SRRI #195

An Analogy-Based Computer Tutor for Remediating Physics Misconceptions

Tom Murray
Klaus Schultz
David Brown
John Clement
University of Massachusetts

Abstract

We describe a computer tutor designed to help students understand physics concepts. The tutor uses a teaching strategy called "bridging analogies" that previous research has demonstrated to be successful in one-on-one tutoring. The strategy is designed to remedy misconceptions by appealing to existing correct intuitions, and extending these intuitions by encouraging analogical thinking. Students were videotaped while using the program and were encouraged to think aloud. The strategy was successful in changing beliefs for some students. We outline suggestions for improving the tutor using artificial intelligence technology.

1 CONTEXT SPECIFIC TUTORING STRATEGIES

We report on a computer simulation of a human tutoring strategy that previous studies have shown to be effective in one-on-one human tutoring. The purpose of the research has been to evaluate the strengths and limitations of an automated version of the strategy, which we envision as ultimately being one of several strategies available to an intelligent computer tutor.

To date, most Intelligent Tutoring Systems (ITS) research has focused on the design of semantically rich learning environments, or on the conceptual analysis of topic areas and tasks. Though these efforts are important, this article addresses the paucity of research dealing with the art of tutoring itself. Though many ITS incorporate expertise in teaching, diagnosis, and communication, few reports of such studies are *explicit* about the following: 1. the teaching strategies used; 2. the assumptions about learning and teaching that underpin these strategies; and 3. the precise rules or algorithms that instantiate these strategies. Sharing this information is conducive to the comparative analysis of tutoring system descriptions. It also encourages the proliferation of research that confirms or extends the findings of studies designed to test or evaluate tutoring systems.

Many existing ITS systems teach skills or procedural knowledge. Examples are: Shute and Bonar (1986) and Streibel, Koedinger, Collins, and Jungck (1986) (scientific inquiry skills); Brown, Burton, and deKleer, SOPHIE (1982) and Clancey's GUIDON (1982) (diagnostic skills); and Anderson, Farrel, and Sauer's LISP tutor (1985) (programming skills). Our research is aimed at remediating *conceptual* difficulties, in contrast to remediating bugs in factual or procedural knowledge. Other ITS efforts are aimed at teaching conceptual knowledge, but they do so by making use of the computer's unique ability to simulate physical systems. In the present work, we use a discourse strategy to appeal to the logical and/or intuitive sensibilities of the learner. Incorporating semantically rich learning environments in our system at this point would cloud the research questions, although maximum learning leverage would of course be obtained in a system using both a rich environment and intelligent discourse.

The computer tutor described here has been designed to simulate a human tutoring strategy (as discussed in Clement 1987). The strategy re-

Research reported in this paper was supported in part by the National Science Foundation Grant No. MDR-8470579

Correspondence and requests for reprints should be sent to Tom Murray, Department of Computer and Information Science, University of Massachusetts, Amherst, MA 01003

medies misconceptions by appealing to existing correct intuitions, and extends these intuitions to a broader context by encouraging analogical thinking. The tutor presents the student with situations that bridge the conceptual gap between simple situations (those for which most students have correct intuitions) and difficult ones (those for which most students exhibit misconceptions). The strategy is invoked when the student is diagnosed to be harboring a misconception, as evidenced by an incorrect answer to a target problem. The tutor then attempts to find an "anchor" situation, that is, an analogous situation for which the student has a correct intuition. For the domain we are investigating, an example target problem is the question of whether a table exerts an upward force on a book that rests on it. Many students exhibit the misconception that rigid objects cannot exert forces. A typical anchoring situation is a book being held in a person's hand. Most students will agree that the hand exerts an upward force on the book. Typically, however, the student does not immediately believe that the two situations above are truly analogous with respect to the questions being asked. Intermediate analogous situations, called "bridging analogies," are used to bring the correct intuitions to bear on the target problem. For example, a bridge between the book on the table situation and the book on the hand situation is a book on a spring. Bridging situations are presented one at a time to the student, chosen according to the student's previous responses, as described later. The student is asked to compare (in his head or out loud) his or her response to the bridging situation with his or her response to the target problem. After the comparison, the student is allowed to change his or her answer to either situation. If his or her answers to the two situations differ, the comparison encourages the cognitive conflict that can motivate him or her to think more carefully or differently about both situations. Regardless of whether his or her answers to the two situations agree or disagree, the comparison encourages the student to consider the (potential) analogy between the two example situations. If the comparison is pedagogically successful, the student will see the target problem in a new light and change his or her response for it. This strategy has proven successful in remediating physics misconceptions in several studies involving human tutoring (Clement & Brown, 1984).

An important characteristic of the bridging analogies strategy is that it withholds giving the student substantive feedback except for situations where there is no alternative. This strategy, which is in concordance with a constructivist teaching philosophy, has been used in classroom teaching and one-on-one tutoring (Clement & Brown, 1984) to encourage active learning and to discourage passive acceptance of information and answer-oriented student behavior, which can lead to shallow understanding in conceptually difficult domains.

Human tutors operate with a battery of tutoring strategies, many of which are relevant only to specific contexts. We envision the bridging analogies strategy as one of many tutoring strategies that an intelligent computer tutor would have at its disposal, selecting each (or combinations of them) according to context applicability rules. We see these strategies as "strong method" solutions to the computer tutoring problem because their appropriateness is context dependent, relying on information about the student, the discourse state, and the topic characteristics. A strong method strategy is a specific algorithm or set of rules. Tutoring guidelines or strategies mentioned in previous research are usually "weak methods." They are more generally applicable, but not as specific as context specific strategies. Context specific strategies should be more potent when used in their intended instructional context. Examples of weak method strategies are: "provide concrete examples of, as well as descriptions of, concepts" (Burton & Brown, 1982); "provide immediate feedback for errors" (Anderson, Boyle, Farrell, & Reiser, 1984); and "let the student learn from mistakes" (Brown, Burton, & deKleer, 1982). Our strategy is also a "strong method" in the sense that it incorporates a substantial amount of domain-specific information, in the form of the example network.

The bridging analogies strategy is particularly applicable to the following context: topic areas with a qualitative, conceptual focus, where deep seated misconceptions exist and where correct intuitions (anchors) also exist. Basic Newtonian mechanics is a domain which contains many such topic areas, one of which provides the focus of our study. Real-world experiences result in a wealth of intuitions about how the physical world works. Some of these are in accordance with scientific laws, and others are in conflict with them.

In its current form, the bridging analogies tutor is not intended to be used in a stand-alone fashion to teach. It is an encoding of a specific teaching strategy which research has shown to be effective in certain contexts. The purpose of the research is to evaluate the strengths and limitations of a computer version of the strategy. Ideally, the effectiveness of each of the strategies employable by a computer tutor should be researched. Our view is that it is desirable to test such strategies independently, first off-line, and then in a prototype program. In this way, a great deal is known about how students react to these strategies before a large effort is expended in building an intelligent tutoring system based upon them. Of course, further research is then needed to determine how to integrate several tutoring strategies into a single system.

Research Questions

At the outset of the research there were several key questions:

- What aspects of the bridging analogies strategy are most and least effective for remediating misconceptions?
- Can a computer tutor successfully use a low-feedback response strategy?

Longer range questions that we are addressing are:

- Is it possible to determine dynamically during a computer tutoring session when to use the bridging analogies strategy and when to switch to another strategy?
- What other tutoring strategies might be effective for teaching science concepts where bridging analogies fail?

2 MISCONCEPTIONS IN SCIENCE EDUCATION

Research has investigated students' conceptions about important mathematics and science concepts as they exist prior to formal instruction (sometimes called preconceptions). A surprising number and variety of misconceptions (incorrect preconceptions) have been uncovered in domains such as electricity (Steinberg, 1983; Fredette & Lochhead, 1980), biology (Wandersee, 1983), statistics (Tversky & Kahneman 1974), and classical mechanics (Halloun & Hestenes, 1985). Often these misconceptions are deep-seeded and interfere with problem solving in the domain. Such incorrect beliefs are not limited to certain age groups or academic areas. Serious misconceptions are common even among students taking calculus-based physics courses and students who are proficient at quantitative problem solving using formulas. These debilitating beliefs must be dealt with explicitly in instruction. They are not simply "erased" when conflicting and correct knowledge is conveyed to the student. Many of these misconceptions are commonly held, so that attempts to discover and remedy the most common of them will benefit a large percentage of students. For instance, in a survey of 112 high school chemistry and biology students who had not taken physics, Clement (1987) found that 76% of the students did not believe that a table exerts an upward force on (or pushes up on) a book resting on it (83% of these students indicated a high confidence in their incorrect answer), a belief that is contrary to classical mechanics.

Several studies have shown that misconceptions such as the ones cited are quite resistant to remediation (i.e., they are "deep misconceptions") at least via traditional instructional methods. Halloun and Hestenes (1985) administered tests of basic concepts in Newtonian mechanics (motion and its causes) to over 1000 college level introductory physics students over a three-year period. The primary diagnostic instrument consisted of multiple choice questions about very basic concepts, such as: "if a ball is shot out of a semi-circular tube, what is the shape of its path after leaving the tube?" (The student was to choose between pictures of five path shapes.) Average scores before instruction were "very low," (about 45%) even for university physics courses where 80% of the students are engineering majors, and 80% have taken calculus. Most importantly, the *post-instruction* scores showed little improvement (11 to 15%). This scant gain was remarkably independent of such variables as instructor, mathematics ability, and final course grade.

Previous studies suggest several possible sources for students' misconceptions. White & Frederiksen (1986) note that misconceptions can result when the cognitive jumps required of a student while learning a new domain are too large or too under-specified. Learners can then fill in the gaps in unpredictable and uncontrollable ways. Similarly, diSessa (1983) discusses how misconceptions can result when learners try to accommodate to new information by incorrectly synthesizing existing fragmented and inconsistent pieces of knowledge. VanLehn (1983) has constructed a detailed theory of how procedural knowledge is modified ("repaired") in an ad-hoc way when an impasse is encountered during problem solving. A similar mechanism may account for some (nonprocedural) misconceptions. Claxton (1985) points out that preconceptions can be founded on physical experiences ("gut science") or on social experiences ("lay science"), and that these concepts can conflict with the formal and abstract concepts of "school science." Students are often perfectly content with or unaware of their conflicting beliefs if the contexts in which these beliefs are used do not overlap, such as in everyday life versus taking exams.

