USING STUDENT MODELLING
TO DETERMINE WHEN AND HOW TO HINT
IN AN INTELLIGENT TUTORING SYSTEM

BY

GREGORY D. HUME

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computer Science
in the Graduate School of the
Illinois Institute of Technology

Approved ____________________________
Advisor

Chicago, Illinois
May, 1995

ACKNOWLEDGMENT

It is a pleasure for me to acknowledge everyone who has helped me in my pursuits. First, and most, is the gratitude I have for my friend, advisor, and mentor: Martha Evens. I will always be one of your disciples. Also, I extend a very special thank you to Joel Michael and Alan Rovick. Your enthusiasm, insight, and wit have made all of our work productive and enjoyable. To the past and current graduate students of the CST project, I am a lucky person to have worked with such a diverse and intelligent group of people. All of my colleagues at Valparaiso University supported me and, in particular, Bill Marion made it possible for me to spend time on this academic journey. Thank you Charley (my wife). You are the one who endures my eccentricities.

This work was supported by the Cognitive Science program, Office of Naval Research under Grant No. N00014-89-J-1952, Grant Authority Identification Number NR4422554, Grant No. N00014-94-1-0338, and the associated ASSERT Grants to the Illinois Institute of Technology. The content does not reflect the position or policy of the government and no official endorsement should be inferred.

G.D.H.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Term
	<u>Cardiovascular Variables</u>:
BR	Baroreceptor
CC	Cardiac Contractility
CNS	Central Nervous System
CO	Cardiac Output
CVP	Central Venous Pressure
HR	Heart Rate
IS	Inotropic State
MAP	Mean Arterial Pressure
RA	Arterial-Resistance
RAP	Right Atrial Pressure
SV	Stroke Volume
TPR	Total Peripheral Resistance
	<u>Expert Tutors</u>:
JAM	Dr. Joel A. Michael, Rush Medical College
AAR	Dr. Allen A. Rovick, Rush Medical College

Abbreviation	Term
	<u>Others</u>:
CAI	Computer Aided Instruction
CF	Confidence Factor
CI-Hint	Convey Information Hint
CLOS	Common Lisp Object System
CST	CIRCSIM-Tutor
CV	Cardiovascular
DKRS	Domain Knowledge Representation System
DLR	Directed Line of Reasoning Hint
EP	Error Pattern
DR	Direct Response
ITS	Intelligent Tutoring System
PT-Hint	Point To Hint
RR	Reflex Response
SS	Steady State

CHAPTER I

INTRODUCTION

A tutor must constantly determine *what* to tutor and *how* to tutor. In order to make these determinations, the tutor must carefully monitor the progress of the student throughout the tutoring session. This activity is referred to as student modelling. One goal of intelligent tutoring systems (ITSs) is to simulate this behavior. This thesis contains (1) a study of experienced tutors (JAM and AAR; all abbreviations can be found on page vii) in the field of cardiovascular (CV) physiology, and (2) the design of student modelling in an ITS. The most revealing behavior of JAM and AAR is their frequent use of hints. One-on-one tutoring, an environment ideally suited for hinting, is known to be a particularly effective teaching paradigm (Bloom, 1984). The crux of this thesis is the relationship between student modelling and hinting in a specific tutoring environment.

1.1 CIRCSIM-Tutor

CIRCSIM-Tutor (CST; v.3) is currently under construction; there have been two previous versions. CST is designed to tutor first year medical students studying cardiovascular physiology in a manner that approximates an expert human tutor's behavior. CST is considered an ITS because it (Burns & Capps, 1988) (1) has a formal representation of the domain knowledge, (2) estimates the capabilities of the student, (3) uses a natural language dialogue, and (4) selects from a variety of strategies to aid the student.

Traditional computer based educational programs are categorized as computer aided instruction (CAI). These programs do not have an explicit representation of the domain knowledge or an explicit representation of the knowledge state of the student. Therefore, CAI programs do not make decisions dynamically (Wenger, 1987); all decisions are explicitly determined by the computer code. CAI systems are often unable to adapt to the particular cognitive needs of the student (Park et al., 1987).

The goal of developing an ITS may be solely to explore the use of artificial intelligence techniques (Park et al., 1987), to study issues related to human cognition, to develop useful programs, or some combination of the above. Many people are working on CST; individual objectives and interests vary. My primary interests are to (1) model the student's behavior that determines what tactics (e.g., hinting) are used and to (2) identify the rules that dictate the use of hints.

1.2 Student Modelling

A student model is a representation of the student's knowledge (VanLehn, 1988) or, as I claim, a representation of all aspects of the student's behavior in the learning situation. For example, *how* a student says something may provide information that is as valuable as *what* a student says. An ITS's student model is a set of data structures that provide information to the computer tutor that contributes to decision making. Student modelling is the diagnostic process that provides the student model with information. For simplicity, both "student model" and "student modelling" are occasionally used to refer to both the data structure and the diagnostic process.

1.3 Hints

The correctness of student responses to problems or questions is an obvious input to the student model. My research suggests that there are other factors that experienced tutors use in modelling the progress of a student. I have observed that JAM and AAR regularly provide hints to students. The ability of a student to respond appropriately to hints is a major determinant of the tutor's selection of tactics.

From my experiences as an educator, I know that hinting is a common tactic and it is prevalent in human tutoring sessions. Yet, there is little literature, outside of my work (Hume et al., 1993, 1995a,b), that analyzes hinting. Hints were incorporated in CST (v.2). There was, however, no research performed on hinting. The resulting hints are monotonous and often inappropriate. Creating more natural hints is a major objective of CST (v.3).

A hint is a rhetorical device that is intended to either:

1. Provide the student with a piece of information that the tutor hopes will stimulate the recall of the facts needed to answer the question, or
2. Provide a piece of information that facilitates the student's making an inference needed to arrive at an answer to a question or the prediction of system behavior.

I have identified ten different forms of hints in two main categories. I have found little mention of hints in the ITS literature and no systematic study of hinting. Therefore, much research needs to be done to evaluate the effectiveness of hints and to identify the rules governing how to hint, and I have begun this research. Aside from identifying categories of hints, I have categorized student responses to hints. From this study, I have outlined rules for generating hints. I claim that (1) the student model is essential for determining *when* and *how* to hint and (2) student responses to hints should be used to update the student model.

1.4 Organization of Thesis

Chapter II is a review of the relevant literature in student modelling and hinting. Chapter III is a description of the CST project. Chapter IV describes (1) an analysis of JAM and AAR's tutoring behavior with respect to student modelling, and (2) a design of student modelling for CST (v.3). Chapter V is a detailed analysis of hinting. Chapter VI contains a summary of the thesis, an argument for the significance of this work, and a discussion of future research possibilities.

CHAPTER II

ITS LITERATURE

This section of the thesis reviews several ITS projects and the literature describing them, with the primary focus on the student modelling process. In 1970 Carbonell published a paper titled "AI and CAI: an artificial approach to computer assisted instruction," which described the initial research into ITSs. What distinguishes an ITS from a CAI system is the use of techniques normally associated with artificial intelligence. Examples of such techniques are explicit representations of knowledge, inferencing mechanisms and natural language dialogue (Wenger, 1987; Clancey, 1987). Ever since the beginning of research on intelligent tutoring the least understood component of an ITS has been the student modeller (Clancey, 1986a; Putnam, 1987).

2.1 Early Attempts at Student Modelling

Carbonell and Collins developed a geography ITS called SCHOLAR (Carbonell, 1970). It had an explicit representation of domain knowledge organized in a semantic net. Mixed initiative, natural language dialogue allowed the student to ask questions. SCHOLAR parsed the student's questions, examined its semantic net, and generated an appropriate answer. For example, the question "Is Buenos

Aires in South America?" would receive a reply of "yes" because Buenos Aires is in Argentina and Argentina is in South America. A subsequent version of SCHOLAR did record correct student answers. The only use of this primitive student model information was to avoid redundancies in question asking (Wenger, 1987; p38).

After Carbonell's death, Stevens et al. (1982) worked on a system called WHY, which focused on the following three questions:

1. How can a good tutor's use of questions, statements and examples be characterized? What is the goal structure of a Socratic tutor?
2. What types of misconceptions do students have? How do tutors diagnose these misconceptions from the errors students make?
3. What are the abstractions and viewpoints that tutors use to explain physical processes?

