category information in memory about those representations than when cued to use specific example problems given during the experiment. The results of Experiment 2 showed that the superior performance in the general category condition was not based on a comparison of the test problems with examples in memory. The results of Experiment 3 showed that the superior performance was not due to learning that occurred during the experiment or to transfer appropriate processing. The General Discussion section considers the nature of students' representation schemas and the question of why college students have only rudimentary schemas for common and widely applicable diagrammatic representations.

Diagrams as Tools for Thinking

You would probably find it helpful to consult a diagram to fix a leaky faucet or learn a new juggling pattern. In fact, many studies have documented the usefulness of domain-specific diagrams in instructional texts for enhancing learning outcomes, especially in technical domains (e.g., Hegarty & Just, 1993; Levin, 1989; Mayer & Gallini, 1990; Sweller, Chandler, Tierney, & Cooper, 1990).

In addition to this static function of illustration, domain-specific diagrams often serve a more dynamic role during problem solving. Kindfield (1993/1994) characterized the quality and use of diagrams generated by people with varying degrees of training in genetics (from a single introductory course to a PhD degree) while they reasoned about the process of meiosis. All of the subjects spontaneously generated diagrams during the course of their problem solving. However, the diagrams constructed by the more advanced subjects focused more on meiosis-relevant features of chromosomes as opposed to their perceptual appearance. In relating the types of diagrams drawn to the problem solving displayed, Kindfield argued that "in essence, the more advanced participants used fine-tuned diagrams as 'tools to think with'" (p. 21).

The proposal that abstract diagrams are "tools to think with" can be extended beyond the life sciences fields. For example, in the mathematical domain, one of Polya's

(1957) heuristics for problem solving is to draw a figure to represent the information (in particular the relations) presented verbally in a problem. The importance of teaching students how to represent problems nonverbally and how to extract information from such representations has been stressed by contemporary mathematics educators as well (e.g., Barwise & Etchemendy, 1991; Goldin, 1985; Lewis, 1989; Zimmermann & Cunningham, 1991).

Spatial Diagram Representations

Much of the recent research on the use of diagrams to support problem solving has focused on diagrams that are specific to narrowly defined content domains such as genetics or geometry (e.g., Kindfield 1993/1994; Koedinger & Anderson, 1990). There is less research examining people's knowledge and use of more general-purpose diagrams. In this article, we focus on three related spatial diagrams—hierarchies, matrices, and networks—that are applicable across a wide variety of contexts. For example, a hierarchy can be used to represent the "familial" relations among members of the animal kingdom, the NCAA basketball tournament pairings, or a corporate power structure. Similarly, a node-link network (path diagram) can be used to represent the flight paths for a particular airline, the food web for a particular ecosystem, or the (hypothesized) structure of semantic memory. And a matrix can be used to represent multiplication tables, Punnett squares in genetics, or a teacher's grade sheet. More generally, these types of representations are useful for solving a wide variety of problems involving analytical (including mathematical) reasoning, as well as for conceptualizing and communicating psychological theories (Butler, 1993).

Several factors make it important for psychologists to understand people's knowledge and use of these types of representations. First, Barwise and Etchemendy (1991)

Experiment 1 was presented at the 34th annual meeting of the Psychonomic Society (November 1993). We thank Cindy Hmelo for her help in writing the problems for the selection task and Casey Craven for her help in collecting the data for Experiment 2. Steve Reed, Brian Ross, Dan Schwartz, and Barry Stein provided thoughtful comments on an earlier version of the manuscript. Correspondence should be addressed to L. R. Novick, Department of Psychology and Human Development, Box 512 Peabody, Vanderbilt University, Nashville, TN 37203 (e-mail: novicklr@ctrvax.vanderbilt.edu).

have argued that nonlinguistic representation is ubiquitous in deductive reasoning both in everyday situations and in formal domains. The kind of reasoning Barwise and Etchemendy have examined most closely is typified by some of the problems found on the analytical reasoning section of the GRE. Indeed, the instructions for that section of the exam indicate that it might be useful to draw a rough diagram while answering some of the questions. Moreover, the appropriate type of representation often is one of the spatial diagrams under discussion here.

Second, because of their intermediate level of generality, spatial diagram representations highlight structural commonalities across problems that are superficially quite different (Novick, 1990; Novick & Hmelo, 1994). Thus by successfully (and appropriately) constructing such representations, problem solvers would be led to see deep similarities among diverse problems that otherwise might not be salient. There is a large literature documenting the importance of structural understanding as a key factor underlying expertise (e.g., Chi, Feltovich, & Glaser, 1981; Schoenfeld & Herrmann, 1982).

Finally, spatial diagram representations facilitate learning and problem solving because they (1) simplify complex situations and illustrations by discarding unnecessary details (see, e.g., Lynch, 1990; Winn, 1989), (2) make abstract concepts more concrete by mapping them onto spatial layouts with familiar interpretational conventions (e.g., Winn, 1989), and (3) substitute easier perceptual inferences for more computationally intensive search processes and sentential deductive inferences (Barwise & Etchemendy, 1991; Larkin & Simon, 1987). A number of studies provide empirical support for the benefits of using spatial diagram representations for learning and problem solving on specific tasks (e.g., Bartram, 1980; Carroll, Thomas, & Malhotra, 1980; Day, 1988; Guri-Rozenblit, 1988; McGuinness, 1986; Novick & Hmelo, 1994; S. H. Schwartz, 1971). For example, Day found that subjects who studied medication instructions in the form of a list did less well at answering both factual and inferential questions about the instructions than did subjects who studied the information in the form of a matrix (56% vs. 78% correct, respectively).

The Nature of Solvers' Spatial Diagram Knowledge

Despite these studies documenting the usefulness of spatial diagram representations for solving particular problems, little is known about solvers' knowledge about these representations. Bransford, Sherwood, Vye, and Rieser (1986) have stressed that in order to be good problem solvers, students need to learn conditionalized knowledge—that is, knowledge that includes information about the conditions for and constraints on its use. At least some of college students' knowledge of spatial diagram representations appears to be conditionalized in this sense. In an early study, S. H. Schwartz (1971) studied the types of representations subjects spontaneously constructed to help them solve deductive reasoning problems in which the goal was to match values on different

dimensions (e.g., to determine which prisoner is in Cell Block E). He found that subjects who used a matrix representation were more likely than those who used either a network or sentence representation to solve the problems. S. H. Schwartz and Fattaleh (1972) then conducted a study in which the problem clues were presented in one of three different formats: sentence, network, or matrix. They found that subjects were much more likely to switch to than from the more efficient matrix representation: 45% of the sentence and network subjects switched to a matrix representation, whereas only 17% of the matrix subjects changed representations.

Knowledge concerning the applicability conditions for spatial diagram representations may come from either or both of two sources. One source is specific examples of situations in which a particular type of representation has proved useful in the past. For example, our subjects typically know that matrices are used for multiplication tables, seating charts, time schedules, calendars, and police maps of a city. If representation knowledge is primarily exemplar based, then selection of an appropriate type of representation to use in a new situation would rely on the mechanism of representational transfer (Novick, 1990; Novick & Hmelo, 1994).

A second possible source of information about applicability conditions is an abstract schema for each type of representation. There is precedence for this proposal; Koedinger and Anderson (1990) found evidence for the use of diagram schemas in solving geometry proofs. A matrix schema, for example, might include the information that matrices are particularly useful when (1) there are items in two distinct sets, (2) the items within a set cannot be combined, (3) all possible combinations of the items across sets must be considered, and (4) the links between items in the different sets are nondirectional (also see Day, 1988, concerning the first and third features). An in-depth discussion of the applicability conditions for the three spatial diagram representations, and hence the nature of the abstract schemas, if they exist, is beyond the scope of this article. We are currently working on this topic (Novick & Hurley, 1997).

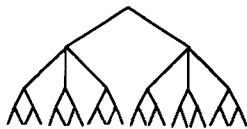
This distinction between schematic and example-specific forms of knowledge is closely related to the current debate about whether people reason and solve problems (in particular domains) using general principles or specific examples (e.g., Malt, 1989; Medin & Ross, 1989; Smith, Langston, & Nisbett, 1992). This is a complex issue whose resolution is likely to depend on the domain being considered, as well as on the level of expertise of the problem solver. It is also possible that people at a given level of expertise in a domain reason using both general principles and specific examples (e.g., Malt, 1989; Medin & Ross, 1989; Novick & Holyoak, 1991; Reed & Bolstad, 1991). This may be especially likely in domains in which abstract schemas are induced from the earlier use of examples by analogy (e.g., Anderson, Farrell, & Sauers, 1984; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Novick & Holyoak, 1991; Ross & Kennedy, 1990).

With respect to the spatial diagrams of interest here, our goal is not to try to pit the exemplar and schema views of knowledge representation against each other. We assume that people have exemplar knowledge, and our own research (Novick, 1990; Novick & Hmelo, 1994) has shown that they are able to transfer such knowledge to new situations. Rather, our goal is to determine whether people also have abstract, schematic knowledge concerning the applicability conditions for the three spatial diagram representations.

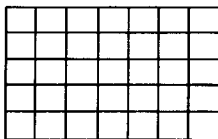
Representations With and Without Canonical Spatial Embodiments

We believe that solvers might have such knowledge about these representations, in part because of S. H. Schwartz and Fattaleh's (1972) results, in part because students from certain academic backgrounds are able to verbalize some of the applicability conditions (Novick & Hurley, 1997), and in part because each of these representations has a consistent, canonical spatial embodiment (e.g., see the diagrams shown in Figure 1) to which it would be easy to "attach" the applicability conditions. The existence of a canonical spatial embodiment means that specific instances of these representations can be generated as simple parametric (i.e., quantitative) variations of each other or of a prototype. Other important types of representations (e.g., parity, parts and wholes) do not have a consistent, canonical spatial embodiment. For example, as shown in Figure 2, the part-whole rep-

A hierarchy or branching structure



A matrix with rows and columns



A network or system of paths

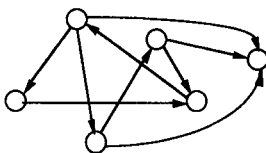


Figure 1. Representation category probes used in Experiments 1-3 for the three spatial diagram representations.

Wholes divided into parts

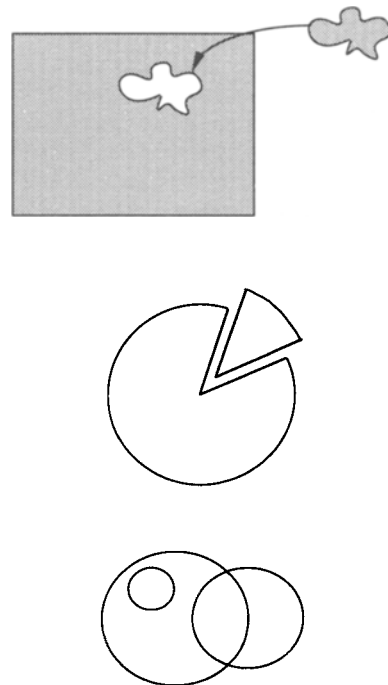


Figure 2. Alternative representation category probes for the part-whole representation. The final probe was used in Experiments 1 and 2.

resentation, which we used in our earlier work on representational transfer (Novick & Hmelo, 1994), has a variety of qualitatively different spatial embodiments. Which one, if any, is most appropriate depends on the type of part-whole relation being considered (see Winston, Chaffin, & Herrmann, 1987). Because we used the part-whole representation in our earlier research, we included it in Experiments 1 and 2 as a contrast case. We excluded it from Experiment 3, however, because it is not needed for drawing conclusions about the three spatial diagram representations. Moreover, differences between representations with and without canonical spatial embodiments, although interesting, are not of central concern in this article.