Many misconceptions result from repeated attempts at comprehending real world phenomena. Some of these beliefs have been used again and again to successfully cope with or explain real world events. Therefore, it is not surprising that some misconceptions are deep seated, and that some are quite common. Some domains, such as physics (McCloskey, 1983; McDermott, 1984) and statistics (Tversky & Kahneman, 1974), are particularly susceptible to the existence of misconceptions.[1] In such domains careful sequencing and explanation of new information, without using knowledge of the student's prior beliefs, is probably not sufficient to lead to understanding. Existing misconceptions must be directly addressed with innovative instructional approaches.

3 PEDAGOGICAL APPROACH

Here we describe the tenets and assumptions which underpin our approach. We have already mentioned several results of previous research: that debilitating misconceptions are widespread and resistant to change; that intuitive anchors exist; and that presenting students with bridging analogies and encouraging comparisons between analogous situations can overcome misconceptions.

[1] Clement (1982) has documented remarkable similarities between students' misconceptions and pre-Newtonian theories such as those of Galileo.

Our approach is also founded on the constructivist paradigm for learning (Piaget, 1972; Lawson & Wollman, 1975; von Glasersfeld, 1988). This paradigm holds that learning is not a passive process of allowing incoming information to be imprinted in the brain, but an active process of trying to maintain equilibrium in a changing environment.

Overcoming deep seated misconceptions involves the modification of existing schemata. For such conceptual change to take place the student must first experience some cognitive dissonance around the misconception. We encourage this dissonance by juxtaposing example situations (or thought experiments) for which the student has given contradictory (according to physics) predictions, and then asking the student to compare them. Since the goal of the tutor is for the student to interpret familiar situations in a new way, rather than simply to convey new information, it is important that the student is actively engaged in applying and critiquing his or her existing beliefs. The teacher cannot change a student's beliefs, the student must reconstruct his or her own beliefs. For this reason we are experimenting with the approach of withholding feedback considerably longer than is the practice in most other pedagogical styles. We not only withhold revealing the correct answers, but we also withhold feedback on whether the student's answers are correct or not. This shifts the emphasis from getting correct answers to thinking about the questions. It allows greater focus on: whether answers *make intuitive sense*, the student's *reasons* for answers, and the *interconnectedness* of knowledge. At the Scientific Reasoning Research Institute (SRRI) the instructional methodology of withholding substantive feedback from students has been used successfully to teach mathematics and science in many tutoring situations (Lochhead, 1983).

The bridging analogies tutor rarely tells the student if the answer is in accordance with the physicist's view. (It *does* reveal the correct analysis of example situations when no bridge is available.) The end effect is that the program generates discourse that has a flavor similar to: "What do you think of [a situation]? OK, now what do you think of [another situation]? Please compare the situations and comment on similarities and differences. Now, what do you think of [a third situation]? How does thinking about it affect your beliefs about previous situations?" Here there is essentially no feedback. The instructional leverage comes from careful sequencing of the example situations.

In teaching other topics, or in teaching other types of knowledge (e.g., skills or factual knowledge) immediate feedback may be necessary. Even in domains such as the one we are dealing with, withholding feedback, if taken to an extreme, of course, has its limitations. The discourse can stray so far from the goal that consolidation becomes difficult. Also, students have varying degrees of tolerance to engaging in instructional dialogue without feedback. Human tutors can sense when the discussion has strayed too far, or when the student's frustration is debilitating. In this study we wish to test the benefits and limitations of using a low-feedback response strategy on a computer tutor.

The bridging analogies strategy is a Socratic teaching method (Collins, 1977) in that it involves reasoning from specific cases, and the focus of the instruction is on strategically posing questions (as opposed to say, conveying information or demonstrating principles). The instructional goals of the bridging analogies tutor are a subset of those addressed by Collins & Stevens' WHY system. Our system aims at remediating deep misconceptions in domains where conceptual anchors exist. The WHY system was designed to teach about causal dependencies in a complex physical system, and also at teaching inquiry reasoning skills. As a subgoal of its major goals, the WHY system addressed the remediation of misconceptions. (Apparently, no data is available on the WHY system's tutoring effectiveness.) It based its remediation on a formal analysis of necessary and sufficient conditions, appealing to the student's logical thinking abilities. Our tutor uses an empirically derived network of example situations, and tries to appeal to the student's intuitions and existing correct (anchoring) concepts. Also our feedback strategy differs considerably from that used by Collins & Stevens.

4 THE BRIDGING ANALOGIES TUTOR

Clement & Brown (1984) have developed a teaching strategy, called "bridging analogies," which utilizes correct intuitions (which they call conceptual anchors) and appeals to students' analogical reasoning to help them extend their correct intuitions to target situations for which they have misconceptions. After the student gives an incorrect analysis of a situation (the original target) that indicates the existence of a misconception (indicating in this case that rigid objects cannot exert forces), he or she is presented with an analogous situation (such as a book in a hand) which is intended to serve as an anchor. In the classroom, situations must be chosen which are expected to be anchors to the majority of the class. In one-on-one tutoring, an anchor can be found with a series of questions. In either case, previous research must establish which situations are useful anchors for a given domain.

If the student answers the anchor correctly he or she has given contradictory answers for the target and anchor situations. In effect, he or she does not see them as analogous. The bridging analogies strategy attempts to bring the student to an understanding of this analogical relationship by presenting a sequence of intermediate analogies (called "bridging analogies"). At some point (or points) the student should be faced with considering two situations for which he or she has given contradictory answers yet which he or she realizes are analogous. The cognitive conflict which results should motivate the student to change his or her mind about the misconceived situation.

For computer implementation, the strategy was formulated as a procedure that traverses a predefined network of example situations according to student responses to questions.[2] The computer version of the strategy cannot respond to creative student insights, but it can tailor its presentation of examples personally for each student. We have used the program for different domains, but below we describe its implementation in teaching about the existence of opposing contact forces in static situations. This is the domain used for the majority of our research.

The Example Situation Network

Available for every example situation are:

- a description of the situation
- a more detailed description

[2] In this section we give an operational definition of the procedure. See Appendix A for a more precise description of the code implementing the strategy.

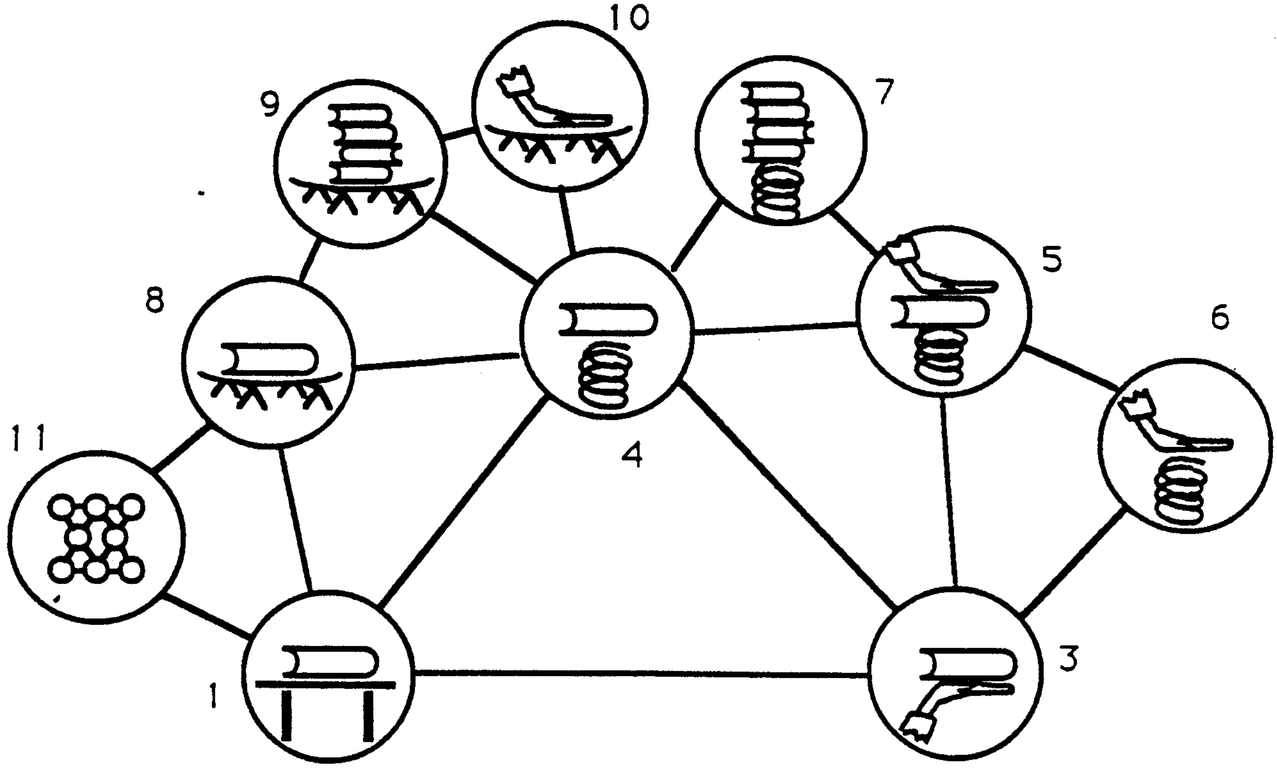


Figure 1. Simplified Network of Example Situations for Existence of Normal Forces

- a question with multiple choice answers
- an explanation of the correct answer
- a picture.