The research moved from SCHOLAR's factual domain (geography) to WHY's causal domain (meteorology). This change of domain and the emphasis on understanding student misconceptions led to the notion of a mental model, an internal representation used by a student or a tutor to reason about a system. Stevens et al. (1982) classified the following categories of misconceptions along with directions to correct them:

1. Factual Bugs. The tutor deals with these by correcting the student. Teaching facts is not the goal of Socratic tutoring, interrelation of facts is more important.
2. Outside-domain bugs. These are misconceptions about causal structure that the tutor chooses not to explain in detail.
3. Overgeneralization. This occurs when a student makes a general rule from an insufficient set of factors (i.e., any place with mountains has heavy rainfall). The tutor will find counterexamples to cause the student to probe for more factors.
4. Overdifferentation. This is when a student counts factors as necessary when they are not. The tutor will generate counterexamples to show that they are not.
5. Reasoning bugs. Tutors will attempt to teach students skills such as forming and testing hypotheses and collecting enough information before drawing a conclusion.

WHY remained primarily a theoretical system (Wenger, 1987); it was only used in experimental settings. It did not address how to follow the student's dialogue and diagnose bugs (Clancey, 1986b). It did uncover important issues in student modelling.

2.2 Bug and Overlay Paradigms

WUSOR is a coach for a computer game called WUMPUS (Yob, 1975). The first version, WUSOR-I (Stansfield et al., 1976), did not employ a student model. Its second version, WUSOR-II (Carr & Goldstein, 1977) added a student model using the overlay paradigm. The essence of this paradigm is that the student's knowledge is considered to be a subset of the expert's knowledge. The student's knowledge is compared with the expert's knowledge to determine what knowledge is missing.

The intelligent tutoring system WEST (Burton and Brown, 1976) is a coach; it makes suggestions to students playing a game that requires mathematical skills. These suggestions come from a student model that uses differential modelling. The student's move is compared with a list of the expert module's moves and student problems are recognized. WEST's domain and objectives were very limited but its use of a student model is very significant. It was used by elementary school students in a control group and data was collected that suggested that a group coached by WEST did better than the uncoached group (Burton & Brown, 1982).

There are two problems with using an overlay model only. First, it assumes that the expert module is complete. However, it is possible for the student to employ a legitimate strategy that is not in the expert module. Second, overlay models do not address the situation where the student misuses or misunderstands information. It assumes that information is only present or missing (Wenger, 1987).

BUGGY (Brown and Burton, 1978) tutored students on subtraction. The system is based on a hierarchical network of skills related to subtraction. Included in this hierarchy are potential misuses of these skills. If a student has an incorrect answer, the system tries to substitute variant skills with correct skills in order to reproduce the student's answer.

The overlay and bug paradigms for student modelling are the most prevalent paradigms for student modelling. The distinction between the two implies that there are two classes of knowledge, declarative and procedural (VanLehn, 1988; Hume, 1992). Declarative knowledge consists of facts of which, typically, a student has a subset of the expert's knowledge. Procedural knowledge consists of procedures or problem solving skills. This distinction can often be useful and practical, but neither paradigm is sufficient to model a student working in a complex domain. Some work has been done in combining the two paradigms (as is done in GUIDON; Clancey, 1987), but the resulting model still separates knowledge that is naturally integrated. For example, a procedure can be represented by a declarative sequence (Hume & Evens, 1992).

2.3 <u>Other Paradigms</u>

I have looked for research on modelling meta-level skills, global and local skills and communication skills. Few ITS's, however, have attempted to model anything about the student other than aspects of the student's progress relating to the expert's domain knowledge. Sherlock II (Lesgold et al., 1993a; Katz et al., 1993) does introduce the idea of global and local variables that represent skill levels. Again, these skills are compared to the skills of an expert. I was hoping to find literature that suggests other aspects of student behavior can be modeled. Specifically, can the following be modeled?:

1. Student frustration.
2. Ability to articulate in the jargon of the domain.
3. Ability to understand the tutor's questions.

2.4 Diagnosis

Student modelling research has focused on diagnosing and modelling the student's knowledge primarily through the overlay or bug paradigms. Little research has been done on (1) measuring the effectiveness of student models (Mark & Greer, 1991) or (2) recording other aspects of student behavior in a student model (VanLehn, 1988).

Is there some other student behavior that can aid in diagnosis? Person et al. (1993) concluded from a study of human tutoring sessions that answers to specific domain questions are the only reliable information for modelling. Their analysis of student questions and student responses to comprehension-gauging questions ("Do you understand?") found that this information was not useful; student answers did not help in measuring student understanding. Only the very best and very weakest students can reliably respond to comprehension-gauging questions. Graesser (1993) also suggests that superficial conversational behavior does not provide useful modelling information.

There is research that suggests there are other areas of student behavior to explore. Chi et al. (1989) concluded that students who explain instructional material to themselves learn more. In a follow up study, VanLehn and Jones (1993) determined that increased learning was due to the students' self discovery of gaps in their knowledge. If prompting the student for explanations helps them to discover gaps in their knowledge, it may also provide information for student modelling.

Fox (1993a), in a study of tutoring dialogue, has observed student behavior that may be useful in diagnosis. Tutors do not immediately interrupt a student when there is a mistake. The tutor provides an opportunity for self correction. This is similar to Chi's self explanation and VanLehn's self discovery of gaps. A student's utterance, during a delay or after a mistake, may provide valuable diagnostic information to the tutor. The student may (1) recognize that there is a problem, (2) know what the problem is, or (3) not know there is a problem. Also, when assistance is provided, it is in the form of a question that provides a resource to the student (Fox, 1993a). This is a partial description of a hint (see Chapter V). The response to such an indirect utterance can also provide diagnostic information about meta-level skills.

Lepper et al. (1992) argue that motivational factors are as important as cognitive factors in tutoring. They suggest that computer tutors be empathetic. One example of this would be for the computer to help the student find his own error. Another example would be to give encouragement when students are frustrated and challenge students when they are ready to address more advanced issues.

Sherlock II (Lesgold et al., 1993a; Katz et al., 1993) provides a student with hints when there is an impasse.

These hints act as reminders; they provide the student with information that is already available. Lesgold et al. argue that the response of the students to hints can be used in the diagnosis of a student. Sherlock II may provide a sequence of hints; each one providing more information than the previous one. Once a hint helps the student, an inference can be made about the underlying student difficulty. For example, if a hint restates the problem and the student asks for another hint then it can be inferred that the student understood the problem. If a student can only proceed after a hint about interpreting test results, then the student's difficulty may be with the interpretation of test results.

2.5 Hints

I have found no thorough and systematic study of hinting. There are *mentions* of hinting in the linguistic, educational, cognitive science, ITS, and tutoring literature. However, in every case, there is no explicit definition or detailed analysis of the activity, and one finds few concrete examples of hints actually generated by speakers.

In the literature that exists on hinting, there are assumptions made by a tutor (or teacher) and the student. For example, the student must interpret the relation between the surface form of an utterance and the intention of the tutor (Florio-Ruane, 1987). The tutor may have more than one purpose for giving a hint. For example, a utterance may be a hint that provides information and supplies an acknowledgment (Evens et al., 1993; Kamsteeg & Bierman, 1985).

While it is always assumed that hints are used to help a student, recent literature suggests that hints may also provide valuable feedback to the tutor. Reiser (1989) observed that human tutors "moderate their control of the interaction to provide sufficient assistance for the student to solve the problem." He also observed that hints, and the feedback from hints, are the tools human tutors use for this monitoring.

Lesgold et al. (1993a) are experimenting with the concept of using hints to monitor the student's progress using Sherlock II. When a student reaches an impasse, the computer will begin to provide hints. Once a hint appears to help the student make progress on a problem, the computer tutor can make an inference: the cause of the impasse was the student's difficulty with the information provided in the hint.

GIL, a LISP programming ITS (Reiser et al., 1988; 1992), provides hints to students when (1) the student reaches an impasse and (2) when the student requests help. The following is a GIL assignment that a student is incorrectly solving:

> Write a program to take a list as an argument and construct a new list with the last element rotated to the front of the list. For example, (a b c d) would become (d a b c)

The tutor, without a student request, provides the following hint:

> Working backwards using APPEND is a good idea, but (d a) and (b c) may not be the best input to use. Breaking (d a b c) into (d a) and (b c) makes the problem harder than it needs to be.

After a period of time the student requests a hint. GIL responds with:

> Try using (d) and (a b c) as the input for APPEND. This input will be easier to get from (a b c d). (a b c) contains the first few elements of (a b c d).