Nevertheless, we want to briefly explain why we believe the part-whole representation is different in kind from the three spatial diagram representations, because we did not necessarily expect the same results for that representation as for the others. The results of some pilot work in our laboratory (Francis, 1995) support this distinction. Subjects were given one of two different diagrams and a verbal label for each of the four representations, and their task was to generate as many examples as they could of situations for which that type of representation could be used to organize information. We coded the number of appropriate examples, as well as the number of different kinds of examples, generated (the coding scheme will be described later). For the three spatial

diagram representations, performance was comparable across the three category probes, suggesting that the two versions of the spatial diagram and the verbal label caused subjects to access a single, at least semicoherent set of knowledge about the corresponding representation.

In contrast, for the part-whole representation, subjects generated more examples and more different kinds of examples when given the verbal label as opposed to a diagram (one of the first two shown in Figure 2). In addition, performance varied for the two diagrams. These results suggest that the totality of one's knowledge about parts and wholes cannot adequately be captured by a single, canonical spatial diagram. Moreover, the superiority of performance in the verbal label condition suggests that subjects' knowledge about this type of representation may not be tied to a spatial embodiment at all. This conclusion seems consistent with the varied interpretations of parts and wholes discussed by Winston et al. (1987). In fact, some of their interpretations seem purely conceptual, without a possible spatial embodiment.

Overview of the Present Research

To assess the nature of people's knowledge concerning the applicability conditions for the three spatial diagram representations, we devised what we will refer to as the *selection task*. In this task, we asked subjects to select the most appropriate type of representation, from a list of alternatives presented, for each of 12–15 short story problems (the number of problems varied across studies). We manipulated across subjects the type of information they were cued to use to help them make their representation selections: (1) whatever general information they had in memory about these types of representations (Experiments 1–3); (2) a single, detailed example problem for each representation that we provided for them (Experiments 1 and 3); or (3) both general category information and the specific examples (Experiment 1). Briefly, our prediction was that if subjects had abstract schematic knowledge about the three spatial diagram representations, their selection accuracy would be greater when they were led to rely on that knowledge than when they were led to rely on the specific examples we provided. On the other hand, if subjects did not have abstract schematic knowledge in memory, the opposite pattern of results would be obtained. These predictions will be discussed in more detail shortly.

It is important to note that our question concerning solvers' knowledge is really one of their competence rather than their performance. It is quite possible that students have some rudimentary knowledge of the applicability conditions for the three spatial diagram representations, but because that knowledge is somewhat fragmented and not completely specified, they are not always able to bring it to bear in solving relevant problems. Thus we do not take subjects' absolute level of performance in selecting appropriate representations on our multiple choice test as a measure of their ability to select appropriate representations when such cues are not available.

The results of Experiment 1 suggest that college students do have some abstract, schematic knowledge in memory about the three spatial diagram representations. In particular, when they were led to base their representation selections on this information, they were more accurate than when they were led to base their selections on detailed worked-out examples. In Experiments 2 and 3, we cast doubt on three alternative explanations of the results of Experiment 1. Experiment 3 also replicates the results of Experiment 1 using a very different method of cuing subjects to access the abstract knowledge they have in memory.

EXPERIMENT 1

Method

Subjects. Sixty Vanderbilt University undergraduates participated in partial fulfillment of course requirements. Ten men and 10 women were randomly assigned to each of three conditions.

Selection task. As indicated earlier, on the selection task subjects had to decide which, if any, of the four types of representations was most appropriate for each of 15 story problems. Each type of representation provided the best structural model for three problems. For the remaining three problems, none of the four representations was appropriate. The correct answer for each problem was determined by considering factors such as whether the concepts were combined factorially and whether pairs of concepts were linked by only one versus multiple paths (see Novick & Hurley, 1997). It has been our experience (also see Novick & Hmelo, 1994) that in many situations the type of representation that provides the best structural model is unambiguous. We tried to choose such problems for this task, as indicated in Appendix A, where we have reprinted one selection problem from each category.

Subjects were told to select the type of representation they thought would be most helpful for understanding each problem. If they thought that more than one type of representation could be used, they were to pick the one that best captured the structure of the problem. If they thought that none of the representations was appropriate, they were to select "none of the above." Subjects rated their confidence in their choices on a 4-point scale (1 = *not at all sure*; 4 = *very sure*). Subjects indicated their responses on a separate answer sheet. The four types of representations, and the none-of-the-above alternative, were indicated at the top of the answer sheet. Exactly how the representation choices were presented varied across conditions, as described in the Procedure section. Subjects were given 1 min, 45 sec to make their selection for each problem.

The 15 problems were collated into booklets in one of four different orders. The orders were essentially random, except for the constraint that the two deductive reasoning problems had to be separated by at least four other problems.¹ To help subjects become familiar with the selection task, the 15 experimental problems were preceded by two practice problems. "None of the above" was the correct representation choice for both of those problems. Subjects were given 2 min to make their selection for the first practice problem. They were given 1 min, 45 sec for the second practice problem, just as they were for the experimental problems.

Design. We varied, across subjects, the type of knowledge subjects were cued to use to help them determine the best type of representation for the problems on the selection task: general category, specific example, and category + examples.

In the general category condition, we gave subjects a task that was designed to help them access whatever general knowledge they had in memory about each type of representation. Subjects received a category probe for each type of representation, which consisted of a verbal label and a diagram (see Figure 1 and the last diagram in

Figure 2). Their task was to generate as many examples as they could of situations for which the indicated type of representation could be used to organize information. Because the labeled diagrams were intended to stand for the type of representation in general, subjects were told that they might need to modify the diagrams somewhat to make them work for their examples. Subjects were given 5 min to generate examples for each type of representation. Given our pilot work (Francis, 1995), we felt that the diagram/label combination provided the best cue we could devise for accessing the breadth of subjects' knowledge about these types of representations (as indicated by the number of appropriate examples and the number of different kinds of examples generated). Because neither diagram used for the part-whole representation in the earlier work was particularly effective, we created a new diagram involving overlapping and embedded circles that seemed more relevant to the problems on the selection task.

In the specific example condition, subjects were cued to focus on a single detailed example problem illustrating the use of each type of representation. This was accomplished by giving subjects four example problems to solve—one for each type of representation. A partially constructed representation was provided with each problem, and the subjects' task was to complete the representation and then use it to solve the problem. Subjects were given 5 min to solve each problem. The problems were pretested to ensure that subjects understood how to complete the representations and were able to use them to get the answers. Previous work, using three of the four example problems (all except the hierarchy problem, which is reprinted in Appendix B), has shown that college students are able to use these problems as the basis for transferring the indicated type of representation to new problems like those used on the selection task (Novick & Hmelo, 1994). Moreover, representational transfer from a specific example problem has been found for all four types of representations (Novick, 1990; Novick & Hmelo, 1994). Thus it seems reasonable to believe that subjects would find the example problems helpful for completing the selection task.

The hierarchy example involved determining a player's starting level in a fantasy game. Different numbers of points were awarded depending on the outcome of spinning a spinner, then rolling a die, then flipping a coin. Different ranges of total points indicated different starting levels, and subjects had to determine how many different combinations of actions enabled the player to start the game at each level. The matrix example involved a girl who was going to get one of three pairs of pants and one of six shirts for her birthday. She preferred certain combinations of shirts and pants, and subjects had to determine the probability that she would get an outfit she preferred. The network example involved a couple who were planning a trip to some tropical islands. Various of the islands were connected by bridges, and the couple wished to go over each bridge exactly once. Subjects were to plan a route to accomplish this. For the part-whole example, subjects were given a list of all the children in a fourth-grade class who collected rocks and a list of all the children in the class who collected shells. They were to determine who collected only rocks, who collected only shells, and who collected both rocks and shells.

Finally, in the category + examples condition, subjects were cued to use both their general knowledge of the four types of representations and the specific example problems. Subjects generated examples for each type of representation; then they solved the four example problems.

Procedure. Subjects participated in small groups of 2–6 (except for 1 person who participated individually) in a single session that lasted approximately 90 min. By necessity, all subjects in a group participated in the same condition.

Because subjects in the general category and specific example conditions completed only one knowledge cue task, it is possible that they could have had an advantage on the selection task because

they would have been less fatigued. Therefore, subjects in those two conditions began the experiment by completing two filler tasks that together took approximately the same amount of time as each knowledge cue task. For the first task, subjects spent 12 min rating the pronounceability of some anagrams. Then they were given 10 min to unscramble 20 anagrams.

After the anagram solution task, subjects in the general category and specific example conditions completed either the example generation or the problem solution task, respectively. When the knowledge cue task was introduced, subjects were told that we were interested in the different types of representations people use to help them understand and solve problems. They were further told that their main task was going to be to determine which of several types of representations would be most helpful for understanding a variety of problems. So that subjects would know what we meant by *type of representation*, they were given the example shown in Figure 3a. They were also shown how this type of representation could be used in several specific contexts (Figure 3b). The cue task (example generation or problem solution) was then introduced as a preliminary task to help subjects become familiar with the four types of representations we were interested in for the main task. The order for presenting the four labeled diagrams or example problems was counterbalanced across subjects so that each representation was encountered equally often in each position and was preceded and followed by each other representation equally often.

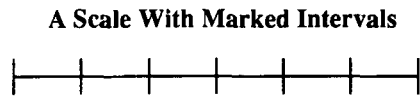
Subjects in the category + examples condition were told that they would be completing two preliminary tasks to help them become familiar with the four types of representations. They did the example generation task first, followed by the problem solution task, so that the problems would not affect the examples generated. Representation order was counterbalanced, as in the other two conditions. Subjects encountered the four representations in the same order for the two tasks.

After completing the knowledge cue task(s), subjects completed the selection task. The representation choices for this task were presented differently for subjects in the different conditions. For the general category condition, we printed the four category probes (i.e., the labeled diagrams) at the top of the selection task answer sheet. For the specific example condition, we printed short names referring to each of the example problems (namely, fantasy game example, clothing example, islands example, rock and shell example) at the top of the answer sheet. We also provided reduced-size copies of the problems themselves (clean copies, not the ones subjects solved) for reference. For the category + examples condition, the labeled diagrams and the example problem names appeared at the top of the answer sheet. As in the specific example condition, subjects in the category + examples condition also received the reduced-size versions of the example problems. In all conditions, the representation choices given at the top of the selection answer sheet were printed in the same order in which the subject encountered them during the preliminary task(s). The none-of-the-above option was always printed last.