The bridging analogies tutor makes use of a network of example situations (see Figure 1). For each pair of linked situations, such as the book on the table (node 1) and the book on the spring (node 4), a bridging situation is defined—a book on a flexible board in this case (node 8). This network is an abbreviated version which we will use to simplify the explanation of the tutor. The network actually used by the program is shown in Figure 2.[3] It starts with the more difficult "fly on the road" situation rather than the "book on the table."

The network is structured so that a bridging situation shares key features with its corresponding anchor and target situations. For example: A table is rigid and has no volition (i.e., can't decide to push). The hand has both mobility and volition. The spring has mobility (flexibility) and no volition. The bridging situations were originally conceived by considering these kinds of feature similarities. But more importantly, the bridges and anchors in the network are based on previous experiments designed to test effective bridges and anchors. The network used in this study was thus in part empirically derived, and not constructed explicitly according to feature-based rules. It was, for the most part, designed and tested before this computer tutoring study was conceived. In this study we have used the existing network with few changes. Our (post-hoc) feature-based analysis of the network appears in the discussion section of this paper.

[3] The initial anchor for the network used in the study, node 0, is a "null node." Null nodes are ignored by the program. In this case 0 was included to force the program to give the book on the table situation directly after the fly on the road situation. The inclusion of this node was, in hindsight, perhaps unnecessary, and its existence in the network complicates an explanation of the network.

Finding the Anchor

When the student's answer to a key target situation indicates a misconception, the bridging analogies strategy is invoked. The goal of the strategy is to bring the student to an intuitive understanding of the target situation, as indi-

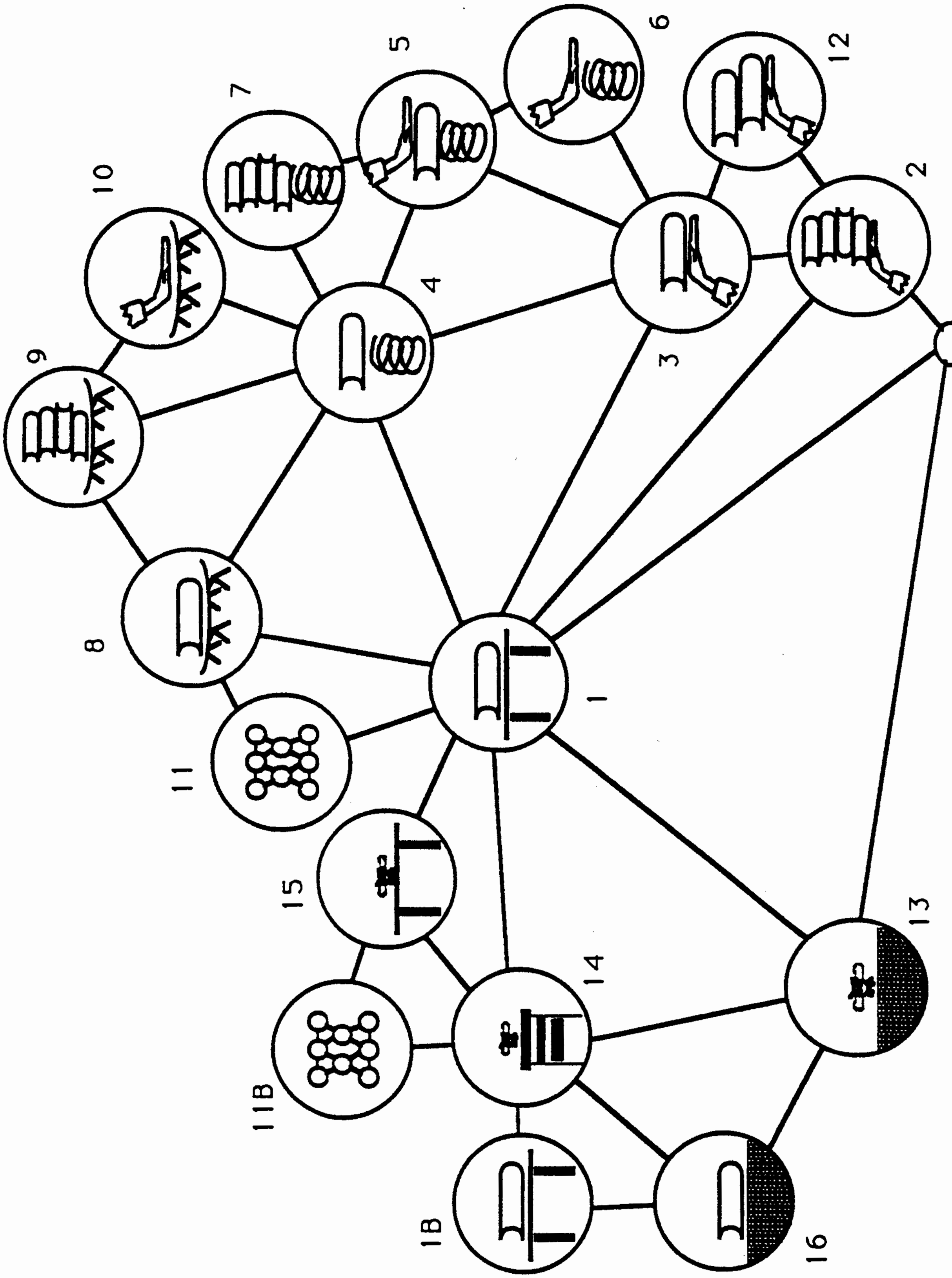
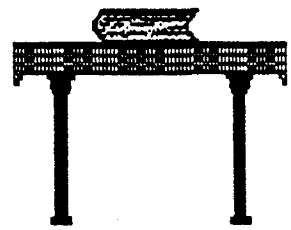


Figure 2. Complete Network of Example Situations

cated by the student changing his or her mind about it (and indicating a high confidence level, as described later). Figure 3 shows a sample dialogue including the book on the table and the book on the hand situations, accompanied by the diagrams shown to the student.

Comparison Questions

At certain points during a tutoring session the computer encourages a comparison of two previously seen example situations. The comparison reminds the student of his or her answer to



A book resting on a table

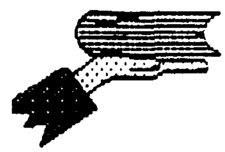
Imagine a medium size textbook resting on a dining room table.[1]

While the book is resting there, the table:
- IS exerting a force up on the book
- IS NOT exerting a force up on the book

The student chooses "IS NOT . . .".

Please rate your confidence in this answer:
blind guess *not very confident* *somewhat confident* *fairly confident* *I'm sure I'm right*

The student chooses "fairly confident".



Many books held in your hand

Imagine that you are holding a textbook in your hand.

While the book is resting there, your hand:
- IS exerting a force up on the book
- IS NOT exerting a force up on the book

The student chooses "IS . . . ".

The confidence question is given again and the student chooses "I'm sure I'm right."

For "a book in your hand" you said:
your hand IS exerting a force up on the book (with high confidence).
But for "the book on the table" you said:
the table IS NOT exerting a force up on the book (with very high confidence).

Explain[2] why these answers are different.

The student then explains her reasoning to the interviewer.

Do you want to change your mind about:
- a book in your hand
- a book on the table
- both
- neither

The student opts to change "neither" answer and the tutor moves on the present the book on the spring situation.

[1]After the description of an example situation the student is asked: "Do you want a more detailed description?" in case there is any confusion concerning the brief description given first. We will not show the detailed descriptions here.

[2]If the program were used without an interviewer, the prompt would be changed to: "Think for a moment about why . . ."

Figure 3. A Sample Dialogue Containing the Original Target and Anchor.

two previous situations, highlighting whether they agree or not. The student is then asked to explain[4] why these answers are the same or different, and is given the option to change his or her mind about either or both answers. See Figure 3 for an example of the comparison question.

The comparison has several purposes. First, it highlights potential contradictions in answers which, if the student sees the situations as truly analogous, fosters the cognitive dissonance (or "disequilibrium") which motivates changes in belief about the misconceived situation, and ultimately, the construction of new concepts. Second, it tests the stability of answers, giving the student a chance to re-think answers. Lastly, it encourages analogical thinking by asking the student to continuously juxtapose different example situations in a given domain. Note that the student is not told that the situations should be analogous. No direct feedback is given after the student answers a question.

After the anchor has been selected, the student is asked to compare the target situation with the anchor situation. The network is designed so that the target problem invokes incorrect intuitions and the anchor invokes correct ones for most students. Even though they are analogous to a physicist, most students are content with their conflicting answers, and will choose not to change either of them.

Bridging Between the Target and the Anchor

Given a target situation (for which a misconception exists) and an analogous anchor situation (for which a correct concept was demonstrated) the tutor next (see Figure 4):

1. Looks up which example situation splits in half the conceptual distance between the two, that is, it chooses a bridge, and asks a question about it,
2. Asks a comparison question to help the student establish the analogical link between the anchor and the bridge, and
3. Asks a comparison question to help the student establish the analogical relationship between the bridge and the target, with the hope that the student will change her mind about the target situation.

[4] The explanation is to his or herself or out loud to an interviewer. No text is typed in at this point. At our current phase of the research, we feel that such information is too difficult to analyze automatically, and the effort of typing in such an explanation may hinder the students' motivation and limit their responses to very short ones.

Since the algorithm is recursive, the status of a given example situation, whether it is a target, anchor, or bridge, can change as the tutoring session progresses. In order to bridge between 3 an 1 (in Figure 1) it may be necessary to bridge between 3 and 4, then between 4 and 1. If this happens, situation 4 is first a bridge (as it is initially asked), then a target (after it is answered incorrectly) and then an anchor (after the student changes her answer to a correct one). Thus, when we refer to a target, anchor, or bridge, we refer to the current status of that example situation.