The protocol of GIL allows the student to request a hint. GIL's hints are attached to problem solving rules so GIL must decide what hint is appropriate when a request is made. Though not explicitly stated, GIL's hints are error driven and designed to allow the student to self discover solutions. As with other references to hinting in the literature, no formal definition of hinting in GIL is provided.

Fox (1993a,b) observes that tutors often reply to students with indirect responses, but does not call this hinting. There appears to be an assumption that a hint provides some information, but not all of the information the recipient needs. Fox's studies involve graduate students tutoring undergraduates in solving physics problems; the published tutorial dialogues show only a few examples of hinting and almost no examples of the kind of complex hinting done by JAM and AAR. However, she does comment that "the tutor and student both make use of strategies which maximize the student's opportunities to correct his or her own mistakes" (Fox, 1993a, p. 122). Certainly this is the intent of JAM and AAR in their use of hints. Fox presents one example of a tutor responding to a student with "a question whose answer will serve as a resource for getting the student unstuck" (p. 124). This is a type of hint generated by JAM and AAR.

Reiser et al. (1988) point out that GIL's hints are more explicit and directive than hints provided by human tutors. McArthur et al. (1990) carried out a detailed analysis of human tutoring of algebra problem solving; many of the tutoring techniques they describe (and the examples of dialogue they provide) look like hints, although the authors do not focus any attention on this phenomenon. Littman et al. (1990) studied experienced tutors (graduate and advanced undergraduate students) doing simulated tutoring of computer programming, and again, little or no hinting is evident in their results. Graesser (1993) has studied graduate students tutoring undergraduates about psychology research methods and he reports that his tutors generate "virtually no occurrences of sophisticated tutoring strategies, such as the Socratic method" (p. 127). He then mentions the occurrence of hinting but shows no examples of this behavior. An examination of two of Graesser's transcripts (provided by him) reveal few hints, most of them simple in construction.

Developers of intelligent tutoring systems (ITSs) have in the past talked about hints and hinting, usually without precisely defining what they mean by this pedagogic tactic. Burton and Brown (1982) describe four levels of hints that their computer tutor WEST will offer to students when help is requested. While they offer a rule for each level of student difficulty, they present only one example of a level 1 hint, and it is unclear what other hints might look like and how they would function. Ohlsson (1987), in describing "principles of intelligent tutoring," discusses six tactics and a number of variations of each; many of these sound like the hints that I have identified (see Chapter V). However, Ohlsson does not explicitly discuss hinting and offers no examples of tutor-student dialogue in which hinting might be observed. Reiser's computer Lisp tutor GIL (Reiser et al., 1989; Merrill et al., 1992) generates hints in response to a student's request for help, or in response to student errors. Here too, however, there is no explicit description of what a hint is, when it is generated (except in response to a request for help), or how it is constructed. Woolf (1992) discusses "tools for representing, acquiring, and reasoning about tutoring knowledge" (p. 48) and includes hints in her "response matrix" (p. 56) for reasoning about tutoring discourse. She too has avoided any concrete consideration of how and when to generate hints.

Why is there so little substantive discussion of hinting in the literature? One possible answer is that other investigators have simply not looked seriously for hints, or at least have not directed much attention to this phenomenon. The occasional reference to "hints" or "hinting," without elaboration, would support this hypothesis. Hinting may be a more common tactic than is evident from published studies of tutoring.

An alternative hypothesis is that there are a great number of factors that account for the differences in the prevalence and sophistication of hinting in the studies available for comparison. Specifically, (1) the domain expertise of the tutors may vary, (2) the tutoring expertise of the tutors may be different, (3) the particular domain in which tutoring is occurring has varied, (4) the nature of the problems being solved by the students is different, and (5) the context in which the tutoring takes place varies.

CHAPTER III

THE CST PROJECT

Allen Rovick (AAR) and Joel Michael (JAM), physiology professors at Rush Medical College, wrote a CAI program called CIRCSIM (Rovick & Michael, 1986). The objective of this program is to assist the student in problem solving and making predictions about the human body's blood pressure regulating system. This program is currently used by medical students at Rush and other institutions. It has been proven to be an effective learning resource (Rovick & Michael, 1992).

The development of CST began in 1989. JAM and AAR started a cooperative project with Martha Evens of Illinois Institute of Technology and several of her graduate students. JAM and AAR are the domain and human tutoring experts of the CST project. The goal of CST was and is to use natural language to tutor problem solving in the same domain as CIRCSIM. The need for a robust student model, more natural text generation, and further enhancements in other modules was the impetus for developing CST (v.3).

CST's primary goal is to help medical students understand the reflex regulation of blood pressure. The human body attempts to maintain a more or less constant blood pressure through its own negative feedback system, a self regulating system. Corrective responses occur when a disturbance is detected. As an example of a negative feedback system, consider the heating system in a house. The temperature of the house (or at least the temperature at the thermostat) is the regulated variable and the heat output of the furnace is an effector variable that will act to determine the temperature in the house.

In the human cardiovascular system there are many variables to be considered, and the number of interactions between these variables is large. Students often find some of these interactions counterintuitive. It is, therefore, difficult for students to learn to reason about the behavior of this system.

When blood pressure is altered, the baroreceptor reflex initiates responses by way of the central nervous system. For example, during a hemorrhage (a loss of blood) a very noticeable response of the body is an increase in the heart rate (HR; see page iv for full variable names and acronyms). HR, inotropic state (IS) and total peripheral resistance (TPR), are the effector variables in the cardiovascular system; they are the variables that are altered (under neural control) to reverse the effect of the disturbance on the regulated variable, mean arterial pressure (MAP).

CST is designed to be used by first year students in medical school. Because relationships between physiological variables in this system are often complex and counterintuitive, virtually all these students need some assistance beyond the lectures and readings. Even when students have learned the facts, they need help to apply them in problem solving. Therefore, CST's domain provides an ideal environment for the study of tutoring and student modelling.

3.1 Methodology

An indispensable tool in the design of CST has been the analysis of human tutoring sessions. These sessions are referred to as keyboard to keyboard sessions. Fifty-eight human tutoring sessions, conducted with tutor and student using PCs in different rooms, have been recorded using a computer program called CDS (Li et al., 1992). Students were presented with problems of the type used in CST and were asked to solve them in a certain way (see Section 3.3). The resulting dialogue has been used to analyze a range of topics including planning (Woo et al., 1991), student initiatives (Sanders et al., 1992), knowledge base issues (Khuwaja et al., 1992), sublanguage issues (Chang et al., 1992; Seu et al., 1991a,b), negative acknowledgments (Evens et al., 1993; Spitkovsky & Evens, 1993) and hints (Hume et al., 1993, 1995a,b).

The subjects for the tutoring experiments were first year medical students at Rush Medical College. They ranged in age from 21 to 37 years with a mean age of 25 years; 32 were female and 26 were male. JAM and AAR believe that all subjects, while paid volunteers, had the expectation that they would learn something from their participation in the experiment. The transcripts of sessions guided the development of CST (v.2).

The examples of tutoring dialogue (all enclosed in boxes) in this thesis have been edited slightly for readability. Distracting typographical errors have been corrected and acronyms for physiological variables have been capitalized. Tutors and students interchangeably use the tokens "+," "i" and "increase" when describing qualitative changes in the CV system. Likewise, the tokens "-", "d" and "decrease" are used interchangeably; the tokens "0", "same" and "unchanged" are used interchangeably.

3.2 The CST Domain

JAM and AAR have developed a representation of the cardiovascular system that is composed of physiological variables connected by qualitative causal linkages. For example, when heart rate increases (and certain other variables remain unchanged) then cardiac output (CO) must increase. This representation is called a concept map. Figure 1 shows the most elementary system representation, the surface level concept map. An arrow from one box to another represents a qualitative causal relationship. A plus sign means the causal relationship is a direct causal relationship (when HR increases, CO increases). A minus sign means the causal relationship is an inverse causal relationship (when CO increases, CVP decreases).