Predictions

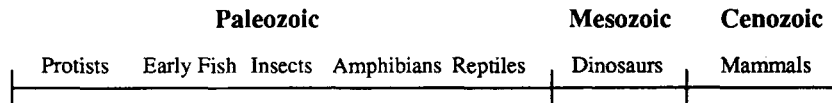
We are interested in investigating whether typical college students have some general (i.e., relatively abstract) knowledge about spatial diagram representations that includes at least rudimentary information about their applicability conditions, or whether their ability to determine the appropriate representation for a problem depends solely on the availability of an appropriate example. The latter hypothesis of solely example-specific reasoning yields two predictions concerning the results of the present experiment: First, we should see very poor selection

(a) Labeled diagram for the example representation:

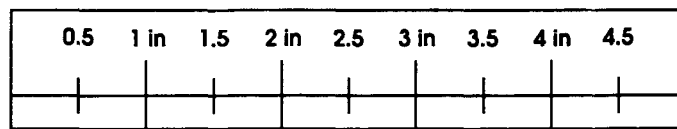


(b) Applications given for the example representation:

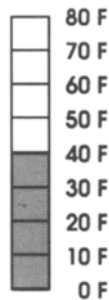
1. A TIME LINE



2. A RULER



3. A THERMOMETER



4. A SPEEDOMETER

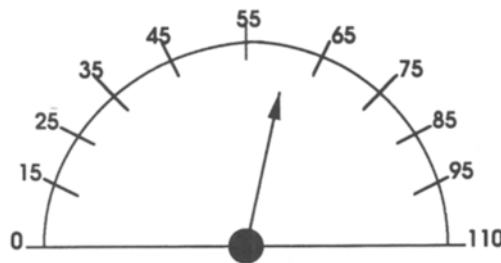


Figure 3. The labeled diagram (a) and applications (b) for the example representation used in Experiments 1-3.

performance in the general category condition, perhaps near chance for some representations, because whatever category information subjects have does not include information about the applicability conditions for the representations. Second, performance in the specific example condition should be better than performance in the general category condition, because even if subjects do not have (accessible) abstract knowledge about when to use the various representations, as argued earlier they should be able to compare the selection task problems to

the readily available, and appropriate, example problems and make their selections on the basis of similarity (see Novick, 1990, and Novick & Hmelo, 1994, for relevant data on this issue).

On the other hand, the hypothesis that subjects have rudimentary schemas for the three spatial diagram representations that include some information about their applicability conditions yields a different set of predicted results. First, selection performance in the general category condition should be significantly above chance. Sec-

ond, performance in that condition should be better than performance in the specific example condition, because the diagrams presented as part of the category probes provide a good summary of the canonical spatial embodiment for each type of representation. Thus they should be effective in helping subjects access the breadth of their knowledge about those representations. Although relying on a single appropriate example to guide representation selections may be effective, relying on more generalized knowledge should be more effective. Third, performance in the general category condition is likely to be significantly below ceiling, reflecting the rudimentary nature of subjects' knowledge of the applicability conditions.

These predictions are based on the fact that abstract schemas provide a more reliable source of knowledge for problem solving than do specific examples. As discussed by Holyoak (1985), abstract schemas do not contain the potentially irrelevant information that specific examples do. Abstract (i.e., context-free) schemas contain only structurally relevant information, so when reasoning using abstract schemas, people need to concern themselves only with whether the current problem fits the structure of the schema. In contrast, if people reason about a problem with reference to a specific example, they must take into account both the structural similarities and the superficial differences between the two problems. Not only does considering superficial differences between the problems (e.g., between their cover stories) add an extra step to the comparison process, but also it may interfere with the attempt to align the problems' structures. Hence schema-based transfer should be more effective than example-based transfer (Holyoak, 1985).

The three problems for which the correct answer is "none of the above," however, do not belong to a coherent category. The only thing they have in common is that none of the four cued representations is helpful, and the example generation task does not cue subjects to access any abstract schemas they might have that would be relevant for these problems. Thus, for this set of problems, there is no basis for predicting any difference between the general category and the specific example conditions.

We expect a similar pattern of performance for the part-whole problems and for the none-of-the-above problems, because, as discussed earlier, no single diagram is able to span the entire category for that type of representation. To summarize, the schema-based reasoning hypothesis would lead us to expect an interaction between the type of representation (spatial vs. nonspatial) and the cue condition.

Results

The results provide support for the hypothesis that subjects have rudimentary schemas for the three spatial diagram representations that include some information about their applicability conditions over the hypothesis that subjects rely solely on example-specific reasoning to select the appropriate representation for a problem. Several aspects of our data support this conclusion.

Absolute level of performance on the selection task.

First, we compared accuracy on problems requiring spatial versus nonspatial representations to both chance and maximum performance. The observed means and *t* statistics for each condition are shown in Table 1. In all cases, actual performance was reliably better than would be expected by chance but reliably worse than the maximum. More importantly, this pattern was obtained in the general category condition for each of the three spatial representations [for the comparisons to chance, smallest $t(19) = 4.75, p < .001$, and for the comparisons to the maximum, smallest $t(19) = -4.00, p < .001$].

Selection task performance as a function of condition and type of representation. Our critical analysis of performance for spatial versus nonspatial representations across conditions was based on discrimination scores. These scores provide a more fine-grained indication than does simple accuracy of a subject's ability to discriminate the situations for which each type of representation is most appropriate. The discrimination score for each representation was calculated as the proportion of appropriate choices of that representation minus the proportion of inappropriate choices. For example, for the matrix representation, the proportion of appropriate choices was computed by counting the number of times

Table 1
Observed Performance on the Selection Task
in Each Condition of Experiment 1 Compared to
Chance Expectation and to Maximum Performance

Type of Representation*	Proportion Correct	Comparison to Chance†	Comparison to Maximum
Specific example			
Spatial	.36	$t(19) = 6.06, p < .001$	$t(19) = -24.03, p < .001$
Nonspatial	.56	$t(19) = 9.25, p < .001$	$t(19) = -11.40, p < .001$
General category			
Spatial	.58	$t(19) = 9.73, p < .001$	$t(19) = -10.58, p < .001$
Nonspatial	.50	$t(19) = 7.16, p < .001$	$t(19) = -11.94, p < .001$
Category + Examples			
Spatial	.61	$t(19) = 15.74, p < .001$	$t(19) = -14.89, p < .001$
Nonspatial	.61	$t(19) = 10.54, p < .001$	$t(19) = -10.11, p < .001$

*Spatial refers to hierarchy, matrix, and network, whereas nonspatial refers to part-whole and none of the above. †Chance performance is .20 correct.

the subject selected the matrix representation for matrix problems and dividing that sum by three, which is the number of matrix problems. Similarly, the proportion of inappropriate choices was computed by counting the number of times the subject selected the matrix representation for nonmatrix problems and dividing that sum by 12, which is the number of nonmatrix problems. On this scale, a score of 0 represents chance performance, and a score of 1 represents perfect discrimination. (A score of -1 indicates consistent confusions among all the representations.) Collapsed across representations and conditions, the mean discrimination score was .42.

For our analysis, we averaged the discrimination scores across representations to get means by condition for both spatial (hierarchy, matrix, network) and nonspatial (part-whole, none-of-the-above) representations. We then conducted a 3 (condition) × 2 (type of representation) mixed analysis of variance (ANOVA) with repeated measures on the second factor. There was a reliable effect of condition [$F(2,57) = 8.68, MS_e = 0.0430, p < .001$] and a marginally reliable effect of type of representation [$F(1,57) = 3.40, MS_e = 0.0237, p < .08$]. More importantly, these two factors interacted [$F(2,57) = 8.81, p < .001$], as shown in Figure 4.

To determine whether the interaction followed the pattern predicted by the hypothesis that subjects have rudimentary schemas for the three spatial diagram representations, we performed a separate contrast analysis for

performance on each type of representation. In each analysis, we tested two orthogonal contrasts. The first contrast compared performance in the general category and specific example conditions. Recall that the prediction is that there should be a difference between these two conditions, in favor of the general category condition, for the spatial representations only. As expected, performance was reliably better in the general category condition than in the specific example condition for the spatial representations [$F(1,57) = 23.31, MS_e = 0.0271, p < .001$]. But for the nonspatial representations, there was no difference between the two conditions [$F(1,57) < 1, MS_e = 0.0395$].

The second contrast compared performance in the category + examples condition with that in the other two conditions combined. For the nonspatial representations, subjects who were cued to make their representation selections on the basis of both their general knowledge and the specific example problems showed better discrimination than subjects who received either type of cue alone [$F(1,57) = 4.19, p < .05$]. That is, adding the detailed example problems to subjects' general knowledge and adding subjects' general knowledge to the detailed example problems improved performance. For the spatial representations, it was also the case that performance in the category + examples condition was better than the average performance in the two single-cue conditions [$F(1,57) = 13.10, p < .001$]. In this case, however, the effect can be attributed entirely to the benefit of adding subjects' gen-

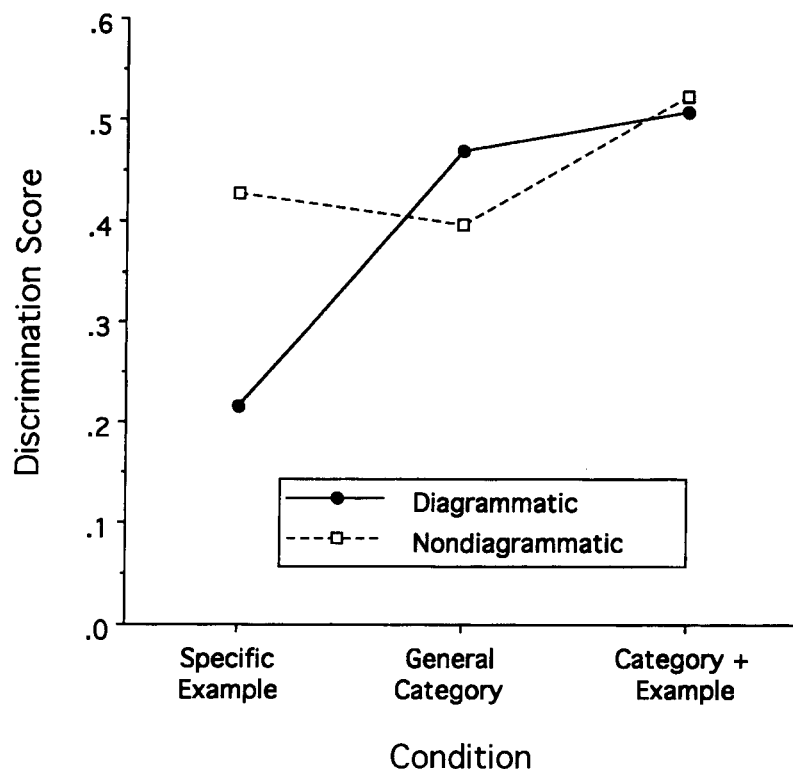
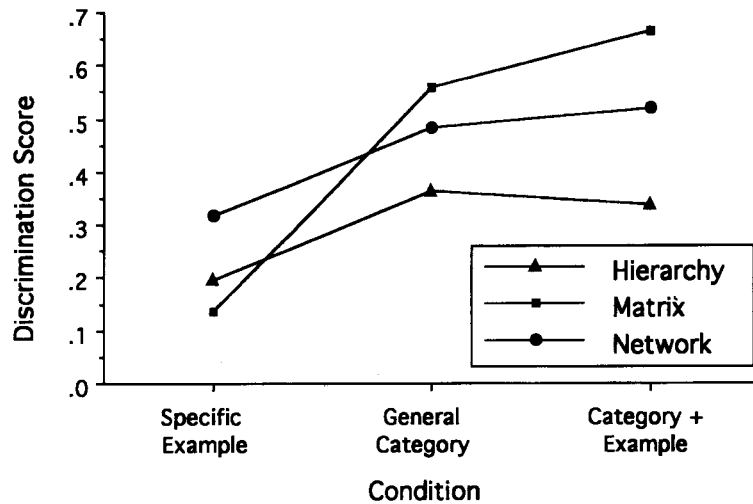


Figure 4. Mean discrimination scores for spatial and nonspatial representations as a function of experimental condition in Experiment 1.