In Figure 1 we see that the first bridge is the book on the spring (node 4). If the student answers incorrectly about the bridge situation (or changes her answer to incorrect as a result of the first comparison), then the tutor sets up a subgoal to tutor for understanding of the bridge situation. In that case we are back in our original condition, having a misconceived situation (the book on the spring), and an anchor (the book on the hand). The entire strategy is invoked recursively, first identifying a new bridge between the two (a hand on a book on a spring in this case). Levels of subproblems and bridges are presented until the student's misconception about the original bridge (the book on the spring) is (apparently) remedied (see Figure 4), and the anchor-to-bridge analogy is established.

Once the bridge situation is understood and the anchor-to-bridge analogy established, we wish to establish the bridge-to-target analogy, thus constructing a path of analogies or conceptual links connecting the anchor with the target. The strategy compares the bridge (the book on the spring in this case) with the target problem (the book on the table) with the hope that the student will change his or her mind about the target. If the student does so we are done. If the student does not opt to change his or her mind about the target, we are back in our original condition, having a misconceived situation (the book on the table) and a correctly conceived one (the book on the spring). This algorithm is invoked again to establish the bridge-to-target anal-

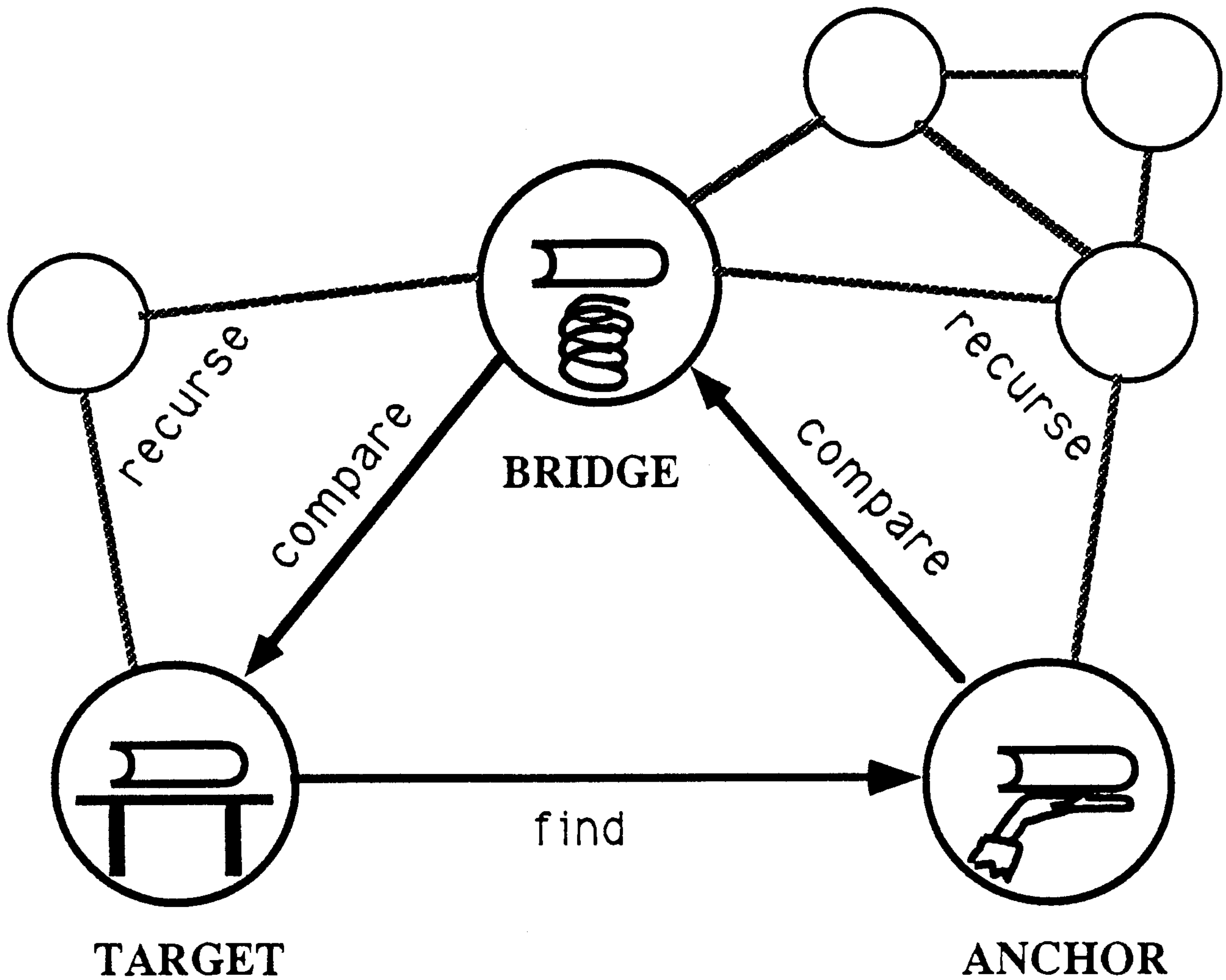


Figure 4. Diagram of the Bridging Analogies Strategy

ogy, beginning by finding a new bridge (in this case the book on a flexible board) and so on.

One perspective on the strategy is that it keeps splitting the conceptual difference between the simplest misunderstood situation and the most difficult understood situation. Eventually, it provides the student with an analogical path from the original anchor to the original target, such that the gaps between neighboring situations are close enough for the student to perceive the analogy. Splitting the difference at each step is more efficient than walking the student through every smallest step of the entire set of example situations linking the target with the anchor. The dark line in Figure 1 shows one student's journey through the network. The path shows the order that situations were understood, rather than the order in which they were introduced. For example, situation 4 was given and answered incorrectly, then situation 5 was given and answered correctly, and then the student opted to re-visit and change his or her mind to a correct answer for situation 4.

Explanations for Example Situations

There are two circumstances where a correct explanation of an example situation is given to the student. One is when the student, after a comparison, opts to change his or her mind about an anchor and changes it from right to wrong. At this point we have no more graceful method for dealing with these "regressions." They occur very rarely in our tests so far.

The second condition warranting an explanation (i.e., the student gets immediate feedback after a wrong answer) is when there is no bridge in the network between the current target and anchor, between nodes 5 and 7, for example. The explanation is given, often suggesting that the two situations are analogous, and the student is asked the problem question again.

A typical example of an explanation is (for node 4): "The spring IS pushing up on the book. When you push down on the spring, the spring compresses some and pushes back."

Confidence Checking

A severe drawback to computer tutoring, as compared with human tutoring, is the computer's inability to ascertain the types of affective information about the student that are evident to human tutors from facial expressions, vocal infections, and so on. Information about students' certainty level, confusion, boredom, excitement, and so on, are crucial to effective human tutors.

Human tutoring studies at the University of Massachusetts have used "how sure are you" or, alternatively, "how much sense does the answer make" scales to gather additional information about the strengths of anchors and the effectiveness of the bridging strategy in changing students' beliefs. Similarly, the computer tutor asks the student to rate his or her confidence in the answer on a five point scale after each question (as shown in Figure 3). The program combines this information with information about the correctness of the answers in deciding whether to ask for comparisons, and whether to bridge (see Appendix A for details).

Since confidence information is important, following each comparison the student is allowed to change his or her confidence about a previous answer as well as changing the answer itself. We have noted in some human tutoring sessions that asking students to commit to a confidence rating causes them to put additional cognitive effort into analyzing the example situations.

Topics

Each network corresponds to one (and only one) topic area. In this paper we describe results for only the topic "existence of contact forces between two stationary objects" whose network is shown in Figure 2. Another topic given to some subjects in this study was "equality of contact forces between two stationary objects." Though we have thus far tested only these two topics with students, the tutor is capable of teaching an arbitrarily long sequence of topics. For each topic area the goal of the tutor is to help the student understand the qualitative physics of a key target situation which exemplifies an expert application of the concept. If the student responds incorrectly to the target situation, then the bridging analogies strategy is invoked. If the student responds correctly to the target problem, the tutor moves to the next topic. If there is any overlap between example situations in the topic areas, the program takes into account previous answers about the situations.

During comparisons the tutor asks the student to explain why her answers to two different problems were the same, or to explain why they were different. For the tutor to be able to do this, all of the problem questions within a network must be similar. This tends to limit topics to relatively small curricular areas, and it also limits the diversity of examples that can be included in a given network. For example, for a network dealing with the existence of static forces all of the questions have multiple choice answers similar to "xxx IS exerting a force up on yyy" and "xxx DOES NOT exert a force up on yyy." For the network dealing with the equality of forces (i.e., that the forces are equal and opposite) all of the problem questions have multiple choice answers of the form "xxx pushes harder," "yyy pushes harder," or "they both push the same amount."

Other Features of the Tutor

The bridging analogies tutor is written in PASCAL and runs on an IBM PC. The tutor has been designed to be domain independent. Domain-specific information is stored in text files separate from the code, and these are processed at run time. The specification of the network structure, as well as information about the example situations, can be easily modified, or new domains created from scratch, by straightforward editing of text files. However, it is not easy to design *effective* networks of anchors and bridges. This requires research iterations with human subjects.

Figure 3 shows examples of the graphic pictures that accompany the descriptions of the example situations. Created using the PC-Paint (trademark of Mouse Systems) program, their inclusion is optional.

All student input is through menu selections. The use of menu-driven options, and of multiple choice answers, allows us to bypass difficult problems of language recognition involved in interpreting student input. Multiple choice answers can include reasons for each of the answers, incorporating items which catch common wrong answers and misconceptions. Usually, reasons for a student's answer are just as important as the correctness of the answer.

Example situations can have two levels of description associated with them. After being presented with a description of a new situation, the student is asked if she wants a more detailed description. The detailed description attempts to address possible areas of confusion.

5 FORMATIVE EVALUATION OF THE TUTOR

The purposes of this phase of evaluating the tutor are to determine what aspects of the simulated strategy are effective. We are also interested in information leading to alternative computer tutoring strategies to be used where bridging is not effective. In addition, we seek information concerning how effective the tutor is for a pair of students using it in collaboration.