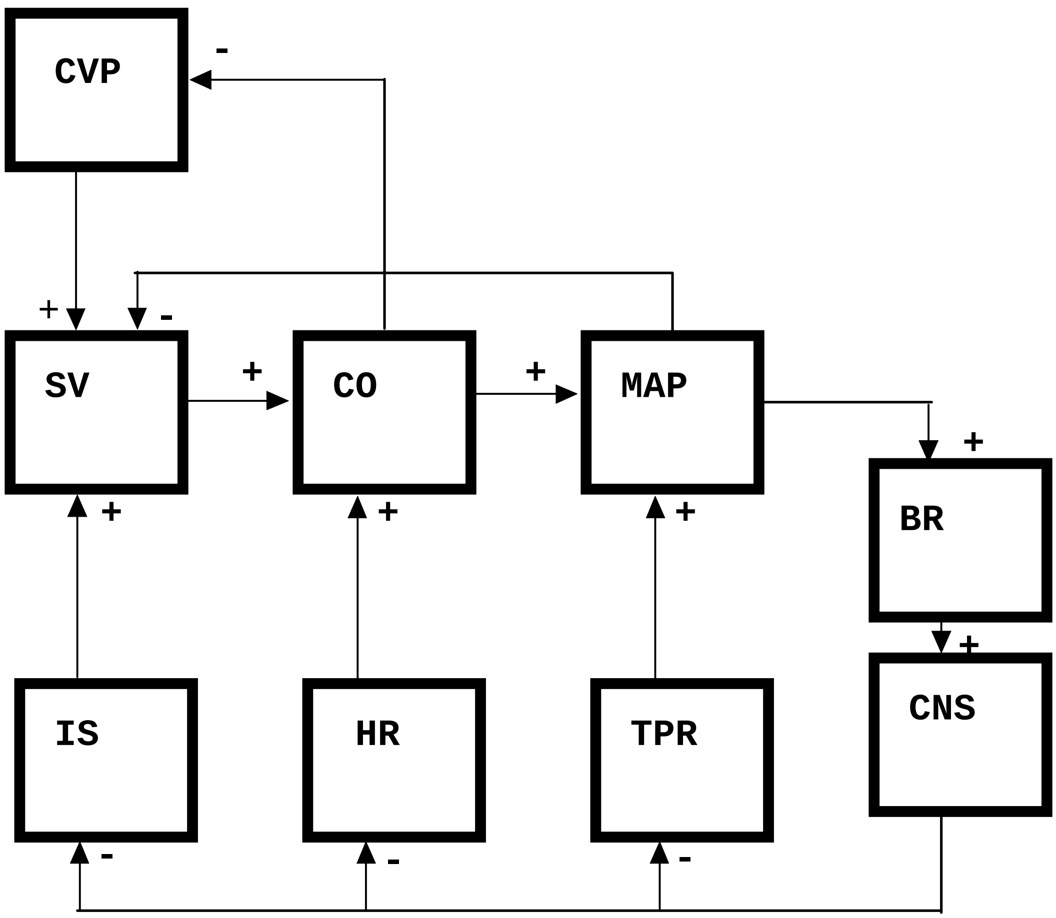

Figure 1. Surface Level Concept Map

One goal of CST is to help the student reason about the causal relationships represented in the surface level of the concept map. There are, however, many causal relationships between physiological variables not represented in this surface level. A study of CST's human tutoring transcripts revealed which relationships JAM and AAR used to explain the surface level relationships (Khuwaja, 1994). These relationships are represented in the intermediate and deep levels of the concept map (see Appendix A). At a deeper level there may be many causal effects between a pair of causally related variables in the surface level concept map. Figure 2 shows the intermediate level expansion of the CO to MAP relationship. JAM and AAR often use these deeper level relationships to (1) explain the surface level relationship and to (2) provide a hint so that the student may be able reason about the surface level relationship.

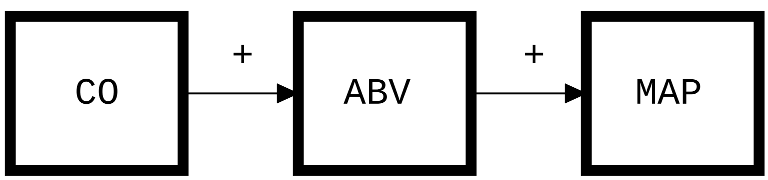

Figure 2. Portion of Intermediate Level Concept Map

Dialogue 1 is an example of the tutor using the intermediate level causal relationships represented in Figure 2 to make an explanation. The tutor is explaining the effects of an increase in CO. On of those effects is it causes arterial blood volume (ABV; intermediate layer) to increase which causes pressure (MAP; surface layer) to increase.

Dialogue 1 from k10-tu-57:

> T: When CO increases it transfers increased quantities of blood from the venous system into the arterial system, decreasing the CBV (central blood volume) and increasing the arterial blood volume (and pressure).

Two design changes that have been made are worth noting so that excerpts of transcripts make sense with regard to terms. The cardiovascular variable Right Atrial Pressure (RAP) was removed from the surface layer of the concept map so that its highest level is the intermediate layer. Central Venous Pressure (CVP) was added to the surface layer of the concept map to replace RAP. Students often confused RAP with MAP. JAM and AAR decided that CVP would cause less confusion involving terminology; there was no significant difference regarding the reasoning about casual effects on the CV system. For similar reasons the term Cardiac Contractility (CC) was replaced with Inotropic State (IS).

3.3 Keyboard to Keyboard Tutoring Experiments

Recent keyboard to keyboard tutoring experiments commenced with a description of a perturbation. A perturbation is a disturbance. For example, the human body may experience a hemorrhage, a blood loss. All perturbations cause a direct response (DR) in the human body that leads to a change in mean arterial pressure (MAP), the regulated variable. In the case of a hemorrhage, MAP will decrease in the DR. Then, the central nervous system responds to the change in MAP. This is the reflex response (RR) phase and, in the case of a hemorrhage, the reflex will increase MAP. The eventual result of the reflex compensation for the perturbation is a new steady state (SS). In the hemorrhage example, MAP in the SS will be lower than before the loss of blood.

Phase: CV Variable	DR	RR	SS
Inotropic State (IS)			
Central Venous Pressure (CVP)			
Stroke Volume (SV)			
Heart Rate (HR)			
Cardiac Output (CO)			
Total Peripheral Resistance (TPR)			
Mean Arterial Pressure (MAP)			

Figure 3. The Predictions Table

A student is asked to identify the primary variable, the first Surface Level Concept Map variable be affected in DR. If the student has problems making this initial prediction, the tutor will engage in an interactive dialogue to help the student. If the student can not make the correct prediction after one or two hints then CST will provide the correct answer. After the primary variable has been identified, the student predicts the qualitative change of the primary variable; it either increases or decreases.

The next step is to ask the student to make qualitative predictions (+ for increase, - for decrease, 0 for no change) in the appropriate boxes of the Predictions Table (PT; shown in Figure 3) for the remaining variables in the DR phase of the response. The Predictions Table contains seven cardiovascular variables from the Surface Level Concept Map (Rovick & Michael, 1992). The rules for making predictions on the PT for each of the three phases (DR, RR, and SS) should be considered as rules of the game. This is because JAM and AAR have imposed these discrete phases on an essentially continuous process. The point of time when the reflex response and the steady state commence is not exact.

There is no representation of a predictions table in CDS; the student fills out a column of predictions on paper. When the student finishes making all the predictions in the DR column of the PT, the tutor prompts the student to type predictions one at a time. Then, the tutor engages the student in an interactive, remedial dialogue. When tutoring about the DR phase is complete, the student is prompted to fill the RR column on paper. Again, the student types those predictions and the tutor initiates an interactive dialogue. Finally, the same protocol is followed for the SS phase. Following this protocol, one perturbation takes approximately one hour.

This prediction activity, in which the student makes seven uninterrupted predictions (six for DR), gives CST an opportunity to isolate student problems expeditiously and record them in the student model (Michael et. al., 1992). In the earliest set of keyboard to keyboard tutoring experiments JAM and AAR regularly responded to students immediately after each incorrect prediction. They have decided that they can tutor more effectively if they wait to see patterns of errors (see Section 4.1). The exception is the primary variable; it is predicted separately in the DR phase. They have determined that it is essential for students to start the DR phase correctly. They can not learn anything from the patterns of errors if the student is, in essence, solving the wrong problem. The CST protocol has evolved as JAM and AAR gained experience with this tutoring task (Khuwaja, 1994).