(a) Spatial Representations:



(b) Nonspatial Representations:

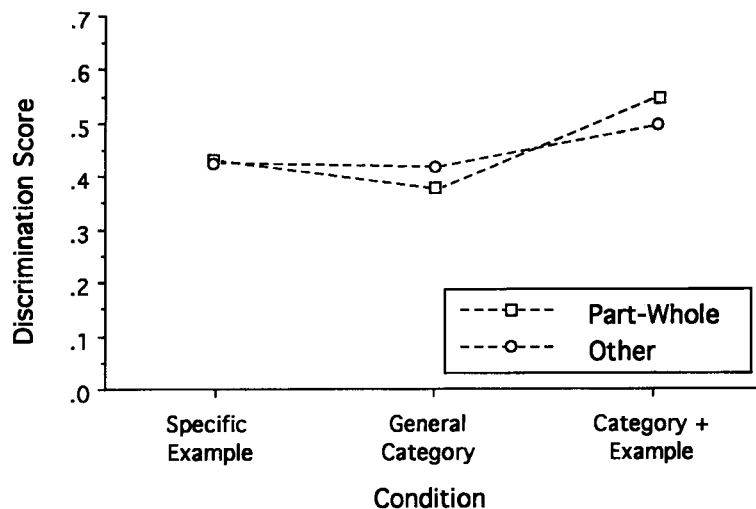


Figure 5. Mean discrimination scores for each type of representation as a function of experimental condition in Experiment 1. The results for the spatial representations are shown in (a), and the results for the nonspatial representations are shown in (b).

eral knowledge to the example problems, because there was no difference between the general category and category + examples conditions [$F(1,38) < 1$, $MS_e = 0.0285$]. Figure 5 shows that the patterns observed for the spatial and nonspatial representations hold for each of the specific representations entering into the composite measures.

Performance on the example problems. It is important to note that the poor performance in the specific example condition for the spatial representations was not due to our use of examples that subjects could not understand. For each example problem, we coded separately whether subjects (1) constructed the representation correctly and (2) got the correct answer. For the matrix, net-

work, and part-whole representations, each measure was scored on a 0/1 scale. Because the hierarchy problem was longer and required more work, and subjects' responses were therefore more variable, we added a partial credit score of 0.5 for each measure for that representation. As shown in Table 2, performance was quite high for both measures for all problems. Averaged across the three problems with spatial representations, 93% of the subjects used the representations correctly, and 76% of the subjects correctly solved the problems. Also, as mentioned earlier, three of the four example problems had been used previously in transfer experiments (and a different hierarchy example was used as well), and the subjects in

Table 2
Performance on the Example Problems in Experiments 1 and 3

Representation/Measure	Experiment 1	Experiment 3
Hierarchy		
Representation use	.92	.73
Accuracy	.70	.62
Matrix		
Representation use	.93	1.00
Accuracy	.83	.86
Network		
Representation use	.93	.95
Accuracy	.75	.86
Part-whole		
Representation use	.97	—
Accuracy	.95	—

Note—The part-whole representation was not used in Experiment 3.

those experiments were able to use each of the example problems as the basis for representational transfer (Novick, 1990; Novick & Hmelo, 1994).

Discussion

In interpreting the results of this experiment, we have assumed that asking subjects to generate examples of situations for which each type of representation could be used to organize information is an effective way of getting them to access whatever general information they have in memory about these types of representations. Under this assumption, the results support our hypothesis that subjects have rudimentary schemas for the three spatial diagram representations that include some information about their applicability conditions. The key finding in this respect is that performance was better in the general category condition than in the specific example condition for the spatial but not the nonspatial representations.

However, the generation task we used in the general category condition differs in multiple ways from the problem solution task we used in the specific example condition, so a variety of alternative explanations for the superior performance in the general category condition are possible. We consider three such explanations here, which will be tested and rejected in Experiments 2 and 3. One possibility is that subjects were able to induce some of the applicability conditions for the spatial diagram representations by generalizing over the examples they generated. According to this explanation, selection performance in this condition was based on abstract knowledge, as we hypothesized, but that knowledge was learned in the experiment rather than being preexisting. This explanation seems implausible to us, because we know of no other study in which subjects were able to do significant generalization and abstraction in less than 5 min (figuring that some of the 5 min was spent in generating the examples), especially without a specific request to attempt to do so. Nevertheless, we evaluated this explanation more formally in Experiment 3 by reducing the example generation time by 70% from 5 to 1.5 min. If the good performance in Experiment 1 was due to learning

that occurred during the generation task, performance in this condition in Experiment 3 should have been considerably lower due to the large reduction in the time available for learning.

A second alternative explanation comes from the observation that in the general category condition, subjects spent much of their time linking specific examples to particular types of representations. Because this is arguably what is required for good performance on the selection task, one might expect better performance on that task in this condition due to transfer appropriate processing (Morris, Bransford, & Franks, 1977). We evaluated this explanation in Experiment 3 by adding a "purer" knowledge elicitation task in which subjects were asked to spend 20 sec familiarizing themselves with each type of representation.

A third alternative explanation is that subjects used the examples they generated as the basis for representational transfer. That is, subjects transferred their knowledge about these particular cases to the problems on the selection task. Because subjects in the general category condition had more examples for each representation to use as the basis for transfer than did subjects in the specific example condition (who had only one example per representation), it makes sense that they would perform better on the selection task. Although the absence of a difference between these two conditions for the part-whole representation is somewhat problematic for this alternative explanation, this account of the results seems to have considerable face validity. Averaged across the four types of representations, subjects generated 3.60 appropriate examples per representation, which was 74% of the total number of examples generated. Although the results varied somewhat across representations, all representations had more than just a single appropriate example generated ($M = 4.55$ for the hierarchy representation, $M = 3.40$ for the matrix representation, $M = 4.10$ for the network representation, and $M = 2.35$ for the part-whole representation). We evaluated this example-based reasoning explanation in Experiment 2.

EXPERIMENT 2

If subjects use the examples they generate as the basis for representational transfer, then variability in characteristics of the examples generated should predict variability in performance on the selection task. For example, (1) subjects who generate a larger number of appropriate examples should do better because they have more possibilities for successful transfer; and (2) to the extent that different types of examples can be identified for each representation, subjects who generate a larger proportion of the possible types should do better because it is more likely that they will have generated a type that matches one of the selection problems. In contrast, if the example generation task gets subjects to access the relevant part of their abstract knowledge base, and they use that abstract knowledge about the properties of the represen-

tations to guide their later representation selections, then we would not expect reliable correlations between aspects of the examples generated and selection performance. This prediction follows from the fact that the abstract properties of a representation (e.g., for the matrix, two sets of items, factorial combination of items across sets, and nondirectional links between items) apply equally to any of the examples of that representation (e.g., multiplication table, restaurant seating chart, calendar, or police map), so more examples generated does not translate into more abstract properties accessed.

It is difficult to provide a strong test of these hypotheses with the Experiment 1 data because there were only 20 subjects in the general category condition. If we were to find null results, as predicted by the schema-based reasoning hypothesis, this could easily be attributed to the small sample. Combining the data across the general category and category + examples conditions is problematic, because in the latter condition selection task performance also might be affected by the examples we gave subjects. Therefore, to provide a better (more powerful) test of the example-based reasoning explanation, we conducted a second experiment in which all subjects participated in the general category condition. We will report the correlational results for the Experiment 1 general category condition alongside the results of Experiment 2.

Method

Subjects. The subjects were 38 Vanderbilt University undergraduates (17 women and 21 men). They were paid \$15 to participate in a single session that lasted approximately 2 h, 15 min.

Materials and Procedure. As in Experiment 1, subjects participated in small groups. The generation and selection materials were the same as those used in Experiment 1. To reduce the length of the experimental session, subjects were given 4 min to generate examples for each type of representation rather than 5 min. The generation probes were presented in one of four orders, as before. After the generation task, subjects completed the selection task, which was conducted exactly as described earlier. Each of the four generation probe orders was combined with each of the four orders for presenting the selection problems. Two to 3 subjects received each combination. After the selection task, subjects completed several other tasks that were part of a separate investigation. Only the generation and selection data are relevant to the issue being addressed here.

Results

Example-based reasoning predictions. This experiment was designed to determine whether the example generation task facilitated performance on the selection task by providing exemplars that could be compared with the selection problems. If this example-based reasoning explanation of the Experiment 1 results is correct, we would expect characteristics of the examples generated to predict performance on the selection task. In particular, subjects who generated a greater number of appropriate examples, or more different categories of examples, should be more likely to have generated examples that were similar to the selection problems and therefore should have performed better on the selection task. Thus for each type of representation, we counted the number

of appropriate examples that were generated, as well as how many different categories of examples were generated. A more fine-grained analysis looked at the relation between generation responses of a particular type and performance on selection problems of that same type.

Coding the generation data. On the basis of examples generated by pilot subjects, we developed a coding scheme that had four to seven categories of examples for each type of representation. The different numbers of categories for the different representations reflect differences in the diversity of the examples generated by our subjects. We make no claim that the categories coded here constitute all possible types of situations for which each type of representation is appropriate. Detailed coding criteria for each of the categories may be obtained from L.R.N. A short description of each category is provided in the next paragraph.

There were seven types of hierarchy examples: descent (e.g., family tree), categorical classification (e.g., trees broken down into deciduous and coniferous and then into oaks, firs, etc.), power structure (e.g., division of power in a corporation), subcomponents (e.g., mathematical factoring: 32 is the product of 8 and 4, 4 is the product of 2 and 2), series of choices (e.g., decision tree), exponential growth (e.g., chain letter), and elimination (e.g., basketball tournament bracket). There were five types of matrix examples: binary (e.g., restaurant seating chart to keep track of which tables have been seated), logical consequence (e.g., multiplication table), associated characteristic (e.g., daily planner), game (e.g., chess board), and graphs, grids, and maps (e.g., grid for pinpointing military targets). There were four types of network examples: maps (e.g., airline routes), associative relationships (e.g., diagram of alliances among a group of nations), flow through a network (e.g., flow chart for a computer program), and directed movement (e.g., diagram of a football play). Finally, there were six types of part-whole examples (based on the analysis provided by Winston et al., 1987): portion/mass (e.g., an hour is part of a day), component/integral-object (e.g., a wing is part of a bird), member/collection (e.g., a juror is part of a jury), place/area (e.g., an oasis is part of a desert), feature/activity (e.g., proofreading is part of writing a paper), and stuff/object (e.g., sugar is part of apple pie).