Experimental Method

A pretest was administered to students in several university classes. A total of 180 students were given four example situations with questions and confidence ratings. None of the questions were the same as the ones in the tutor's network. One situation was similar to the target problem, involving contact forces on a medium sized object resting on an inflexible object. Another was similar to the initial anchor, where one object was held in a hand. To conduct a formative evaluation of the tutor we were looking for students for whom the particular network we used was designed, that is, those who demonstrate understanding of the anchor and lack of understanding of the target problem. Eighty four percent of the students answered the anchor-like problem correctly, and 68% answered the target-like problem incorrectly. Sixty six percent got both the target wrong and the anchor right. These figures are in rough agreement with the results of previous studies (Clement & Brown, 1984).

Of the students who took the pretest, 53% volunteered to be subjects for a one-hour videotaped interview study, for which they would be remunerated 5 dollars if chosen. The distribution of correct answers on the pretest did not differ significantly between those who volunteered and those who did not.

Twenty five students were selected for interviews. The selection of students was biased toward those for whom we thought the strategy, and the network used, were most appropriate. Those who answered the target incorrectly with high confidence and the anchor correctly with high confidence were given preference. Three of the sessions involved pairs of students using the tutor (they were asked to discuss each question and decide together what answer to give). There were technical problems (including power outages), and several students unexpectedly answered the original target correctly with high confidence right from the start, in which cases the bridging analogies strategy is not invoked. In the end we had a total of 15 sessions (3 of which involved pairs of subjects), lasting approximately 45 minutes each, with useful data from the book on the table network.

Data Analysis

Three types of data were collected: notes taken by the interviewer during the sessions, a computer trace of the tutor's actions and user responses, and videotaped recordings of the sessions. We have completed a first pass analysis of the videotaped recordings (see Schultz, Murray, Clement, & Brown (1987) for a protocol analysis report). Here we will outline our method for analyzing the computer traces. The trace records important events in a readable form (see Appendix B for an example). This information was transferred (by hand) to a diagrammatic representation, as in Figure 5. Circled nodes are those visited. Next to each circled node the correctness of the answer ("OK" for correct, "X" for wrong) and the confidence rating (from 1-blind guess to 5-I'm sure) are recorded. Single line arrows go from targets or anchors to the next bridge. Double lined arrows go from bridges to targets, and

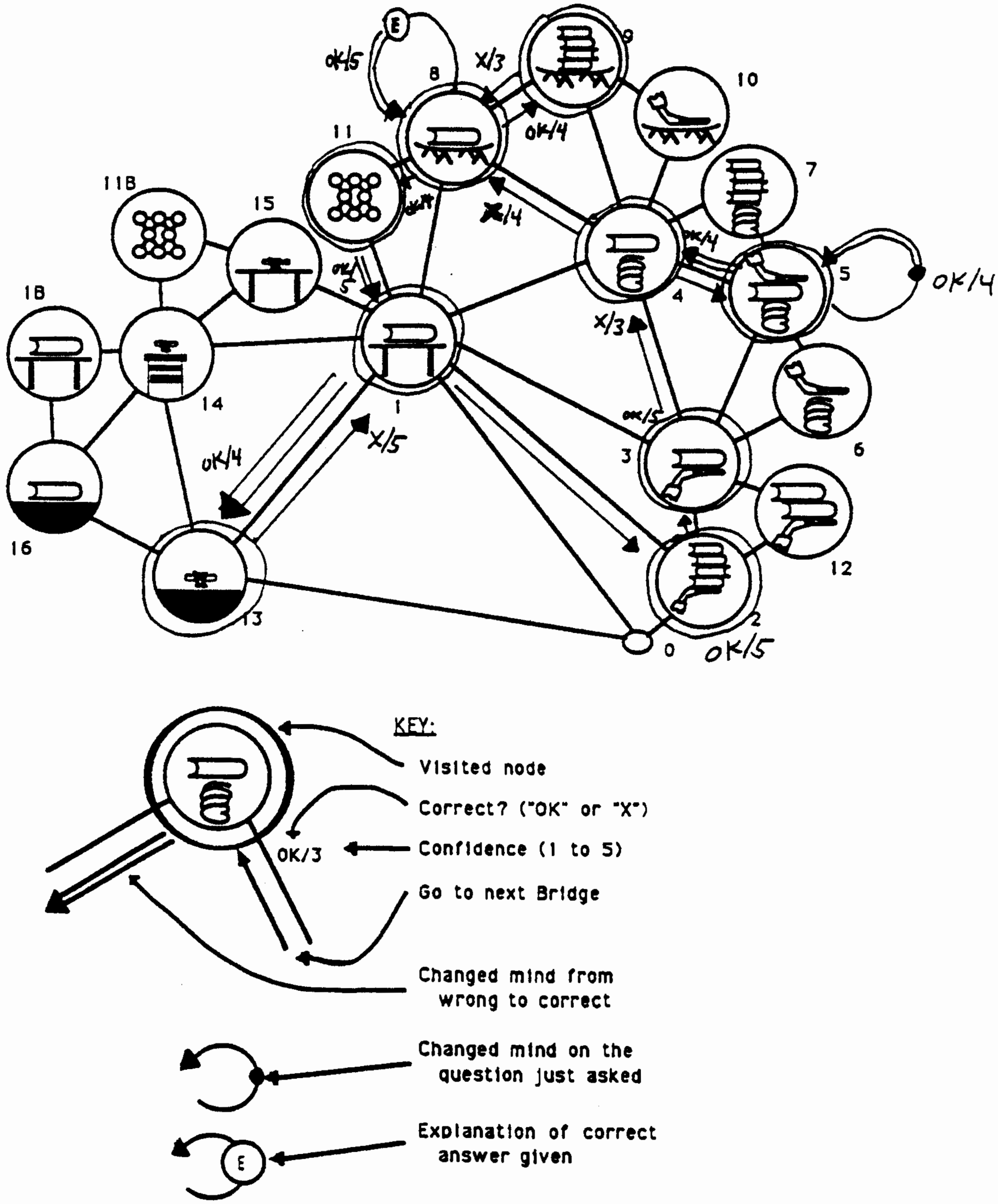


Figure 5. Diagramatic Representation of the Program Trace

show where the subjects changed their minds from wrong (or correct with low confidence) to correct (with medium to high confidence) for the problem pointed to. Circular lined arrows which connect a node with itself represent changing their minds about the question most recently asked, or, if an "E" is shown, they were given an explanation and asked the question again (as would happen if no bridge were available between two nodes). Not represented in the diagram, but easily inferred from the information in it, are the places where the tutor asks for a comparison between the most recently given situation and a previous one. Recall that after each such comparison the student is asked if she wants to change her mind on either or both situations.

Data analysis is greatly enhanced by the diagrammatic representation. The subjects' path through the example network, their changes of mind, and where explanations were needed, are highly visible. The regions of the net where the subjects needed the most bridges are apparent. From this we can infer which attributes of the original target contributed to their misconception.

For each subject's network, the following information was summarized: the number of nodes visited, the number of explanations given, the number of changes in mind, the total change in the original target problem from the beginning to the end of the session, the "critical points" (defined below), and the "focal regions" (defined below).

The network was divided into four "focal regions," each addressing a different key attribute of the target problem. The "focal regions" data specifies in which regions the subject visited two or more nodes.

"Critical points" changes give a more reasonable estimate of the student's thinking than the total changes of mind, and are defined as follows. Often, one realization on the part of the subject leads to changes of mind for several situations. For example, a student who is being shown a hand on a flexible table situation (node 10) is four recursive levels deep in the algorithm, having answered incorrectly questions about a book on a table, a book on a flexible board, and many books on a flexible board. Assume that the hand on the board is a critical analogy for this student. After answering it correctly the student may realize that several of his or her previous answers should be changed. As the algorithm unwinds (see Figure 4), the student would be given three comparison questions in a row, and opt to change the previous answer for each of these to the correct one with high confidence. When the subject changes his mind in a "chain" like this, all changes can be the result of a single change in belief. "Critical points" are the first change in any chains of one of more changes.

Summary of the Results

Following are our observations from the protocols and the computer traces. Since this evaluation of the tutor is formative in nature we did not perform a detailed statistical analysis, nor does our sample size permit this kind of analysis. Much of our interpretation of the data is specific to the observations. To improve readability we will fold these interpretations in with the data in this section. Interpretations specific to observations follow those observations and are in italics. Some observations are accompanied by illustrative quotations from the videotape protocols.

At the end of the sessions all but one of the 15 subjects had high confidence in the correct answer to the original target situation, and most of their verbal reasons indicated that the correct answer made sense to them at an intuitive level. Nine of them started the session with high confidence in the wrong answer to the original target. (The rest started with low confidence in the wrong answer to the original target.) *This is an indication that most of the subjects improved their understanding of Newton's third law as it apples to static situations. An exemplary quote is given below*:

> *[A student is considering the book on the table problem.]*
> *Um . . . I still don't think that . . . unless we're talking about molecules. I mean, unless we're talking about the book on the table [pause] would actually mean that there* are *molecules pushing up on the book. I guess just pushing back* would *be exerting a force. [pause] Well, actually it* can't *be wrong if the molecules are pushing up, then it* has *to be exerting some kind of force on the book.*

Analysis of the critical points of the session shows that the analogical comparisons alone (without an explanation) motivated at least one change of mind from an incorrect to a correct answer for 50% of the subjects.