The PT structures the student's thought process because it corresponds to the phases of the baroreceptor reflex. Also, the order of variables in the PT helps organize the student's thought processes. IS and CVP (PT entries 1 and 2 in Figure 3) are determinants of SV (entry

3). SV and HR (entries 3 and 4) are determinants of CO (entry 5). CO and TPR (entries 5 and 6) are determinants of MAP (entry 7). In a sense, the PT helps the student break a larger problem into several smaller problems. It is unreasonable to assume that a student can consistently make seven correct predictions per phase without an understanding of the physiological concepts.

3.4 CST

Analysis of keyboard to keyboard tutoring experiments and CST (v.2) tutoring sessions influenced the development of a protocol that has been used in recent keyboard to keyboard experiments. The CST (v.3) uses this protocol. The various cognitive activities of the tutors are separated into modules. The student model is updated upon completion of each phase in the PT. The instructional planner (see 3.3.2), after examining the student model, determines the tutor's goals, strategies and tactics. A natural language dialogue commences. The student modeller monitors this dialogue and updates the student model constantly to provide the instructional planner with information.

The primary modules making up CST (v.3) are the domain knowledge representation system (DKRS), the student modeller, the instructional planner, the text generator, the input understander, the judger, and the screen manager.

There are several data stores: the student model, the lexicon, the tutoring history and the discourse history. The data stores can provide information to any module. Therefore, information used by several modules needs to be stored in only one place. Descriptions of the DKRS and the primary procedural modules follow; the student modeller is fully discussed in Chapter IV. Figure 4 shows the high level components of CST (v.3).

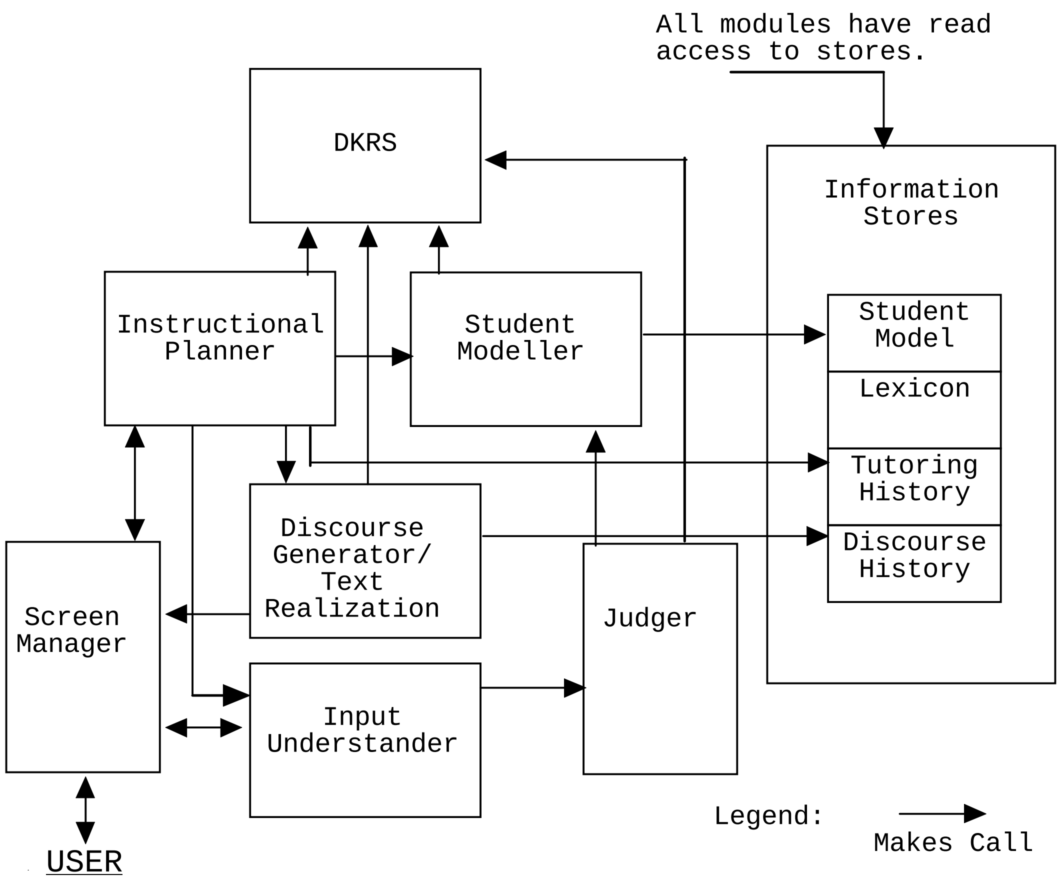

Figure 4. CST's Modules

3.4.1 <u>The Domain Knowledge Representation System</u>. The domain knowledge representation system (DKRS) contains a formal representation of cardiovascular knowledge. It also executes procedures that represent CST perturbations. In other words, it solves the CST problems or makes the correct predictions. The knowledge in the DKRS is essential to the student modeller. In fact, many of the relationships represented in the student modeller are based on relationships represented in the DKRS (see Chapter IV).

There are two major components to the DKRS: the knowledge base and the problem solver. The knowledge base contains a collection of physiology facts including the three levels of the concept map (see Section 3.2). The problem solver contains the algorithms used to solve CST problems and answer questions. Any modules that needs information may access the DKRS.

3.4.2 <u>The Discourse Generator and Surface Realization Module</u>. The discourse generator receives messages from the instructional planner (see Section 3.4.6). The intent of these messages is to cause it to plan and generate text for display on the screen. Sometimes the intended text is a short utterance, sometimes a long explanation. Sometimes the message signals the start of a multiturn dialogue. The discourse generator puts together a discourse plan and sends that plan to the surface realization component of the module. The surface realizer generates utterances and sends them to the screen manager (see Section 3.4.5). The discourse generator may consult the student model (see Chapter IV) to help formulate its plan.

3.4.3 The Input Understander. The input understander parses messages typed by the student and produces a logic form A logic form is an unambiguous representation of a concept. For example, the utterances "HR went up," "HR is increased" and "increase in HR" are represented by the same logic form.

The input understander also tries to determine the intention of the student. If the student was trying to answer a question, the logic form is sent to the judger (see Section 3.4.4). If the student is trying to ask a question (a student initiative), the logic form is sent to the instructional planner (see Section 3.4.6).

3.4.4 The Judger. The primary purposes of the judger are (1) to decide the correctness of a student response, (2) make that information available to the instructional planner, and (3) to make calls to the student modeller. The judger receives logic forms, representing student responses, from the input understander. A student response can be correct, incorrect, partially correct, or not an answer to the current question (e.g., a student initiative). If the student response is an answer, the judger determines what portion of the answer is correct, what portion is incorrect and what portion of the correct answer is missing. This is recorded in a system variable called LAST-REPLY. The instructional planner has access to this variable. Then, the judger sends a message to the student modeller to update the local assessment (see Section 4.9); it may send a message to the student modeller that will update the overlay model (see Section 4.3).

3.4.5 <u>The Screen Manager</u>. The screen manager is CST's interface to the student. It displays messages sent to the screen from the discourse and text realization module and it captures student inputs and sends them to the input understander. It controls where text and graphics are displayed on the screen. It also manages the entries into the predictions table and sends them to the student modeller.

3.4.6 <u>The Instructional Planner</u>. The instructional planner is, in a sense, the control center of CST. It makes tutoring decisions by examining its own rules and by consulting the student model. It selects domain topics, determines tutoring strategies, and initiates dialogue by sending messages to the text generator.

The role of the instructional planner is illustrated in the following CST scenario:

1. The instructional planner sends a message to the screen manager to collect predictions for one phase of the PT.

2. The instructional planner sends the predictions received from the screen manager to the student modeller so that the student model may be updated.

3. The instructional planner determines a tutoring strategy examining its own rules and by consulting the student model. For example, it may focus on an error pattern (see Section 4.1) and decide to use an interactive dialogue.

4. The instructional planner selects a tactic, perhaps a hint (see Chapter V). It instructs the discourse generator and surface realization module to form the appropriate utterance and have it sent to the screen manager.

5. The instructional planner sends a message to the input understander to parse the student response received from the screen manager and return the resulting logic form.

6. The instructional planner sends the logic form to the judger so that it may (1) determine the correctness of the response and (2) call the student modeller so that it can update the student model.

7. The instructional planner reevaluates its strategy. In essence, the instructional planner goes back to

step 3 unless it decides it is time to terminate tutoring the current phase.