L.R.N. and M.F. independently coded the generated examples from 30 subjects (from Experiments 1 and 2, plus the experiment reported by Francis, 1995). The agreement between coders was fairly high: a mean correlation between coders (collapsed across the four representations) of .81 for the number of appropriate examples (i.e., tokens) generated, a mean correlation of .76 for the number of different categories (i.e., types) of examples generated, and a mean level of agreement on the categorical coding of individual responses of 78%. Analyses of the generation data from Experiment 1 were based on M.F.'s coding of the data. Because the generally low correlations in that experiment could have been due in part to unreliability in the coding, especially given that

Table 3
Descriptive Data for the Number of Tokens and the Proportion of Types
Generated for Each Representation in Experiment 2 and in the General
Category Condition of Experiment 1

Representation	Number of Tokens				Proportion of Types			
	Experiment 2		Experiment 1		Experiment 2		Experiment 1	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hierarchy	3.66	1.55	4.55	1.85	.36	.15	.41	.12
Matrix	3.61	1.78	3.40	1.27	.46	.19	.36	.14
Network	3.08	1.60	4.10	2.22	.53	.23	.54	.30
Part-whole	2.08	1.19	2.35	1.50	.18	.12	.19	.08
<i>M</i>	3.11	1.10	3.60	0.92	.38	.11	.38	.12

Note—*N* = 38 for Experiment 2; *N* = 20 for Experiment 1.

one set of analyses we conducted required agreement at the level of the number of examples generated in specific categories, all of the generated examples from Experiment 2 were coded independently by both L.R.N. and M.F., and disagreements were resolved by discussion.² As will be shown below, these coding differences seem to be largely irrelevant, because the results are remarkably consistent across studies for all measures.

Table 3 shows the number of tokens and types generated for each representation by all subjects in Experiment 2 (*N* = 38) and by the general category subjects in Experiment 1 (*N* = 20). The types measure is expressed as a proportion because the number of possible types varied across representations. As mentioned earlier, an average of 74% of the examples generated per representation by the Experiment 1 general category subjects were appropriate. In Experiment 2, an average of 81% of the examples generated were appropriate.

Global correlations between the generation and selection tasks. One set of analyses correlated the number of tokens (i.e., appropriate examples) generated for a particular representation with the selection discrimination score for that representation. A second set of analyses correlated the number of different types of examples generated with selection performance. The results of these analyses for both experiments are shown on the right-hand side of Table 4. Table 4 also provides descriptive data for the selection task for both experiments. The correlations can be summarized simply: None were reliably

different from zero. Moreover, correlations that were positive and relatively large in one experiment were either zero or slightly negative in the other experiment. Thus, neither set of analyses provides any support for the example-based reasoning explanation of selection performance in the general category condition of Experiment 1.

A finer grained comparison of performance on the generation and selection tasks. It is possible, however, that a finer grained analysis might yield evidence of example-based reasoning. One might argue that generating many examples or many different types of examples will be helpful only if the examples generated belong to the same categories as the problems on the selection task. Therefore, we categorized the selection problems using the generation coding guides described earlier. The three hierarchy problems belonged to three different categories: descent, series of choices, and categorical classification. The matrix problems belonged to two different categories: associated characteristic (two problems) and logical consequence. The network problems belonged to two different categories: associative relationships and flow through the network (two problems). The part-whole problems belonged to three different categories: member/collection, component/integral-object, and place/area. A final set of analyses correlated the number of examples generated of a particular type with accuracy on selection problems of that type (computed as proportion correct to maintain the same scale across categories). In addition to displaying these correlations, Table 5 pro-

Table 4
Performance on the Selection Task and Correlations Between the Generation and
Selection Tasks for Each Representation in Experiment 2 and in the
General Category Condition of Experiment 1

Representation	Selection Task				Correlation Between		Correlation Between	
	Discrimination Score				Tokens and Selection		Types and Selection	
	Experiment 2		Experiment 1		Experiment 2	Experiment 1	Experiment 2	Experiment 1
Hierarchy	.26	.29	.36	.27	.07	.12	.27	-.08
Matrix	.61	.32	.56	.34	-.02	.33	-.15	.15
Network	.32	.20	.48	.31	.14	-.26	.01	.05
Part-whole	.46	.32	.37	.26	.23	-.04	.15	-.19

Note—In Experiment 2, the critical value of *r* for *p* < .05 with *df* = 36 is approximately .32. In Experiment 1, the critical value of *r* for *p* < .05 with *df* = 18 is approximately .44.

Table 5
Correlations Between the Number of Tokens of a Particular Type Generated and Accuracy on Selection Problems of That Type in Experiment 2 and the General Category Condition of Experiment 1

Representation	Type of Example	No. Tokens Generated of This Type				Proportion Correct on This Type of Selection Problem				Token by Selection Correlation	
		Experiment 2		Experiment 1		Experiment 2		Experiment 1		Experiment 2	Experiment 1
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Hierarchy	Descent	1.16	0.82	1.50	0.76	.61	.50	.85	.37	.29	-.09
	Series of choices	0.34	0.58	0.25	0.91	.26	.45	.25	.44	.16	.36
Matrix	Categorical classification	0.26	0.55	1.00	1.12	.29	.46	.30	.47	.01	.30
	Associated characteristic	1.74	1.13	1.30	1.49	.82	.32	.70	.38	.09	.07
Network	Logical consequence	0.47	0.60	0.35	0.59	.66	.48	.80	.41	.11	-.13
	Associative relationships	0.42	0.68	0.50	0.69	.39	.50	.45	.51	.29	-.23
Part-Whole	Flow through the network	1.29	0.96	1.85	1.42	.47	.31	.60	.31	.26	-.02
	Member/collection	0.24	0.49	0.25	0.55	.47	.51	.45	.51	.41*	.51*
	Component/integral-object	0.66	0.94	0.70	1.17	.63	.49	.65	.49	.07	-.10
	Place/area	0.16	0.44	0.05	0.22	.53	.51	.30	.47	.10	-.15

* $p < .05$.

vides descriptive data for both experiments for the number of tokens generated and the proportion of correct selections.

Again, the correlational results are easy to summarize: With only one exception, discussed in the next paragraph, none of the more specific correlations were reliably different from zero. And again, correlations that were positive and relatively large in one experiment were small or negative in the other experiment. Thus the more fine-grained analysis also fails to support the example-based reasoning explanation of selection task performance, at least for the three spatial diagram representations that are of primary interest here.

The only reliable correlation between performance on the generation and selection tasks involved the part-whole member/collection category. As predicted by the example-based reasoning explanation, the number of member/collection examples generated was reliably associated with accuracy on that type of selection problem ($r = .41$, $p < .02$, in Experiment 2; $r = .51$, $p < .03$, in Experiment 1; the member/collection selection problem is printed in Appendix A). However, there was no relation between performance on the generation and selection tasks for the other two types of part-whole selection problems. These results will be discussed in the next section.

Discussion

In interpreting the results of Experiment 1, we made the assumption that the generation task facilitated performance on the selection task for the three spatial diagram representations (hierarchy, matrix, and network) by helping subjects access their general knowledge about (i.e., abstract schemas for) those types of representations. That is, we interpreted the results as providing evidence for the existence of schemas for the three spatial diagram representations that contain at least some rudimentary information about their applicability conditions.

An alternative explanation for the results can be offered, however. In particular, the generation task might have facilitated performance by providing subjects with

several examples for each type of representation to use as the basis for representational transfer. In contrast, subjects in the specific example condition had only a single example for each type of representation. The present study was designed to test this alternative explanation. If the example-based reasoning explanation is correct, we would expect characteristics of the examples generated to predict performance on the selection task. All analyses conducted on the Experiment 2 data were also conducted on the data from the general category condition of Experiment 1. Without exception, the two data sets showed the same results.

For the three spatial diagram representations, there was no evidence that subjects in the general category condition relied on example-based reasoning to choose the most appropriate representation for the selection problems. In the global analyses, neither the number of appropriate examples nor the proportion of types of examples generated was reliably correlated with selection accuracy. In the fine-grained analyses, the number of appropriate examples generated of a particular type was not reliably correlated with accuracy on the selection problem(s) of that type. Note that we are not claiming that students do not have exemplar-based knowledge for the three spatial diagram representations or that they never use such knowledge. We believe that they do have such knowledge and that they do use it in certain circumstances (e.g., especially when the known example and the current problem share salient superficial features). We are arguing only that this was not the basis for the superior performance in the general category condition in Experiment 1.

We believe we are justified in drawing this conclusion from the null results for several reasons. First, the results were obtained in two independent replications. Second, we had adequate reliability for our coding of the examples subjects generated. Third, our data are powerful enough to detect a reliable relation between the variables under consideration when one exists: The one pair of variables (involving the part-whole representation) that was reliably correlated was reliably correlated in both

experiments. Thus these correlations cannot easily be dismissed as Type I errors.

It is less clear how to interpret the results for the part-whole representation. Students may have some abstract knowledge about this type of representation, but we did not see strong evidence of that in Experiment 1. We know from Novick and Hmelo's (1994) research that they can use example-based reasoning (although they may need some help to do so), and the one reliable correlation in Experiment 2 here supports this conclusion, but we did not find overwhelming evidence for example-based reasoning from the other correlational analyses in this study. It may be that students' knowledge about the part-whole representation is particularly fragile or fragmentary. Additional research is needed to address this issue, but it will not be our focus here.

EXPERIMENT 3

As described in the Discussion section of Experiment 1, there are two additional alternative explanations for the superior selection task performance in the general category condition of Experiment 1. The training explanation proposes that subjects in that condition formed generalizations from the examples they generated and therefore learned new abstract information as a result of completing our knowledge cuing task. According to this explanation, subjects relied on abstract knowledge to guide their performance on the selection task, as we claimed, but that knowledge was learned in the experiment rather than being preexisting. Another alternative explanation is that subjects in the general category condition performed better on the selection task due to transfer appropriate processing. In particular, the generation task gave them practice at linking specific examples to types of representations, which arguably is the main goal of the selection task.

Experiment 3 was designed to evaluate these two alternative explanations while replicating the general pattern of results found in Experiment 1. Subjects completed the selection task after having their knowledge cued in one of three ways. The specific example condition was essentially the same as that condition in Experiment 1. Subjects were given a single detailed example problem to solve for each representation. The general category (generate) condition was the same as the general category condition in Experiment 1 except for a large reduction in the amount of time allotted for generating example uses for each representation, from 5 to 1.5 min. If the training explanation is correct, performance in this condition should be much worse in this experiment than in Experiment 1, and likely worse than performance in the specific example condition as well, because the time available for learning has been drastically reduced. Finally, we added a new general category (familiarize) condition to address the transfer appropriate processing explanation. For this condition, we devised an orienting task that seemed

likely to provide a "purer" cue for subjects' general knowledge about the spatial diagram representations. Subjects were shown the labeled diagram category probes and were simply asked to familiarize themselves with each of the representations for 20 sec. If the transfer appropriate processing explanation is correct, then we would expect performance to be significantly worse in the general category (familiarize) condition than in the specific example condition, because the former subjects are highly unlikely to have had the benefit of having performed a task with similar processing demands prior to the selection task.

Because our main interest is in the nature of people's knowledge concerning the three spatial diagram representations, there is no compelling reason to include the part-whole representation in this experiment. Therefore, to reduce the length of the experimental session, we removed that representation from this study. That meant that subjects performed the knowledge cue task for only three representations, and the three part-whole problems were removed from the selection task.