> *[Student is asked to compare situations 8 and 1]*
> *. . . and it's not sturdy at all, and so it* has *to be exerting some kind of force up to keep . . . [laughs] This is like blowing my whole theory. The book on the board, the book is pressing down on the board, and it's bending it. In order for the board not to bend all the way, it has to exert some kind of pressure, but it kind of makes me think about the book on the table exerting force.*

This suggests that bridging and analogical comparisons are effective instructional strategies for many students at least locally (i.e., if the analogy *is not too distant), though perhaps not when used exclusively.*

Nine of the subjects needed at least one explanation at some point in the session. *These indicate places in the network where another bridge could be added.*

There were two instances of what we call regression, in which a student, after a comparison question, attempted to change his or her mind about an anchor from correct to wrong. Also, there was one case in which a student did not answer correctly about the original anchor with medium or high confidence (in which case the program had to explain the answer to the original anchor). *That these numbers are low indicates that the anchor used was an effective starting place for remediating the primary misconception.*

Most of the subjects needed to pass through a node which described all matter as being composed of molecules connected by springy bonds (node 11). Since designing the network we have come to regard such bridges as "microscopic models." Microscopic models of physical phenomena may give students a deeper causal explanation of the phenomena than other bridges. *It is not clear whether their misconception concerning the book on the table could have been remedied for these subjects without the use of a microscopic model.*

On the average, 7.5 example situations were needed to bring the subject to a successful conclusion of the session. *This supports the need for many nodes in the networks, and indicates that the original anchor and original target were indeed distant analogies.*

Most of the explanations that were given occurred in the focal region of the molecular model example. *This indicates that more example situations are needed to expand this area of the network.*

Only one of the 15 subjects spent any time in the focal region of the network between the book on the spring and the book on the hand. *This suggests that the book on the spring would serve equally well as the original anchor. (This finding is in agreement with the results of previous experiments (Clement & Brown, 1984).*

Included in our data is a compilation of the reasons students gave for their answers to selected example situation questions. We will not discuss these in detail here, but note that this data could be used to improve the network by including new situations which address common misconceptions indicated by these reasons.

> *I mean, I know that the fly is staying there by some force, but it's gravity, it's not that the . . . I don't think that the road is pushing up on the fly. It's gravity that's keeping the fly there. You know what I mean?*

We did not observe that asking confidence questions caused a significantly greater cognitive involvement in the subjects, as we did for some previous human tutoring sessions. We could not observe the contrary either, because the interviewer encouraged subjects to think aloud and give reasons for answers, which they did fairly constantly when answering the example situation questions, *before* they were asked the confidence question. *The question of whether the confidence question stimulates additional reflection needs further study.*

There was no evidence of subjects being overly annoyed or frustrated by the low feedback nature of the response strategy. In general, they seemed to accept this style without question and remained focused on thinking about the situations and comparisons given.

> *[One student's comment on the tutor's feedback:]*
>
> *It's helpful when it goes back and tells you to look at your answers, and then, once you've thought about it . . . it forces you to think about it, because it keeps bringing it back, and then if you're totally wrong on something . . .* then *it will say to you, you know,* this *is the way it is, and you're like—it makes you think at first, instead of kind of giving you a hint at first. I think that's better.*

Since our experimental setup is somewhat dissimilar to a student using a computer tutor by herself, conclusions based on this observation are limited. The data shows no evidence that a low feedback strategy is ineffective when implemented on a computer tutor.

We observed that, though the algorithm used to traverse the network was not complex (see Appendix A), the presentations of example questions and comparison questions did not seem to be predictable or become rote to the subjects. If one knew or inferred the algorithm being used, then it would be possible to infer whether a previous answer was correct by observing the pattern of comparisons given. This did not happen in our sessions. Only in two cases did the students show any sign of thinking about what was going on in the background—in the "tutor's head." In neither case was this a distraction. One case is given below:

[Two students discussing the book on a flexible board problem (node 8)]

[Student A:] Well, if we're going to stick by our last answer, I think we'd have to say the board is exerting a force up on the book . . . [proceeds to give reasons]

[Student B responds:] You know what they're going to say to that *one, don't you? They're gonna give you the book on the table example. Then what are you going to say?*

Three of the sessions involved pairs of subjects using the tutor. They were asked to discuss their beliefs about the questions and come to an agreement about an answer. These sessions were quite animated. There was a high degree of motivation, verbalization, and serious reflection about the problems. Only rarely did the interviewer need to intervene and ask for verbalization of thoughts or reasons for answers. Here is a sample interaction:

[Two students discuss the impact of the molecular model previously described to them]

C: Because the table is, uh, not springy.

D: But the molecules are, and the table is made of molecules.

C: But, depending on the kind of wood, assuming that it's a stronger wood, the molecules aren't as springy, because they said there were springy molecules . . .

D: Not as *springy, but wouldn't they still be springy, some what? They'd have to be.*

C: They'd be springy, but not to the weight of a book. Maybe to the weight of a couple of anvils.

D: Um hmm. Maybe not as *springy to a book, maybe there is a very slight force there . . .*

C: Yeah.

D: . . . giving way, given even how slight it may be, maybe there's still a force there.

In order for the bridging analogies strategy to be effective, students must be motivated to think seriously about the example situations presented. We do not know whether a version of the bridging analogies tutor designed to be used in a stand-alone fashion (without an interviewer present) would be sufficiently engaging for students, but our observations suggest that a stand-alone version may be sufficiently engaging for pairs, or even groups, of students.

Students' comments about the program, elicited at the conclusion of the session, were in general favorable regarding the effectiveness of the learning experience and their enjoyment in using the tutor.

It's a really great program, because you write, you put down what you think your answer is without having any clue, and then you put, you know, if you're confident or not, and then you go back and it keeps going back saying "why, why, why," and it gives you different examples.

6 DISCUSSION AND CONCLUSIONS

The current implementation of the bridging analogies strategy has served well as a research vehicle. Videotaped tests showed that the strategy can engage students and bring about significant changes in belief. The tutor was designed to be used as a research vehicle, and this evaluation has given us information concerning what aspects of the current implementation are not flexible or complete enough to act as a stand-alone tutor.

In this section we will first give recommendations, based on the results summarized in the previous section, for extensions of the present study and straightforward improvements to the existing tutor. Then we will discuss longer range possibilities for a more complete re-design of the tutor using artificial intelligence techniques. The knowledge representation and inferencing power of AI technology may provide the leverage needed to design a robust tutor which uses bridging analogies as its central strategy.

Further Research Using the Bridging Analogies Tutor

The results of this study have several indications for continuing research on the bridging analogies tutor:

- add more bridges to the network in areas where students needed explanations;

- incorporate one or two levels of hints to be given before the explanation reveals the correct answer to the student;
- construct networks for other topic areas in which to test the strategy;
- conduct further tests with pairs of students;
- test how the introduction of the molecular model near the beginning of the session (it currently is introduced near the end) affects student's beliefs;
- incorporate the reasons observed in students' thinking aloud explanations into multiple choice answers.

Designing the Example Situation Network

An example situation network can be described in two ways: in terms of example difficulty, or in terms of example features. We will discuss each below. Since the final version of the network used was influenced by empirical results from human tutoring studies, our analysis is post-hoc. In this section we will analyze pedagogically important characteristics of the network and offer some design principles that can guide the design of such networks manually or that could be incorporated into a system which automates the process of choosing or generating appropriate bridges.

First we will discuss the network in terms of example difficulty. Traversing the net along the outer edge will touch every example situation, yielding a sequence that can be interpreted as starting with the original anchor situation, and as progressing from easier to harder problems (see Figure 1). Each bridge is intended to split the *conceptual distance* between two other situations. The bridging analogies strategy is seen in this light as efficiently searching the space of examples using a divide and conquer approach. The sequence of examples between the current target and the current anchor is divided at the bridge specified by the network. The network thus encodes the pedagogical knowledge of estimated conceptual distance (as opposed to the cardinal distance in a linear sequence of examples, where the third would be half way between the first and fifth).

Though the divided conceptual distance perspective on the example net is valid for the network used here, a feature dimension based perspective is more interesting from an AI viewpoint. Research has uncovered several reasons students give for believing that the table cannot push up on the book, but a hand *can* push up on a book. Each reason involves an attribute (or attributes) which distinguishes the two examples. Some of these reasons are: a person is *doing* something, but the table is just in the way; a person's arm can move to adjust to the book's weight, but the table cannot; an arm's purpose is to exert forces, but a table's is not. Thus the student sees one or several features of objects as being relevant to its ability to exert forces, all of which are irrelevant from a physicist's standpoint. The network factors out these possible features by providing example situations with different sets of relevant and irrelevant features. For example, the spring can move and is designed to exert a force, but cannot decide on its own to volition to do so.

We will analyze example situations using a methodology similar to Rissland, Valcarce, and Ashley's (1984) analysis of example cases in case-based reasoning. We distinguish key dimensions along which the example situations differ. The dimensions are shown in Figure 6. These dimensions are pedagogically relevant abstractions over the surface features of the examples. For example, springs, palm-up hands, and flexible boards are all surface features which map to "movable" in the dimension space. The dimensions have qualitative values as shown in Figure 6. For example, volition has two possible values, true and false; solidity has five: + + + (the road), + + (a rooftop), + (tables and boards), and − (springs and hands). Each dimension has a polarity which determines whether increasing its value is likely to result in a problem situation which is more difficult or less difficult. Increasing the solidity or gravity dimensions tend to make the problem harder (i.e., the existence of the upward force is less apparent), so they have positive polarity. The remaining dimensions are all of negative polarity.

Our feature dimension analysis of the relationships among the network nodes reveals that the selection of an appropriate bridge may involve some rather complex inferences. No simple feature-based algorithm was found which would account for all of the bridges in the network. Given target and anchor situations, the generation of an appropriate bridge is an underconstrained problem. There exist many potentially applicable bridges. Below we identify several

Dimension	Values	Polarity	Definition
solidity	–,+,++,+++	pos	solidity of the substrate
gravity	true, false	pos	is gravity the most salient force?
salience	–,0,+,++	neg	size of the force involved
motion	true, false	neg	moves visibly when force is applied
volition	true, false	neg	can "decide" to exert a force
mechanism	–,0,+	neg	its purpose includes pushing or pulling

Figure 6. Proposed Feature Dimensions for the Example Network

example selection strategies that would account for the bridges in the empirically derived network.

Intermediate Examples. The most pervasive strategy is creating a bridge by changing the value of one or more dimensions of the anchor toward the target. For example, the bridge between the book on the table (node 1) and the book in the hand (node 3) is the book on the spring (node 4). This differs from the anchor by a change in the volition dimension from true to false.