3.5 CST's Use of Object Oriented Programming

CST uses Allegro Common LISP in the PC version, Procyon Common LISP in the Macintosh version. Both the DKRS and the student modeller use an extension of Common LISP called CLOS (Common LISP Object System). CLOS supports object oriented programming. Some advantages of object oriented programming are (Keene, 1989):

1. It provides a natural way to design the computer program's objects so that they have the same properties as the real world objects they model.
2. It provides modularity.
3. It provides simple interfaces so that separate modules can easily communicate.
4. It is extensible; it creates a higher level language for the programmer.
5. It is easy to make enhancements and modifications.

In this paradigm, objects are hierarchically organized. The simplest form of this organization is a tree. An object may, however, have more than one parent. An object inherits properties from its parent object, called a class. Classes are also objects. An object that is a leaf node is called an instance. Instances represent actual data. For example, there are instances for every CV variable, every causal relationship, and every perturbation.

There are two types of properties an object may have: data and procedure. A property that is data resides in a slot. A procedural property is called a method. A method is a small program. When one object wants to communicate with another object, it sends that object a message. When a message is sent, the receiving object searches its slots and methods in order to reply. If it does not find the appropriate slot or message, it forwards the message to its parent. The parent object follows the same algorithm. When an appropriate slot or method is found, a reply is sent to the original sender. If a slot is found, the enclosed data is sent. If a method is found, it is executed. Generally, the original sender does not need to know if the message sent retrieves data from a slot or executes a message. It never knows if the slot or method was found in the object or in an ancestor object.

Figure 5 displays a small portion of the DKRS's object hierarchy. The 'CO to CVP inverse causal relationship' object is one of the instances of the 'inverse causal relationship object' (a class of objects). Two of its slots are called antecedent and consequence. Their slot values are the CV variables CO and CVP. A method, called 'propagate causal effect' is attached to the 'inverse causal relation' object. The module that is responsible for solving CST's problems determines the values for the CV variables one at a time. Once the qualitative change of CO is determined, this module sends a message, 'propagate causal effect,' to the 'CO to CVP inverse causal relationship' object. A method that matches the message is found in its parent object. The method is executed and the qualitative change of the CV variable in the consequence slot (CVP) is determined.

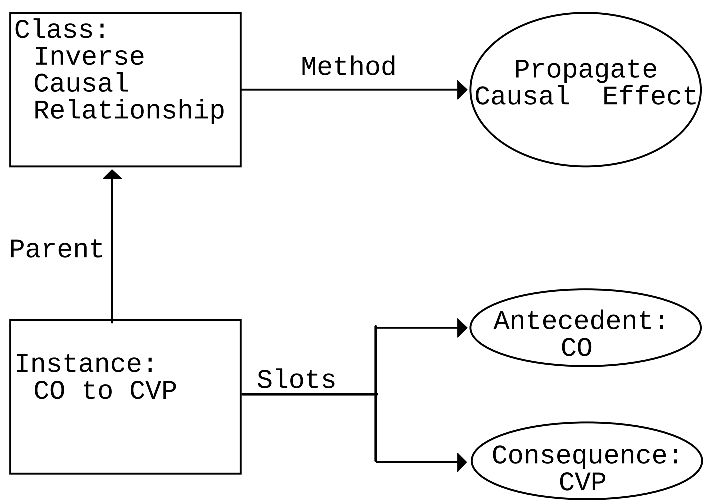

Figure 5. An Example of CST Objects

CHAPTER IV

THE DESIGN OF STUDENT MODELLING IN CST (v.3)

The primary purpose of a student model is to provide information to the instructional planner about the student. Therefore, the design of student modelling in CST (v.3) has proceeded concurrently with the design of the instructional planner. CST's student model represents both (1) the student's understanding of domain material and (2) variables that reflect other aspects of the student's progress. Examples of other aspects include responses to the various tutoring tactics employed, use of hedges (when the student shows a lack of confidence in an answer), and the time elapsed between inputs. The design of the student model is based upon justifications for tutoring decisions offered in interviews by JAM and AAR while reviewing tutoring transcripts.

The student modeller provides services to (1) the instructional planner to help create the tutoring plan and (2) the discourse generator to determine the nature of CST's contributions to the dialogue. The discourse generator will pose questions and acknowledgments differently to students according to past performance. Aside from providing services to these two modules, the student model may be used for other purposes. These include (1) providing statistics for future research, (2) validating planning and student modelling rules, (3) storing the student model on disk for future analysis and use in a subsequent tutoring session, and (4) initializing the student model from a disk. Number 4 may be useful if a student returns for a subsequent tutoring session.

4.1 <u>Model of Error Patterns</u>

Interviews with JAM and AAR, and studies of their tutoring transcripts, reveal that they plan the interactive dialogue by focusing on the student's most severe observable problems. A severe problem is one that, if not tutored, will limit the student ability to make correct predictions in the remainder of the session (and presumably in the future). The first step in identifying these problems is to examine the predictions made and compare them with the correct predictions. This produces a list of surface level errors. Table 1 is an example of possible student predictions and the actual changes produced by a perturbation. In this case, the surface errors are the predictions for variables SV, CO, and MAP. All of the other student predictions are correct.

These surface errors need further analysis to determine the problems that gave rise to them. One error, followed by flawless causal reasoning, often creates many erroneous predictions. For example, consider the qualitative changes, in correct causal order, from Table 1:

CVP +, SV +, CO + , and MAP +. In this case, CVP is the primary variable (the first PT variable that changes due to the perturbation) and it increases in value. In this example, the student incorrectly predicts SV and applies correct reasoning to the subsequent relationships. One error in causal reasoning produced three surface errors. Conversely, simply guessing is likely to produce a number of correct responses.

Table 1. An Example of Student Predictions for the DR Phase

CV Variable	Student Prediction	Actual Change	Errors
IS	0	0	
CVP	+	+	
SV	-	+	X
HR	0	0	
CO	-	+	X
TPR	0	0	
MAP	-	+	X

JAM and AAR use the surface errors to identify error patterns. Each error pattern (EP) corresponds to a physiological concept. While these concepts vary in importance and complexity, the student must understand all concepts to consistently make correct predictions. There are several categories of error patterns. Some error patterns arise from incorrect reasoning about qualitative causal relationships. For example, there is a direct causal relationship between SV and CO; CO is directly proportional to SV (refer to Figure 1). If SV increases then CO increases and if SV decreases then CO decreases. When a student fails to make predictions for SV and CO that exhibit this causal relationship, an error pattern is sensitized. To sensitize an error pattern is to make that error pattern visible to the instructional planner and the other CST modules. The student modeller sensitizes error patterns after each column (DR, RR or SS) is filled with predictions and later updates the status of these error patterns during interactive tutoring.

Some error patterns can arise when making predictions about inverse causal relationships. For example, CVP is inversely proportional to CO. If CO increases then CVP decreases and if CO decreases then CVP increases. Other error patterns relate to the behavior of neural variables. The neural variables are those controlled by the nervous system; they are the effector variables in this negative feedback system. There are a finite number of error patterns (see Section 4.2). The error patterns that would be sensitized from the predictions in Table 1 are the error patterns that correspond to the following relationships:

1. The direct causal relationship from CVP to SV,

2. The direct causal relationship from SV to CO, and
3. The direct causal relationship from CO to MAP.

An error pattern represents a physiological concept that the student about which the student has appeared to reason incorrectly. The error pattern number 1 (above) means a qualitative change in CVP causes the same qualitative change in SV. The error pattern number 2 (above) means a qualitative change in SV causes the same qualitative change in CO. The error pattern number 3 (above) means a qualitative change in CO causes the same qualitative change in MAP.

4.2 <u>CST's Error Patterns</u>

The following list contains representations of physiological concepts. The tutor can determine, by simple observation, if student predictions violate any of these concepts. Therefore, an error pattern is attached to each of these concepts.

Direct Causal Relationships:

- CVP to SV.
- IS to SV.

- SV to CO.
- HR to CO.
- TPR to MAP.
- CO to MAP.

Inverse Causal Relationships:

- CO to CVP.
- MAP to SV.
- MAP in DR phase to TPR in RR phase.
- MAP in DR phase to HR in RR phase.
- MAP in DR phase to IS in RR phase.

Multiple Variable Direct Causal Relationships:

- MAP = TPR x CO.
- CO = HR x SV.

Others:

- Neural variables remain unchanged in DR phase (except for primary neural variables).
- Neural variables in RR make the opposite qualitative change from MAP in DR (except for clamped neural variables which remain unchanged).
- All neural variable change in RR phase (except clamped neural variables which remain unchanged).