Method

Subjects. Sixty-five Vanderbilt University undergraduates participated in partial fulfillment of course requirements. Approximately equal numbers of males and females were randomly assigned to each of three conditions.

Design. We varied across subjects the type of knowledge subjects were cued to use to help them determine the best type of representation for the problems on the selection task. Two conditions used orienting tasks that were intended to cue subjects' general (i.e., abstract) knowledge about the three spatial diagram representations: general category (generate) and general category (familiarize). The third condition gave subjects specific example problems to use to help them perform the selection task. Twenty-three subjects participated in the general category (generate) condition, and 21 subjects participated in each of the other two conditions.

The specific example condition was identical to the specific example condition in Experiment 1 except that subjects were given 6 min to complete the example problem for each type of representation instead of 5 min. Extra time was allowed for this task because a number of subjects were not able to complete the hierarchy example problem within 5 min. Also, with additional time to study the example problems, subjects may form richer representations of them that would improve their performance on the selection task. The general category (generate) condition was identical to the general category condition used in Experiments 1 and 2, except that subjects were given only 1.5 min to generate as many examples as they could for each type of representation.

Finally, the general category (familiarize) condition was devised to provide a way to cue subjects' general knowledge about the spatial representations without encouraging them to perform a specific task with the representations that could affect their performance on the selection task for reasons other than those hypothesized. Subjects were shown the labeled diagram category probes used in the general category (generate) condition. However, their instructions were to simply familiarize themselves with the given representation for 20 sec "so that you'll have clearly in mind what each one is like when you do the next task."

To determine how subjects interpreted these instructions, 9 additional subjects completed just the familiarize task and then wrote down what they had thought about for each representation. For all

three representations, a majority of the subjects thought about properties of the diagrams (67% for hierarchy, 56% for matrix, 78% for network). Sometimes these thoughts concerned structural properties of the diagrams (e.g., one-to-many breakdown for the hierarchy, directional links for the network), and sometimes they involved counting features of the diagrams (e.g., the number of rows and columns for the matrix, the number of nodes and branches for the hierarchy). In contrast, only a minority of subjects generated specific examples: 1 subject (11%) for the network (the path of a journey), 1 subject (11%) for the matrix (a multiplication table), and 3 subjects (33%) for the hierarchy (probability problems, bureaucracy, and family tree). If we include nonspecific example responses (e.g., "It reminds me of a biology class" for the network, "growth" for the hierarchy), the counts remain at 1 subject each for the network and matrix representations and increase by 1 to 4 for the hierarchy representation. No other types of responses occurred with notable frequency across representations.

Procedure and Materials. Subjects participated in small groups of 2–8 (except for 1 subject who participated individually) in a single session that lasted about 60 min. All subjects in a group participated in the same condition. Subjects first received the knowledge cue task appropriate to their assigned condition. As in Experiment 1, the cue task was introduced as a preliminary task to help subjects become familiar with the three types of representations we were interested in for the main task. The cues relating to each of the three spatial representations were presented in one of three orders, with each order being used approximately equally often in each condition. Thus each representation was encountered approximately equally often first, second, and third.

After the knowledge cue task, subjects completed the selection task, which was essentially the same as in Experiment 1 except for a few minor changes. For the general category (generate) and specific example conditions, the representation choices were printed at the top of the answer sheet, as described in Experiment 1. Representation choices for the general category (familiarize) condition were presented in the same manner as for the general category (generate) condition. As mentioned earlier, the part-whole problems were omitted. In addition, one of the network problems was replaced with a new problem because performance on that problem was very high in all conditions in Experiment 1 ($M = 80\%$ correct). It is more difficult to find differences across conditions when performance is very high or very low. Finally, one of the hierarchy problems was replaced with a new problem because the original problem had a cover story that was similar to the cover story for the corresponding example problem in the specific example condition (both problems were about games). It is well known that problem solvers easily notice and make use of superficial similarities between problems. In contrast, our interest was in the type of knowledge solvers rely on when there are no obvious superficial similarities available to guide performance. The problem booklet consisted of 12 story problems, 3 for each of the spatial representations and 3 for none of the above. Ten of the 12 problems were identical to problems used in Experiment 1.

The 12 problems were collated into booklets in one of four different orders. The orders were random with the constraint that no more than two problems with the same representation could appear in each half of the problem booklet.

Predictions

This experiment was designed to replicate Experiment 1 for the general category (generate) and specific example conditions and to show that the general category (generate) knowledge cuing task did nothing beyond cuing subjects' abstract knowledge for the three spatial representations (i.e., to discredit the training and

transfer appropriate processing explanations described earlier). Therefore, we predicted that subjects would perform better on the selection task in both the general category (generate) and the general category (familiarize) conditions than in the specific example condition. Moreover, we predicted that subjects in the two general category conditions would perform similarly.

Results and Discussion

Selection task performance as a function of condition and type of representation. As in Experiment 1, our analysis of selection task performance was based on discrimination scores. Again, we averaged the discrimination scores for the hierarchy, matrix, and network representations to get means by condition for the spatial representations. In this experiment, the nonspatial means came solely from the none-of-the-above problems. As shown in Figure 6, the results support our predictions, thereby replicating Experiment 1 while discrediting the training and transfer appropriate processing explanations of the earlier results. Thus the present results provide further support for the hypothesis that subjects have rudimentary schemas for the three spatial diagram representations.

To analyze the data, we conducted a contrast ANOVA that enabled us to directly test the predictions outlined earlier. In this analysis, the main effect of condition was divided into two orthogonal pieces—(1) a comparison of the two general category conditions, combined, with the specific example condition and (2) a comparison of the general category (generate) condition combined, with the general category (familiarize) condition. The condition \times type of representation (spatial vs. nonspatial) interaction was similarly divided into two orthogonal pieces—the interaction of representation type with each of the two condition contrasts.

There was no overall difference between the spatial and nonspatial representations [$F(1,62) < 1$, $MS_e = 0.0630$] or between the two general category conditions [$F(1,62) < 1$, $MS_e = 0.1240$]. Moreover, the type of general category condition did not interact with the type of representation [$F(1,62) < 1$]. Thus, performance was statistically equivalent in the two general category conditions for both representation types. There was a reliable overall difference between the general category condition and the specific example condition [$F(1,62) = 4.41$, $p < .04$], with better performance occurring in the two general category conditions. More importantly, as predicted, there was a reliable interaction between the general category versus specific example conditions and the type of representation [$F(1,62) = 6.37$, $p < .02$]. As shown in Figure 6, subjects in the two general category conditions performed much better than subjects in the specific example condition for the spatial representations but not for the nonspatial none-of-the-above category. As in Experiment 1, the large difference between the general category and specific example conditions was obtained for each of the spatial representations separately: means of .58 versus

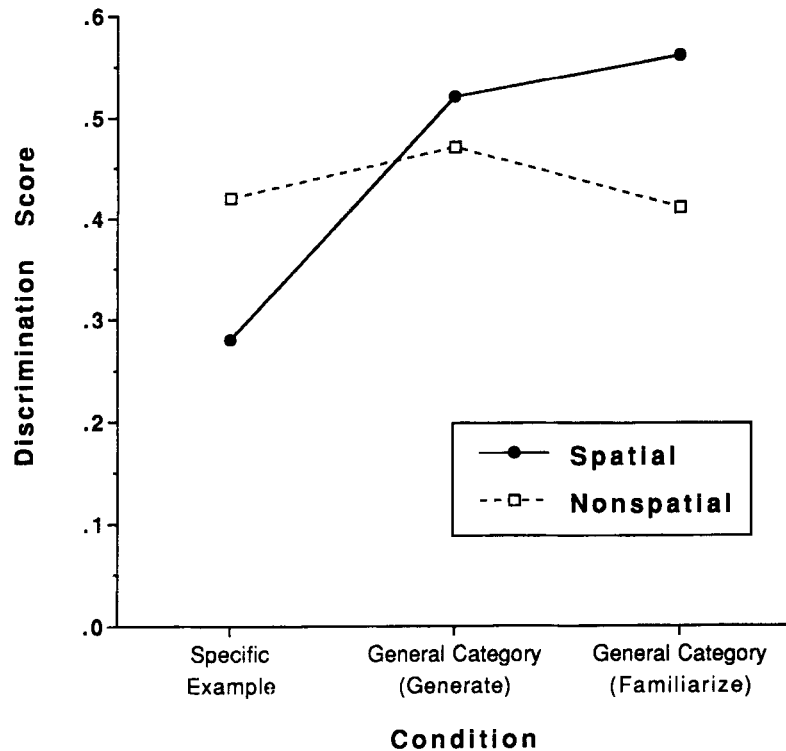


Figure 6. Mean discrimination scores for spatial and nonspatial representations as a function of experimental condition in Experiment 3.

.29, respectively, for hierarchy; means of .58 versus .32, respectively, for matrix; and means of .46 versus .24, respectively, for network.

Reconsidering the training and transfer appropriate processing explanations. The comparable performance in the two general category conditions supports our interpretation that both the example generation and “familiarize” tasks served as cues to subjects’ abstract knowledge in memory about the three spatial diagram representations. Moreover, the better performance in the general category (familiarize) condition than in the specific example condition discredits the transfer appropriate processing explanation of the Experiment 1 results, and the better performance in the general category (generate) condition than in the specific example condition discredits the training explanation. With respect to the generate examples knowledge cue task, we find it highly implausible that subjects would have had enough time (never mind the inclination) to abstract out any of the applicability conditions for the representations from the examples they generated during the 1.5 min allotted for example generation. Moreover, the spatial diagrams we are investigating are used frequently for representing information in newspapers, textbooks, and elsewhere. If our college-student subjects were able to form useful abstractions for these representations during 1.5 min of generating examples, surely they would already have done so on the basis of their prior experiences.³

Is the poor performance in the specific example condition due to using misleading example problems?

We have argued that relying on abstract schemas improves selection task performance compared to relying on specific examples. A possible alternative explanation for our results is that the example problems were misleading or difficult to understand and consequently impaired performance. The results of several experiments reported in Novick and Hmelo (1994), however, provide evidence that having students solve the types of example problems we used here does not interfere with performance, but in fact improves performance. Novick and Hmelo had subjects solve word problems for which a spatial diagram representation would be useful for organizing the information presented, after having solved a set of problems that either included or did not include a relevant example problem. Subjects had to figure out on their own which of the example problems, if any, was relevant to each test problem. For test problems requiring each of the three representations, subjects were more likely to use the appropriate spatial diagram representation after having solved the relevant example problem.

These results are directly applicable to the present study for several reasons. First, in both studies the example problems differed from the test problems in terms of the specific type of each representation used (according to the coding scheme discussed in Experiment 2). Second, Novick and Hmelo’s (1994) subjects were told that

the example problems might be helpful for solving the test problems, which is similar to the experimental situation in the present experiments in which subjects knew the possible representation choices. Thus, it is highly unlikely that the poorer performance in the specific example condition was due to the use of misleading examples.

Neither were our examples difficult to understand. As in Experiment 1, we coded whether subjects in the specific example condition (1) correctly used the representations given for the example problems and (2) correctly solved those problems. As shown in Table 2, performance on the example problems was similar to that found in Experiment 1. The slight variation in performance across the experiments (better for hierarchy in Experiment 1, better for matrix and network in Experiment 3) was most likely random. Averaged across the three problems, 89% of subjects used the representations correctly and 78% of subjects solved the problems correctly.