Dimension Importance. Deciding which dimension to change may be constrained by assigning an importance or priority to the example dimensions. This importance may reflect how much or how often a dimension contributes to a misconception. For example, for the bridge described above, the target and the anchor differed along the salience, motion, mechanism, and volition dimensions, and only volition was changed in the anchor to make the bridge. It may be that the belief that "no volition implies no force" is the most common misconception. Another possibility is that the volition misconception is the *easiest* to remedy, so it is attempted first.

Constraints on Example Dimensions

A program which automatically generates intermediate examples must be able to decide: 1. *how many* dimensions in the anchor to change, 2. *which* dimension(s) to change, and 3. *how much* to change each by (where the dimension has more than one value that lies between the anchor and target). There exist some constraints on the values a dimension can take which may help with these inferences. A "true" value on volition implies a "true" value on mechanism. That is, all things which can decide to exert a force have a mechanism for doing so. For example, the bridge between the book on the table and the book on the spring is the book on a flexible board (node 8). The board differs from the spring in two dimensions, solidity and mechanism, but this can be seen as a change in the solidity dimension which necessitated the change in the mechanism dimension.

Also, if example situations are being selected from a predefined set, then the system is constrained to using combinations of dimension values that are available in the example knowledge base. If examples are being generated on the fly, there will be additional naive physics constraints.

Extreme Cases. In situations where either the target or anchor is an extreme case of the other, an example may be chosen which is intermediate along the dimension which is extreme. For example the bridge between many books in the hand (node 2) and a book in the hand (node 3) is two books in the hand (node 12). The target and anchor differ only in the salience dimension, and each has an extreme value in that dimension. The bridge has an intermediate value in the salience dimension. This strategy is similar to the "slippery slope" type of argumentation discussed in Rissland et al. (1984). Another example of using extreme cases occurs when the *bridge* chosen is an extreme case—chosen to focus the student's attention on a particular dimension. For example, the bridge between the book on the flexible board and the book on the spring is many books on a flexible board (node 9). The target and anchor differ only in the solidity dimension, and only by one value. The only change that one could make to bring the anchor closer to the target would yield a value identical to the target. Therefore the intermediate example method is inapplicable. The student does not notice that a book on a flexible board and a book on a spring are two examples of the same phenomenon. She may not notice (in her own thought experiment) that the board would bend a bit under the weight of the book. In this case a bridge is selected which focuses attention on this fact by giving an extreme case in the salience dimension. This example shows that the student will not always perceive the same feature dimensions in an example situation as the instructional designer. The designer categorized the flexible board as having the motion attribute, while a student may not.

Personal Analogies. Another bridge (or anchor) selection strategy noted is encouraging the student to imagine himself in the place of one of the objects and noticing whether he would "feel" a force. Examples are the hand on the flexible board (node 10) and the hand on the spring (node 6).

Microscopic Models. As noted in the results section, the introduction of the molecular model of matter (node 11), that is, that matter is composed of molecules connected by springy bonds, was a critical point for many subjects. This indicates that introducing an explanation at a causal level was important for them. Other studies raise the importance of models in teaching and understanding physical phenomena (deKleer & Brown, 1980; Stevens & Collins 1977; White & Frederiksen, 1986). The introduction of the hand and the spring examples can be seen as offering a model of springiness with which the student can view seemingly rigid objects in a new way. These examples provide an explanation for why the table exerts a force—that is, that it has some finite, though imperceptible, amount of springiness. The microscopic model provides a yet deeper explanation. It explains or gives a model for how rigid objects can be springy. Our use of the term "model" is more limited than the "runable mental models" discussed in deKleer & Brown and elsewhere. The strategy we are proposing here (i.e., using a microscopic model as a bringing analogy) is not a complex mentally runable mechanism or causal chain, but an explanatory model.

Above we have suggested several strategies for the selection of bridging examples. These strategies could also be applied in instructional contexts where the overall strategy is one of generalizing or refining a concept by presenting new examples. In that case new examples build incrementally on previous examples, as compared to bridging analogies where new examples bridge between a target and an anchor.

Using strategies like the ones identified above to select bridges would enhance the flexibility of a computer tutor which used the bridging analogies strategy. Such a tutor would not be constrained to having only one possible bridge between a target and anchor. But as we have seen, defining these strategies is difficult. And the question of when to use which strategy is an open question needing further investigation. Also, using an empirically tested network of bridges has the advantage of being more reliable. The "intelligence" of an AI inference system for selecting bridges would be limited, so that the system might give spurious or nonoptimal examples.[5]

Multiple Tutoring Strategies

An important area of ITS research is representing tutoring strategies and selecting multiple strategies (Clancey, 1982; Stevens & Collins, 1977; Woolf & McDonald, 1984; Woolf & Murray, 1987). Such strategies can be sensitive to students' knowledge state and learning needs (as diagnosed in a pretest or inferred incrementally with a student modeling mechanism). They can also be sensitive to the pedagogical properties of the topic or task given to the student. Obviously, other strategies are needed for topic areas with characteristics for which the bridging analogies strategy was not intended, such as teaching skills or teaching conceptual knowledge where there are no common anchors. Also, our data indicates that, for some students, alternative strategies are needed to remedy conceptual difficulties in the physics domain we used. Explicitly representing and using the example selection strategies proposed in the last section may benefit some students, but we have little data as yet telling us when to employ these strategies. In addition, the focus on analogical thinking and low feedback may be ineffective for some students.

The addition of a student modeling mechanism might enhance the system's performance. Below we outline how such a mechanism might

[5] Note however that tutoring expertise is unlike some other domains being examined by AI or cognitive science. There is usually no great risk taken in making a nonoptimal inference. Compare a tutor presenting a nonoptimal example situation with the more serious error of a doctor prescribing a nonoptimal drug.

be used to select among alternate strategies. Though our analysis of the protocols is still in progress, we have tentatively identified several types of student behavior that a student modeling system might make inferences from. Some subjects consistently report low confidence levels on their answers, whether correct or incorrect. They may be uneasy with the low feedback nature of the current strategy, or may have only weak intuitions on which to base their answers. For such students one could provide more positive feedback to correct anchor cases. Some subjects self-corrected only after being given explanations. If this pattern is noticed, a microscopic or other explanatory model might be given early on in the session. For students who do not change their answers even after being given an explanation of the scientific viewpoint, the bridging analogies strategy may be inappropriate (for use in the domain in question). Also, for some students the tutor may not be able to find an anchor. In such cases the tutor could switch to a more didactic form of tutoring, or present examples explicitly labeled as positive or negative exemplars of the concept being taught. Students who show a pattern of changing their answers from incorrect to correct with high confidence, without being given explanations, may be best suited to the bridging analogies strategy. We may be able to infer that they are proficient and confident at using analogies to reason about situations. The depth of bridges (and subbridges) a student is given corresponds to the number of incorrect answers she has given in a row without having had a change of mind or being given a correct explanation. The depth of bridging that a student is allowed to be given before receiving an explanation or hint could be limited according to how well the student is suited (as described above) to the bridging analogies strategy.

The example selection strategies proposed in the previous section might benefit from a student model that recorded students' reasons for their answers. This information could be ascertained using natural language interpretations of students' response to the comparison questions (assuming that the AI state of the art allows robust language understanding). Alternatively, the most common reasons given for questions could be identified experimentally and put into a multiple choice form for students to select from after answering a question.

Guidelines for Future Research

Below we summarize guidelines for future research of this type:

1. Rely on previous cognitive studies of a tutoring strategy.
2. Include knowledge about common misconceptions and their remediation in the tutor.
3. Perform a formative evaluation of the computer tutor, including taped sessions, early on, before large effort is expended in coding a robust tutor.
4. Be explicit about the tutoring strategy used.
5. Research the strengths and limitations of a context specific ("strong method") strategy rather than more general or less well-defined strategies.
6. Test context specific strategies in isolation before combining them in one system.
7. Use self-rated confidence measurements to provide additional information about the strength of student answers.

As mentioned previously, our focus on the process of tutorial interaction contrasts with many ITS research efforts which focus on domain task analysis, modeling domain expertise, or designing learning environments. Another comparison involves the fact that the bridging analogies tutor is aimed at changing students' intuitive beliefs and remediating conceptual difficulties, while most existing ITS systems are designed to teach or debug skill knowledge. We foresee other context-specific tutoring strategies being developed to teach science. The bridging analogies strategy uses analogies to extend the locus of applicability of existing intuitions. It is most applicable to instructional contexts where qualitative knowledge is being taught, commonly held misconceptions exist, and anchoring analogies exist. Other strategies aimed at teaching conceptual knowledge may also be important. Causal model evolution (White & Frederiksen, 1986) and induction from exemplars (Tennyson, 1980) are two examples. Learning science requires the acquisition of many types of knowledge. Strategies are also needed to teach factual knowledge, problem-solving skills, scientific inquiry skills, and to guide students' exploration of learning environments that provide new types of experiences.

In summary, there is evidence that the strategy works effectively for many students, and that the specific network used was effective (though the data indicates places where it could be improved and extended). We are encouraged by the levels of cognitive involvement exhibited by the students while doing the thought experiments, especially in contrast with the more passive role that students tend to take in traditional explanation oriented methods of teaching, or in computer aided instruction that provides immediate feedback for wrong answers. The formative evaluation provided useful information concerning: 1. the improvement of the specific example network used, 2. improving the bridging analogies strategy, and 3. issues in using AI technology to design a more flexible and robust tutor that incorporates the bridging analogies strategy.

We have relied on previous cognitive studies of misconceptions and tutoring strategies in formulating our ideas, and this research exemplifies the importance we place on such studies in the early stages of designing intelligent computer tutors. Further work is needed to codify human tutoring strategies, test their effectiveness in computer learning environments, and determine how to orchestrate the different strategies in computer tutoring sessions.

REFERENCES

Anderson, J., Boyle, F., Farrell, R., & Reiser, B. (1984, August). Cognitive principles in the design of computer tutors. *Proceedings of the Cognitive Science Society.* Austin, TX.