- All variables change in RR phase (except clamped neural variables which remain unchanged).
- MAP in RR makes the opposite qualitative change from MAP in DR
- MAP in SS makes the same qualitative change as MAP in DR.
- The SS qualitative change for a variable is the algebraic sum of the qualitative changes for that variable in DR and RR.
- The identity of the primary variable for each perturbation (the first predictions table variable affected by the perturbation).
- The qualitative change of the primary variable is determined by the perturbation.
- All clamped variables remain unchanged in RR.

4.3 Overlay Model

The DKRS has an explicit representation for every causal relationship represented in the surface (or top), intermediate and deep levels of the concept map. The two types of causal relationships are:

1. Direct Causal Relationships. Example: An increase in CO causes MAP to increase.

2. Inverse Causal Relationships. Example: An increase in CO causes CVP to decrease.

CST's representation of inverse and direct causal relationships includes the following slots and potential values:

1. Level: a list containing either (top intermediate deep), (top intermediate) or (top). A variable in the deep level of the concept map is in all three levels.
2. Direction: either direct or inverse.
3. Antecedent: a physiological variable. Example, CO for the relationship CO directly affects MAP.
4. Consequence: a physiological variable. Example, MAP for the relationship CO directly affects MAP.
5. Medium: either Physical-Chemical or Neural.

CST requires students to make predictions about variables at the surface level. Therefore, when errors are made, violations of surface level causal relationships may be observed. This is when error patterns are sensitized. The representation of every CST surface level causal relationship is linked to the representation of its corresponding error pattern. The members of each pair, the corresponding causal relationship and error pattern, while residing in different modules, are linked. Therefore, the student modeller has access to the causal relationships in the DKRS.

The following are slots attached to each error pattern:

1. Prediction history.
2. Tactical history

The prediction history value is updated after a column in the predictions table is completed. Its value is a list that contains a string of Cs (for correct) and Ws (for wrong). An algorithm can be applied to this string to produce a confidence factor (see Section 4.6) The value of the tactical history slot is a string of Hs (for hint) and Es (for explanation). This is a history of how a sensitized error pattern was resolved (see Chapter V).

During the tutorial dialogue, the student has the opportunity to demonstrate his/her understanding of causal relationships at all three levels. Attached to DKRS's causal relationships are the following slots for the student model's overlay model:

1. Dialogue History.
2. History of knowledge of antecedent.
3. History of knowledge of consequence.
4. History of knowledge of actual antecedent.
5. History of knowledge of actual consequence.

These slots are attached to every CST causal relationship and are updated by the judger during the tutoring dialogue. The value of each of these slots is a list that contains a string of Cs (for correct) and Ws (for wrong). An algorithm can be applied to this string to produce a confidence factor (see Section 4.6).

The following is the list of phase overlay model slots. These are attached to the DR phase, RR phase and SS phase:

1. History of correct predictions.
2. History of incorrect predictions.
3. History of correct error patterns.
4. History of incorrect error patterns.

Each of these slots contains a number. They are initialized to zero at the start of a session. Upon completion of a column in the predictions table, their counts are updated. For example, if a student correctly predicts HR in RR then the value of RR's history of correct prediction slot is incremented. If a student violates the SV/CO direct causal relationship in DR then the value of DR's history of incorrect error pattern slot is incremented.

The following is a list of overlay slots attached to each predictions table variable:

1. Prediction history.

2. Primary prediction history.

3. Primary prediction direction history.

If, as an example of a variable's prediction history slot, a student correctly predicts HR in any column, a C is added to a list of Cs and Ws for the prediction history slot attached to HR. If, as an example of a variable's primary prediction history slot, CVP is the primary variable and the student incorrectly predicts the primary variable then a W is added to CVP's primary prediction history slot. If, as an example of a variable's primary predictions direction history slot, a student correctly predicts the direction (increase or decrease) of a primary variable, a C is added to its primary predictions direction history slot.

After extensive interviews with JAM and AAR it has been determined that they do not make use of an elaborate overlay model. They simply do not have the necessary memory capability. I have, however, designed the student modeller to record all observable phenomena. The inclusion of this overlay model will make future experiments possible.

4.4 Sensitized Classes of Error Patterns

Error patterns are initially sensitized after predictions are made and are updated during the ensuing dialogue. If an error is detected, associated error patterns are assigned a sensitized class. There are seven classes. The first two classes are assigned after the student has filled in a column of the predictions table. During dialogue, class 1 and class 2 error patterns may be re-sensitized as one of the other five classes. Each EP class provides important information to the instructional planner.

An error pattern is sensitized as class 1 when the student has violated a cardiovascular relationship. For example:

Student predicts: CO $\xrightarrow{+}$ and MAP $\xrightarrow{-}$

Actual changes: CO $\xrightarrow{+}$ and MAP $\xrightarrow{+}$

In this case the student violated the direct causal relationship between CO and MAP.

An error pattern is sensitized as class 2 when there is at least one error but the relationship is not violated. For example:

Student predicts: CO $\xrightarrow{+}$ and MAP $\xrightarrow{+}$

Actual changes: CO $\xrightarrow{-}$ and MAP $\xrightarrow{-}$

In this case the student did not violate the direct causal relationship between CO and MAP. However, the predictions for the particular problem are in error.

An error pattern is sensitized as class 3 when it was a class 1 error pattern but all errors associated with the error pattern were corrected during tutoring. The relationship would then not be currently violated and no errors would exist. The instructional planner may make use of the knowledge that this error pattern was initially sensitized as class 1.

An error pattern is sensitized as <u>class 4</u> when it was a class 1 error pattern, but all error(s) associated with the error pattern were corrected while another error pattern was being tutored. For example:

Student predicts: SV $\xrightarrow{+}$ CO $\xrightarrow{-}$ MAP $\xrightarrow{+}$

Actual changes: SV $\xrightarrow{+}$ CO $\xrightarrow{+}$ MAP $\xrightarrow{+}$

The student originally has two class 1 error patterns: 'SV directly affects CO' and 'CO directly affects MAP. If, while the first error pattern was being tutored, CO was changed to + then the student would have corrected the predictions for the second error pattern. Even though no errors are currently present, the instructional planner may want to test the student's knowledge about this relationship.

An error pattern is sensitized as <u>class 5</u> when it was a class 1 error pattern but, while another error pattern was being tutored, one or more errors were corrected. The relationship is not currently violated but at least one error still exists. For example,

Student predicts: HR $\xrightarrow{+}$ CO $\xrightarrow{-}$ TPR $\xrightarrow{-}$ MAP $\xrightarrow{+}$

Actual changes: HR $\xrightarrow{+}$ CO $\xrightarrow{+}$ TPR $\xrightarrow{+}$ MAP $\xrightarrow{+}$

This creates two class 1 error patterns: violations of 'HR directly affects CO' and 'MAP = TPR * CO'. If the instructional planner addresses the HR to CO error pattern, and it is corrected, then that error pattern becomes class 3. The student would now have a prediction of CO +. The 'MAP = TPR * CO' error pattern still has an error (TPR) but the cardiovascular relationship is not violated.

A <u>class 6</u> error pattern was a class 2 error pattern (wrong answer but no relationship violated) but while tutoring it, all errors were corrected. The relationship is not currently violated and no errors exist. The instructional planner may use the knowledge that this error pattern was sensitized as class 2.

A <u>class 7</u> error pattern was a class 2 error pattern. While tutoring another error pattern, predictions were changed such that the relationship is now violated. It is, in essence, now a class 1 error pattern.

4.5 <u>Attributes of Error Patterns</u>

Every error pattern has nine attributes. The values of these attributes provide information to the instructional planner.

1. Prediction History. This will be a chronologically ordered list of C's and W's (C for correct, W for wrong). If a student exhibits an understanding of

a concept when making predictions then a C is assigned to the prediction history of the corresponding error pattern. Likewise, a W is assigned if the student exhibits a misunderstanding of a concept.

2. Tactical History. The purpose of this attribute is to provide information as to how a sensitized error pattern was resolved. For example, the student might correct a prediction after the tutor provided a hint. Another typical scenario is that the tutor explained the concept after several hints failed to help the student in correcting erroneous predictions. The value of this slot is a chronologically ordered list of Hs (for hint) and Es (for explanation).

3. Confidence Factor. This attribute is calculated using the information in the error pattern's prediction history. It reflects the level of the tutor's confidence that the student understands the corresponding physiological concept. The possible values are the integers from -2 to +2. A +2 means a high confidence that the student understands; a -2 means a high confidence that the student does not understand (see Section 4.6).