GENERAL DISCUSSION

Summary of Results and Methodological Considerations

The results of the three experiments reported here provide evidence that students are able to reason about the applicability of spatial diagram representations such as matrices, hierarchies, and networks on the basis of abstract schematic (i.e., general category) knowledge. In particular, when subjects were led to rely on such knowledge, they were more accurate at selecting the most appropriate type of representation for a problem than when they were led to rely on a single detailed example problem for each representation (Experiments 1 and 3). Moreover, adding the detailed examples to subjects' schematic knowledge produced no increment in selection performance (Experiment 1). The results of Experiment 3 indicated that subjects possessed abstract schematic knowledge about the three spatial diagram representations prior to the start of the experiment and did not simply learn that information during completion of the example generation knowledge cue task. Experiment 3 also called into question the alternative explanation that the good performance in the general category condition of Experiment 1 was due to transfer appropriate processing from the knowledge cue task to the criterion (selection) task. Finally, several analyses failed to provide any evidence for example-specific reasoning when subjects were cued to use their general category knowledge (Experiments 1 and 2).

In addition to providing greater insight into students' knowledge and use of spatial diagram representations in problem solving, which we will discuss in more detail shortly, the present results introduce two new methods for cuing general category knowledge. First, asking subjects to spend a few minutes (as few as 1.5 in Experiment 3) generating examples of situations for which a particular type of representation is appropriate enabled them to access their general knowledge about that representation. We are not aware of any other research that has

used this paradigm for cuing general category knowledge, so its range of application is unknown. Second, asking subjects to simply spend 20 sec familiarizing themselves with a representation was equally effective at cuing their general knowledge. The effectiveness of these simple tasks seems consonant with Smith et al.'s (1992) analysis of the types of results that are likely to be obtained when people possess abstract, rule-based knowledge in a domain.

On Schemas for Spatial Diagram Representations

We interpreted the present results as providing evidence for the existence of abstract schemas for the three spatial diagram representations that contain at least some rudimentary information about the applicability conditions for those representations. These representation schemas seem to be at an intermediate level of generality, much like the pragmatic reasoning schemas studied by Cheng and Holyoak (1985). They are not specific to a particular circumscribed domain, as are (for example) the permutations formula, Newton's Second Law, or a diagram of a pulley system; nor are they extremely general, as are rules of logic. Rather, each type of spatial diagram representation works best for a particular type of relational structure, regardless of the problem domain in which that structure is embedded (also see Novick & Hmelo, 1994). For example, a hierarchy is appropriate when the concepts can be organized into levels and there is only one route between any two concepts (Novick & Hurley, 1997). This type of structure is applicable to situations in a variety of content domains, but there are strong constraints on its use. For example, a hierarchy is not appropriate if there are both one-to-many and many-to-one connections between the concepts, and therefore multiple paths exist between at least some pairs of concepts. In such situations, one would likely use a network (or possibly a matrix).

Because hierarchy, matrix, and network representations have a consistent, canonical spatial embodiment, the problem structure best suited for each type of representation can be easily summarized by a "prototype" spatial diagram. Thus solvers could begin to construct rudimentary representation schemas by using the prototype diagrams as convenient "hooks" to which other relevant information could be attached. For the three representations of interest here, relatively simple parametric variations are sufficient to modify a representation prototype to capture the structure of essentially any problem for which that type of representation is appropriate. For the matrix representation, for example, solvers might have a prototype in which the numbers of rows and columns are represented by variables. Applying the prototype to a particular problem would then simply require instantiating those variables with appropriate values from the problem. Similarly, for the hierarchy representation, the number of levels and the number of branches at each level could be specified as variables. And, given a network with a certain number of nodes and links, applying that type of representation to a new situation simply involves

building the same kind of structure with different numbers of nodes and links and different connections between the nodes.

Nevertheless, there are clear limits to solvers' knowledge about the applicability conditions for the three spatial diagram representations, or to their ability to call upon that knowledge, because solvers often do not spontaneously use these representations when appropriate (e.g., Novick, 1990; Novick & Hmelo, 1994; S. H. Schwartz, 1971). In the present experiments, subjects in the general category conditions selected the appropriate spatial diagram representation roughly 61% of the time. Part of the difficulty solvers may have in determining what type of spatial diagram to use is that these representations are quite similar. A hierarchy is a special case of a network, and any information that is most efficiently represented using a hierarchy, matrix, or network can also be represented, although less efficiently (often much less efficiently), using either of the other two types of representations. VanderStoep and Seifert (1993/1994) have shown that greater similarity between principles makes it more difficult for students to learn their applicability conditions.

Reasoning About Conceptual Representations

The results for the conceptual part-whole representation provided an exception to the finding of strong effects of schema-based reasoning in the selection of an appropriate representation for a problem. In a similar vein, Novick and Hmelo (1994) found that transfer of a part-whole representation from an example problem for use in solving a test problem was more difficult than was transfer of any of the three spatial diagram representations. Taken together, these results may indicate that solvers' knowledge and use of conceptual representations in problem solving differ qualitatively from their knowledge and use of spatial representations. Part of the difficulty in constructing a good part-whole schema may be that there are so many different kinds of part-whole relations that have their own distinct structural features (see Winston et al., 1987). Additional research with other types of nonspatial representations will be required to better understand the differences between solvers' knowledge and use of spatial versus nonspatial representations in reasoning and problem solving, and why those differences occur. A parity representation (see Gick & McGarry, 1992; Kaplan & Simon, 1990) or compensation representation (see Wertheimer, 1959/1982) might be a possible candidate for future research to supplement the present findings using the part-whole representation.

Why are Students So Poor at Using Spatial Diagram Representations?

We mentioned earlier that there are clear limits to solvers' knowledge about the applicability conditions for the three spatial diagram representations, or to their ability to call upon that knowledge in problem-solving situations. It is important to consider why students have difficulty reasoning diagrammatically. One possibility is

that spatial diagram representations are too sophisticated a tool for students with the level of mathematical and logical training possessed by our subjects. We would not be surprised to find that third graders cannot use algebraic representations or that seventh graders do not understand integration. Perhaps the analogous situation exists with respect to the use of spatial diagram representations by college students. Alternatively, it may be that diagrammatic reasoning is a neglected topic in math and science (including social science) classes, so most college students have had only limited opportunity to learn how to use diagrammatic representations (e.g., Barwise & Etchemendy, 1991).

Several pieces of evidence support the conclusion that solvers' deficiencies in diagrammatic reasoning are likely to reflect their lack of systematic instruction in this domain. The results of the study reported in Novick and Hurley (1997) indicate that advanced computer science majors and advanced students double-majoring in mathematics and secondary education know considerably more about the applicability conditions for the three spatial diagram representations than do students with non-mathematics-related majors.

D. L. Schwartz (1993) found in one study that seventh graders, in contrast to ninth graders, very rarely used a network to represent many-to-one relations. Therefore, in a second study, he taught an experimental group of seventh graders three types of representations: matrices, networks, and permutation lists. A control group of students learned about cartesian graphs instead of networks. Several weeks later, during a weekly problem-solving day in their class, all students were given problems to solve involving the transmission of information. They were not told anything about the helpfulness of constructing visual representations for these problems. The effect of the earlier instruction was remarkable: 50% of the experimental subjects, but only 8% of the control subjects, constructed network representations for the test problems.

It is likely that children are able to begin learning about spatial diagram representations long before the age of 12. For example, the teacher for a preschool class of 4-year-olds helped her students construct a matrix representation to illustrate the ownership relations that existed between the students and shoes of different colors. The rows were labeled with the children's names, the columns were labeled with various shoe colors, and check-marks inside the cells denoted ownership. After they constructed this matrix, the teacher helped the children reason about the information contained in it. In particular, the children determined who had the most colorful shoes and which shoe color was most popular.⁴

Future Directions

The present experiments have demonstrated that (college-age) problem solvers have rudimentary schemas for the three spatial diagram representations studied here (hierarchy, matrix, and network), and that they use those schemas to help them determine the most appropriate

type of representation for a problem. That is, students have some information about the applicability conditions for these types of representations. On the basis of the present data, however, we cannot say anything very specific about what students know about each of the three spatial diagram representations. Further progress in understanding the processes of selecting, constructing, and using these representations in problem solving, as well as the interrelations among these processes, would seem to be predicated on the ability to more precisely assess solvers' knowledge.

Before that issue can be addressed, it will be necessary to have a theory that specifies the similarities and differences among the three spatial diagram representations. This theory must specify the types of problem structures for which each type of representation is best suited. That is, the theory must specify the applicability conditions for each of the representations. We are currently working on this topic (Novick & Hurley, 1997).