Brown, J.S., Burton, R.R., & deKleer, J. (1982). Pedagogical, natural language, and knowledge engineering techniques in SOPHIE I, II, and III. In D. Sleeman & J.S. Brown (Eds.), *Intelligent tutoring systems*. New York: Academic.

Burton, R., & Brown, J.S. (1982). An investigation of computer coaching for informal learning activities. In D. Sleeman & J.S. Brown (Eds.), *Intelligent tutoring systems*. New York: Academic.

Clancey, W.J. (1982). Tutoring rules for guiding a case method dialogue. In D. Sleeman & J.S. Brown (Eds.), *Intelligent tutoring systems*. New York: Academic.

Claxton, G. (1985). Teaching and acquiring scientific knowledge. In T. Keen & M. Pope (Eds.), *Kelly in the classroom: Educational applications of personal construct psychology*. Montreal: Cybersystems.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*, 66–71.

Clement, J. (1987, July). Overcoming students' misconceptions in physics: The role of anchoring intuitions and analogical validity. *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*. Ithaca, NY.

Clement, J., & Brown, D. (1984). *Using analogical reasoning to deal with 'deep' misconceptions in physics* (Tech. Rep. No. 95). Amherst: University of Massachusetts, Scientific Reasoning Research Institute.

Collins, A. (1977). Processes in acquiring knowledge. In R. Anderson, R. Spiro, & W. Montagne (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Erlbaum.

deKleer, J. & Brown, J.S. (1980). *Mental models of physical mechanisms* (Tech. Rep. CSI-3). Palo Alto, CA: Xerox Palo Alto Research Center.

diSessa, A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Erlbaum.

Fredette, N., & Lochhead, J. (1980). Student conceptions of simple circuits. *The Physics Teacher, 18*, 194–198.

Ginsburg, H., & Opper, S. (1979). *Piaget's theory of intellectual development: An introduction*. Englewood Cliffs, NJ: Prentice-Hall.

Halloun, I.A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics, 53*, 1043–1055.

Lawson, A. & Wollman, W. (1975). Physics problems and the process of self- regulation. *The Physics Teacher* (pp. 470–475).

Lochhead, J. (1983, September). Constructivist approaches to teaching mathematics and science at the college level. *Proceedings of the Fifth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Vol. 2*. Montreal, Canada.

McCloskey, M. (1983). Intuitive physics. *Scientific American, 48*, 122–130.

McDermott, L. (1984). Research on conceptual understanding in mechanics. *Physics Today*, 24–32.

Piaget, J. (1972). *The principles of genetic epistemology*. New York: Basic Books.

Rissland, E., Valcarce, E., & Ashley, K. (1984, August). Explaining and arguing with examples. *Proceedings of the American Association of Artificial Intelligence*. Austin, TX.

Schultz, K., Murray, T., Clement, J., & Brown, D. (1987, July). Overcoming misconceptions with a computer-based tutor. *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Vol. 3*. Ithaca, NY.

Shute, V, & Bonar, J. (1986, August). Intelligent tutoring systems for scientific inquiry skills. *Proceedings of the Cognitive Science Society*. Amherst, MA.

Steinberg, M.S. (1983, June). Reinventing electricity. *Proceedings of the International Seminar on Misconceptions in Science and Mathematics*. Ithaca, NY.

Stevens, A., & Collins, A. (1977). The goal structure of a Socratic tutor. *Proceedings of the National Association of Computer Management Conference*. Seattle, WA.

Streibel, M., Stewart, J., Koedinger, K., Collins, A., Jungck, J. (1987). MENDEL: An intelligent computer tutoring system for genetics problem solving, conjecturing, and understanding. *Machine-Mediated Learning, 2(1 & 2)*. 129–159.

Tennyson. R.. & Park. O. (1980). The teaching of concepts: A review of instructional design research literature. *Review of Educational Research, 50*. 55–70.

Tversky. A.. & Kahneman. D. (1974). Judgment under uncertainty: Heuristics and biases. *Science. 185*. 1124–1131.

von Glasersfeld. E. (1988). *Constructivism in mathematics education*. Hillsdale. NJ: Erlbaum.

Van Lehn. K. (1983). *Felicity conditions for human skill acquisition: Validating an AI-based theory* (Tech. Rep. CIS-21). Palo Alto. CA: Xerox Palo Alto Research Center.

Wandersee. J.H. (1983. June). Students' misconceptions about photosynthesis: A cross-age study. *Proceedings of the International Seminar on Misconceptions in Science and Mathematics*. Ithaca. NY.

White. B.. & Frederiksen. J. (1986. August). Intelligent Tutoring Systems Based Upon Qualitative Model Evolution. *Proceedings of the American Association of Artificial Intelligence*. Philadelphia. PA.

Woolf. B.. & McDonald. D. (1984. August). Context-dependent transitions in tutoring discourse. *Proceedings of the National Conference on Artificial Intelligence*. Austin. TX.

Woolf. B.. & Murray. T. (1987. July). A framework for representing tutorial discourse. *Proceedings of the American Association of Artificial Intelligence*. Seattle. WA.

APPENDIX A
Program Data Structures and Algorithm

Domain information is stored in text files. making the bridging analogies tutor domain-independent. There are three domain files: the Topic File. the Network File. and the Situation File. The Topic File determines the sequence of topics (each corresponding to one example network) to be presented (this paper discusses only one of the topics). and specifies which of the examples in the topic's network will serve as the original anchor and original target situations.

The Network File contains triplets of numbers. each number representing an example situation node. Each triplet defines a unique bridge between two other nodes. and defines the network as in Figure 2. The table thus defined is searched when the program bridges between a target and an anchor*.

The Situation File contains text for the example descriptions. detailed example descriptions. questions. multiple choice answers. correct answers. an explanation for the correct answer. and a pointer to a bitmap picture file. This information is organized using keywords. and is parsed and loaded into PASCAL record structures when the program is executed. These record structures also keep track of the student answers and confidence ratings for each node.

The bridging analogies strategy is an algorithm for traversing the network. The algorithm has been described operationally in Section 3. It appears in pseudo- code below.

An Algorithm for the Bridging Analogies Strategy

The procedure Teach-Topic is called iteratively for each topic in the Topic File. The Bridge procedure contains the algorithm for traversing the network.

The program is capable of "finding" the original anchor. but this capability was switched off (and the original anchor used is the one specified in the Topic file. as described above) in the test runs we report in this article. The Find-anchor procedure finds an appropriate original anchor by presenting the student with all the situations analogous to (i.e.. linked to) the original anchor in order of increasing simplicity. until the student answers one correctly with high confidence. For large networks this may be more efficient than assuming that the best original anchor is the very simplest example.

Procedure Teach-Topic (original-target. original-anchor)

Describe-and-question(original-target)

Describe-and-question(original-anchor)

Compare-and-allow-change(original-target. original-anchor)

IF anchor wrong or conficence<3
THEN Explain(original-anchor)
IF target is right with confidence>2

THEN RETURN {no misconception}
ELSE Bridge (original-target. original-anchor)

Procedure Bridge (target. anchor)
{The anchor was previously answered correctly with

* "Wild cards" are incorporated. though they were not used in this study. They allow the designer to specify the use of a certain bridge whenever a specified node is the target (or anchor). regardless of the anchor (or target).

confidence >2 and the target was previously answered incorrectly or with confidence <3}

{1. find the bridge}

Find-a-bridge-between(target, anchor)
IF no bridge exists
THEN Explain(target) AND RETURN
Describe-and-Question(bridge)

{2. establish bridge-to-anchor analogy:}

IF bridge wrong or confidence<4
THEN Compare-and-allow-change(bridge, anchor)

IF anchor changed to wrong or confidence<3
THEN Explain(anchor)

IF bridge wrong or confidence<3
THEN bridge(bridge, anchor) {first recursive call}

{3. establish target-to-bridge analogy:}

Compare-and-allow-change(target, bridge)

IF target is wrong or confidence<3
THEN bridge(target, bridge) {second recursive call}

{4. At this point the target should have been answered correctly}

RETURN

Procedure Explain (an-example)

{itteratively gives the explanation and re-asks the example's question until the student answers correctly, or until given three explanations in a row, in which case the student is told exactly what to enter as the answer(!)}

(Note: Confidence ratings from 1 to 5 are for 'blind guess' to 'I'm sure' as in Figure 3.)

APPENDIX B
Example Section of a Trace File

Following is part of the computer trace of the session shown in Figure 5. All student input is in multiple choice form on the computer. A "!!" after a comparison record highlights where a student changed the answer.

START SESSION:
program version: 11/18/86.1
Name: Willard O. Oz
Session ID: 292

Topic: Existence of static normal forces
Original target, Original anchor: 13 0
Time: 11:47:15

. . . .
. . . .

ASK 13 fly on the road
Response: B (not OK) Confidence: 5

ASK 1 book on the table
Response: B (not OK) Confidence: 5

ASK 2 many books in hand
Response: A (OK) Confidence: 5

Comparison: 2 and 1: no change

**BRIDGE: T 13 A 2 B 3 11:50:35*
ASK 3 one book in hand
Response: A (OK) Confidence: 5

Comparison: 2 and 1: no change

**BRIDGE: T 1 A 3 B 4 11:50:47*
ASK 4 book on spring
Response: B (Not OK) Confidence: 3

Comparison: 4 and 3: no change

**BRIDGE: T 4 A 3 B 5 11:51:17*
ASK 5 hand on book on spring
Response: A (OK) Confidence: 2

Comparison: 5 and 3: change 5 !!

ASK 5 hand on book on spring
Response: A (OK) Confidence: 4

Comparison: 5 and 4: change 4 !!

ASK 4 book on spring
Response: A (OK) Confidence: 4

Comparison: 4 and 1: no change

**BRIDGE: T 1 A 4 B 8 11:53:03*
ASK 8 book on board
Response: A (OK) Confidence: 4

. . . .
. . . .