REFERENCES

- ANDERSON, J. R., FARRELL, R., & SAUERS, R. (1984). Learning to program in LISP. *Cognitive Science*, *8*, 87-129.
- BARTRAM, D. J. (1980). Comprehending spatial information: The relative efficiency of different methods of presenting information about bus routes. *Journal of Applied Psychology*, *65*, 103-110.
- BARWISE, J., & ETCHEMENDY, J. (1991). Visual information and valid reasoning. In W. Zimmermann & S. Cunningham (Eds.), *Visualization in teaching and learning mathematics* (pp. 9-24). Washington, DC: Mathematical Association of America.
- BRANSFORD, J., SHERWOOD, R., VYE, N., & RIESER, J. (1986). Teaching thinking and problem solving: Research foundations. *American Psychologist*, *41*, 1078-1089.
- BUTLER, D. L. (1993). Graphics in psychology: Pictures, data, and especially concepts. *Behavior Research Methods, Instruments, & Computers*, *25*, 81-92.
- CARROLL, J. M., THOMAS, J. C., & MALHOTRA, A. (1980). Presentation and representation in design problem-solving. *British Journal of Psychology*, *71*, 143-153.
- CHENG, P. W., & HOLYOAK, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391-416.
- 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*, *13*, 145-182.
- CHI, M. T. H., FELTOVICH, P. J., & GLASER, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.
- DAY, R. S. (1988). Alternative representations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 261-305). San Diego: Academic Press.
- FRANCIS, M. (1995). *Symbolic representations: Relations among four measures of knowledge*. Unpublished master's thesis, Vanderbilt University, Nashville.
- GICK, M. L., & MCGARRY, S. J. (1992). Learning from mistakes: Inducing analogous solution failures to a source problem produces later successes in analogical transfer. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 623-639.
- GOLDIN, G. A. (1985). Thinking scientifically and thinking mathematically: A discussion of the paper by Heller and Hungate. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives* (pp. 113-122). Hillsdale, NJ: Erlbaum.
- GURI-ROZENBLIT, S. (1988). The interrelations between diagrammatic representations and verbal explanations in learning from social science texts. *Instructional Science*, *17*, 219-234.
- HEGARTY, M., & JUST, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory & Language*, *32*, 717-742.
- HOLYOAK, K. J. (1985). The pragmatics of analogical transfer. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 19, pp. 59-87). New York: Academic Press.
- KAPLAN, C. A., & SIMON, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*, 374-419.
- KINDFIELD, A. C. H. (1993/1994). Biology diagrams: Tools to think with. *The Journal of the Learning Sciences*, *3*, 1-36.
- KOEDINGER, K. R., & ANDERSON, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, *14*, 511-550.
- LARKIN, J. H., & SIMON, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*, 65-99.
- LEVIN, J. R. (1989). A transfer-appropriate-processing perspective of pictures in prose. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 83-100). Amsterdam: Elsevier.
- LEWIS, A. B. (1989). Training students to represent arithmetic word problems. *Journal of Educational Psychology*, *81*, 521-531.
- LYNCH, M. (1990). The externalized retina: Selection and mathematization in the visual documentation of objects in the life sciences. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice* (pp. 153-186). Cambridge, MA: MIT Press.
- MALT, B. C. (1989). An on-line investigation of prototype and exemplar strategies in classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 539-555.
- MAYER, R. E., & GALLINI, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*, 715-726.
- MCGUINNESS, C. (1986). Problem representation: The effects of spatial arrays. *Memory & Cognition*, *14*, 270-280.
- MEDIN, D. L., & ROSS, B. H. (1989). The specific character of abstract thought: Categorization, problem solving, and induction. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 5, pp. 189-223). Hillsdale, NJ: Erlbaum.
- MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, *16*, 519-533.
- NOVICK, L. R. (1990). Representational transfer in problem solving. *Psychological Science*, *1*, 128-132.
- NOVICK, L. R., & HMELO, C. E. (1994). Transferring symbolic representations across non-isomorphic problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 1296-1321.
- NOVICK, L. R., & HOLYOAK, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 398-415.
- NOVICK, L. R., & HURLEY, S. M. (1997, November). *College students' knowledge about three spatial diagram representations*. Paper presented at the annual meeting of the Psychonomic Society, Philadelphia.
- POLYA, G. (1957). *How to solve it* (2nd ed.). Princeton, NJ: Princeton University Press.
- REED, S. K., & BOLSTAD, C. A. (1991). Use of examples and procedures in problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 753-766.
- ROSS, B. H., & KENNEDY, P. T. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 42-55.
- SCHOENFELD, A. H., & HERRMANN, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *8*, 484-494.
- SCHWARTZ, D. L. (1993). The construction and analogical transfer of symbolic visualizations. *Journal of Research in Science Teaching*, *30*, 1309-1325.
- SCHWARTZ, S. H. (1971). Modes of representation and problem solving: Well evolved is half solved. *Journal of Experimental Psychology*, *91*, 347-350.
- SCHWARTZ, S. H., & FATTALEH, D. L. (1972). Representation in deductive problem solving: The matrix. *Journal of Experimental Psychology*, *95*, 343-348.
- SMITH, E. E., LANGSTON, C., & NISBETT, R. E. (1992). The case for rules in reasoning. *Cognitive Science*, *16*, 1-40.
- SWELLER, J., CHANDLER, P., TIERNEY, P., & COOPER, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, *119*, 176-192.

- VANDERSTOEP, S. W., & SEIFERT, C. M. (1993/1994). Learning "how" versus learning "when": Improving transfer of problem-solving principles. *The Journal of the Learning Sciences*, 3, 93-111.
- WERTHEIMER, M. (1982). *Productive thinking* (Rev. ed.). Chicago: University of Chicago Press. (Original work published 1959)
- WINN, W. (1989). The design and use of instructional graphics. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 125-144). Amsterdam: Elsevier.
- WINSTON, M. E., CHAFFIN, R., & HERRMANN, D. (1987). A taxonomy of part-whole relations. *Cognitive Science*, 11, 417-444.
- ZIMMERMANN, W., & CUNNINGHAM, S. (1991). Editors' introduction: What is mathematical visualization? In W. Zimmermann & S. Cunningham (Eds.), *Visualization in teaching and learning mathematics* (pp. 1-8). Washington, DC: Mathematical Association of America.

NOTES

1. We initially constructed the orders to reflect the constraint that two problems with the same type of representation had to be separated by at least two other problems. This constraint was not completely realized, however, due to an error in numbering the problems.

2. For the hierarchy, matrix, and network representations, an "appropriate example" category was added after the authors coded the Experiment 2 data to help resolve disagreements that reflected alternative reasonable interpretations of a subject's response. (A "generic response" category for the part-whole representation, which serves the same purpose, already existed in the coding scheme.) The appropriate example category was not counted in the computation of the proportion of types for Experiment 2 because in most cases it was impossible to tell whether these ambiguous responses reflected a new type of example for the subject. Thus the number of possible types for each representation was the same in both experiments.

3. We would like to thank Tom Ward for suggesting this argument.

4. We would like to thank Mia Walker for providing this opportunity for us to examine representation use in young children.

APPENDIX A

Examples of the Selection Task Problems (One Problem per Representation)

Hierarchy representation. The following problem falls into the descent category.

Fifty years ago, Alex Johnson and Martha Jennings got married. Both of them had family heirlooms that were passed between the generations according to well-specified rules. At the time of the marriage, Alex Johnson owned a Stradivarius violin and a summer house at the beach.

- The Stradivarius violin goes to the oldest male child. If the current owner has no male children, then it goes to the oldest male child of the current owner's oldest brother who has a male child. If the current owner doesn't have any brothers or if the brothers don't have any male children, it goes to the oldest male child of the current owner's oldest sister who has a male child.
- The summer house at the beach always goes to the oldest child, but must alternate between male and female ownership. If the current owner doesn't have a child of the appropriate sex, then it goes to the current owner's oldest sister or brother whose oldest child is the appropriate sex. If none of the current owner's sisters or brothers has an oldest child of the appropriate sex, it goes to the current owner's oldest child regardless of sex.

At the time of the marriage, Martha Jennings Johnson owned a jeweled brooch.

- The jeweled brooch goes to the youngest female child. If the current owner has no female children, then it goes to the youngest female child of the current owner's youngest sister or brother who has a female child.

Alex and Martha have four children. The siblings, from oldest to youngest, are Ellen, Peter, Richard, and Sarah. Ellen and her husband have two children, Tony, the oldest, and Jessica, the youngest. Peter is a Catholic Priest and therefore doesn't have any children. Richard and his wife have two children: Susan and her younger brother Joshua. Sarah and her husband have three children: Alan, Becky, and Michael, in decreasing order of age. All of Alex and Martha's grandchildren are either married or of marriageable age, but none have any children yet. The family treasures change hands as soon as an eligible child is born. What was the order of passage of each of the three treasures? That is, in order, which people received the Stradivarius violin, the summer house at the beach, and the jeweled brooch?

Matrix representation. The following problem falls into the associated characteristic category.

Samantha is an 8-month-old infant who is a *very* fussy eater. She doesn't eat very much in one sitting, so her parents feed her small meals at different times of the day. It turns out that Samantha always eats on even-numbered hours (e.g., 8 a.m. or 2 p.m. but not 7 a.m. or 3 p.m.). She drinks 2 ounces of formula every 4 hours beginning at 6 a.m. She eats a third of a 6-ounce jar of fruit every 6 hours beginning at 6 a.m. She is fed a fourth of a cup of cereal every 6 hours beginning at 8 a.m. Finally, she eats half of a 6-ounce jar of baby vegetables at noon and then again at 6 p.m. The family is going on an outing to the state park. They are planning to leave the house at 7 a.m. and return at 7 p.m. How many ounces of food and drink do Samantha's parents need to bring for her?

Network representation. The following problem falls into the associative relationship category.

Four sets of brother/sister (i.e., fraternal) twins attending a baseball card show got together to trade cards among themselves. Each trade consisted of two children exchanging a single card. In other words, no child ever gave up or received more than one card in a trade. After all of the trading was completed, you asked everyone in the group *except* Susan Post how many cards he or she had traded. Each of the remaining seven children gave a *different* answer to your question. That is, one of the children you questioned traded 6 baseball cards (this child was Andrew Corley), one traded 5 cards, one traded 4 cards, one traded 3 cards, one traded 2 cards, one traded 1 card, and one traded 0 cards. You also discovered that no child traded cards with his/her twin, and no child traded cards with the same person more than once. Given this information, how many cards did Susan Post trade?

Part-whole representation. The following problem falls into the member/collection category.

A famous story-teller, recently returned from the land of Paraval, is recounting her travels there. The inhabitants of Paraval are most interesting and unusual. According to the story-teller:

All Bandersnatches are Boojums.

All Boojums are Snarks.

Some Snarks are frumious animals.

No Snarks breakfast at five o'clock tea.

A young boy and his friends are fascinated by the tale. Afterwards, they debate among themselves whether it is possible to conclude that:

(a) No Boojums breakfast at five o'clock tea.

(b) No frumious animals breakfast at five o'clock tea.

(c) Some Snarks are Bandersnatches.

- (d) Some frumious animals breakfast at five o'clock tea.
- (e) Some Boojums are frumious animals.

Can you help the children answer these questions?

None of the above. Far away, across many seas, is the island of Liars and Truars. Members of the Liars tribe always lie. The Truars always tell the truth. On the island of Liars and Truars, you meet three people named Alice, Ben, and Charlie. You ask Alice whether she is a Liar or a Truar. She answers in the local dialect, which you do not understand. Then you ask Ben what Alice said. Ben, who speaks English, says "Alice said she's a Liar." You then ask Ben about Charlie. "Charlie is a Liar too," Ben insists. Finally, Charlie adds, "Alice is a Truar." Which tribe does each person belong to?

APPENDIX B
Hierarchy Example Problem

In a certain fantasy game, each player's starting level is determined by the following sequence of actions. First, the player spins a spinner that is divided into two sections. The two sections have the following words on them, which are worth the indicated numbers of points:

Aardvark	+4
Butterfly	+7

Next, the player rolls a pyramidal (four-sided) die. The four sides are marked with the following symbols, which specify the indicated actions to be taken:

\$	double existing points
*	lose all points
?	stay at same number of points
#	+5 points

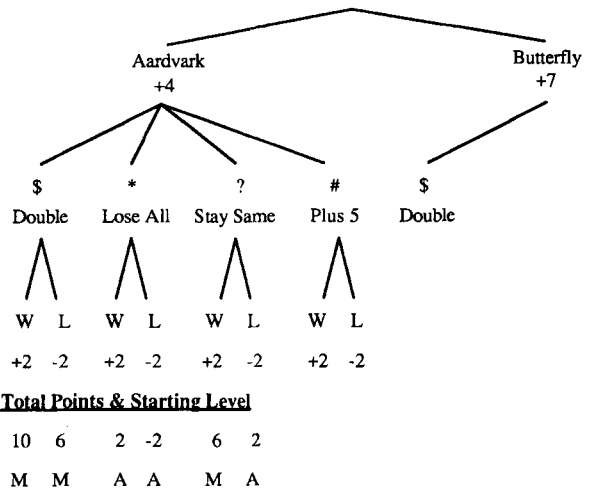
Finally, the player flips a coin marked *win* (+2 points) on one side and *lose* (-2 points) on the other side. The total number of

points a player has after taking these three actions determines the player's starting level in the game, as follows:

Starting Level	# Points
alchemist (A)	less than 4
magician (M)	4-10
juggler (J)	11-16

How many different combinations of actions enable the player to start the game as an alchemist, as a magician, and as a juggler?

Complete the following representation to solve the problem:



The number of different combinations of actions that enable the player to start the game at each level are: Alchemist (A) _____, Magician (M) _____, Juggler (J) _____.

(Manuscript received August 28, 1995;
revision accepted for publication March 13, 1998.)