4. Sequence Confidence. If (1) two predictions violate a causal relationship and (2) they were

predicted in the correct order and (3) the second of the two variables was predicted immediately following the first, then this attribute is set to "t" (for true). For example, a student might predict HR + and several predictions later select CO -. This sensitizes an error pattern that corresponds to the direct causal relationship from HR to CO. However, it is possible that this error is due to a misunderstanding that is not related to the relationship between HR and CO. When a student makes these two predictions (HR +, CO -), but in succession, then the tutor has a high degree of confidence that a misunderstanding of the HR to CO relationship caused the error pattern to be sensitized.

5. Sensitized Class. See section 4.4.

6. Prediction Errors. This is a list of the variables and their predictions that gave rise to the error pattern to be sensitized. For example, if HR is correctly predicted as + and CO is incorrectly predicted as - then the list of prediction errors for the HR to CO error pattern contains: CO -.

7. Student Difficulties. A student difficulty is a potential underlying cause for a sensitized error pattern (see section 4.8 for a more detailed description of student difficulties). For example,

consider a student who predicts CO + and then predicts CVP + These predictions suggest that the student may think an increase of blood exiting the heart has an immediate effect on the volume of blood entering the heart. This would be true if human blood vessels had properties like copper pipes, which they do not. A potential student difficulty in this situation is that the student does not realize that the cardiovascular system is a compliant structure.

8. Severity. Some physiological concepts are more fundamentally important than others. The reason that JAM and AAR determine that a concept is important is because the concept must be correctly applied in many problem solving situations. The severity values are the integers from three down to one. The error pattern corresponding to a very important concept is assigned a severity of three.

9. Frequency. Some error patterns are known, by the tutors, to be more frequently sensitized than others. For example, students regularly have an initial difficulty with the inverse causal relationship from CO to CVP. In this case, the corresponding error pattern is assigned a frequency value of three. Again, the possible values are the integers from three down to one.

4.6 <u>Confidence Factors</u>

Shim (1991) explored the use of Confidence Factors (CFs) for a previous version of CST. The purpose of a CF is to provide the Instructional Planner with an estimate of the likelihood that a student understands a concept. There are two necessary components in a CF: a response history and a time dependent function. The response history is a sequence of Cs and Ws representing correct and wrong. When a student demonstrates an understanding of a concept, a C is added to the response history. A W is added when a student demonstrates difficulty with a concept. A CF is obtained by applying the function to the history. The theory behind Shim's function is that older responses should carry less weight, new responses more weight. For example, consider a student whose response history is (C C W W), where the leftmost response is the most recent. This scheme, while seemingly counter intuitive, arises from the LISP programming environment. This student is more likely to be able to apply this concept than a student whose response history is (W C W C).

Shim's CF function yields a real number between zero and one. After interviews with JAM and AAR, I have modified CFs slightly to better reflect the tutor's behavior. First, JAM and AAR use very coarse grained heuristics and this should be reflected in the CF. Second, very old responses carry no weight at all. The current algorithm for the CF function can be found in Table 2. The response history column contains no more than the three most recent values. The leftmost value is the most recent.

Table 2. Confidence Factor Algorithm

RESPONSE HISTORY	CONFIDENCE FACTOR
No history	No value
(C)	0
(W)	-1
(C C)	1
(C W)	0
(W C)	-1
(W W)	-2
(C C C)	2
(C C W)	1
(C W C)	0
(C W W)	0
(W C C)	0
(W C W)	-1
(W W C)	-2
(W W W)	-2

While this algorithm is coarse grained, it does not violate the spirit of Shim's algorithm. Every time a C is added to the response history, the CF goes up (except when the last three responses were already C). Every time a W is added to the response history, the CF goes down (except when the last two responses were already W).

CST (v.3) contains a variety of prediction history and dialogue history slots. The first is updated after the student makes predictions, the others after a segment of the dialogue. This algorithm will be applied to both types of history slots. For example, if the instructional planner wants a CF for the students ability to predict the CO to MAP direct causal relationship, an algorithm is applied to the appropriate history slot.

Again, my goal is to model the behavior of JAM and AAR. A worthwhile future experiment would be to test Shim's CF algorithm against the one I have designed. This experiment might help answer an interesting question: if a tutor has perfect memory, will that produce better tutoring?

4.7 Student Difficulties and Error Patterns

After predictions are made, JAM and AAR select an error pattern to tutor. They usually attempt to tutor about the corresponding physiological concept. Sometimes a substantive student difficulty (not just an inadvertent mistake) is exposed during dialogue. In these cases, they try to remedy the student difficulty. Dialogues 2 and 3, excerpts from tutoring transcripts, show the distinction between tutoring the physiological concept that corresponds to an error pattern and tutoring the cause of the error pattern, a student difficulty.

In Dialogue 2, the observed error pattern corresponds to the inverse causal relationship between CO and CVP. The dialogue suggests that the student difficulty was an

Dialogue 2 from K30-tu-146:

T: Explain why you think that CVP increased.

S: Because the MAP in the system increased

T: What are the determinants of CVP?

S: It was the cardiac output that affected CVP. But I don't understand why MAP wouldn't affect it too.

T: MAP is not distributed all the way over to the right atrium. Instead it's the atrial volume that becomes the overriding effect.

S: Ok I understand

incorrect notion that there is a direct causal relationship between MAP and CVP. Because this was discovered by the tutor, the student difficulty, not the error pattern, is addressed.

Dialogue 3 is an example of a tutor addressing an error pattern. The observed error pattern corresponds to the inverse causal relationship between MAP and SV. JAM and AAR claim that in most situations it is more expedient to tutor this physiological relationship associated with the error pattern than attempt to isolate a student difficulty. The effort to isolate an underlying cause is counterproductive.

Dialogue 3 from K30-tu-160:

T: Stroke volume is affected by MAP because MAP is the afterload on the ventricle. Increasing MAP increases the afterload. How would that affect SV?

S: Stroke volume would decrease

T: Yes.

There are any number of reasons why a student may make errors and expose error patterns. Sometimes the underlying cause is merely an inadvertent mistake, called a slip; sometimes the student is just simply confused. The student may be missing a piece of information, may use an incorrect piece of information, or may be unable to correctly apply a piece of information. In any case, underlying causes are called student difficulties in CST. The distinction between these classes of student difficulties is prevalent in the student modelling literature. Much of it is not, however, a part of the diagnostic process used by JAM and AAR. For example, the difference between declarative student difficulties (i.e., a missing piece of knowledge) and procedural student difficulties (i.e., an inability to apply rules) is not a factor in JAM and AAR's diagnostic process. Therefore, all terminology used in the student model for CST (v.3) avoids existing terminology that may suggest a distinction between classes of knowledge. Specifically, the model of student difficulties I propose is similar to Brown and Burton's (1982) bug paradigm; but it is also significantly different. I purposely call underlying causes of error patterns "student difficulties" to avoid any connotations of the bug paradigm.

JAM and AAR claim that they do not deliberately look for, or diagnose, underlying misconceptions. While the bug paradigm is prevalent in ITS literature, JAM and AAR's behavior is consistent with other tutors as found in the literature of human tutoring. Lepper and Chabay (1988) claim tutors do not perform error diagnosis; they only attempt to get the student to recognize errors. Putnam (1987) found, in a study of arithmetic tutors, that tutors looked for the nature of the student's difficulty only 7% of the time. Graesser (1993) notes that tutors do not tailor their dialogue to the knowledge and misconceptions of students; they drill students with predefined examples. He concludes "the effectiveness of tutoring cannot be attributed to the implementation of a sophisticated pedagogy."

What is the reason for this behavior? My analysis of the behavior of JAM and AAR suggests the following:

1. It is time consuming to identify substantive student difficulties (i.e., not just a slip). This problem is magnified by the keyboard to keyboard environment. In any environment, however, this diagnosis takes away from actual tutoring.

2. There is not always an identifiable student difficulty. Many errors in this domain are the result of confusion, not an identifiable misconception or missing piece of information.

JAM and AAR determine *what* to tutor by identifying error patterns. They attempt to follow a protocol in which the student makes six or seven predictions without interruption. Because of the dynamics of human tutoring, they sometimes fail to adhere to this protocol. Of the eighteen transcripts I have analyzed, there are 58 times that:

1. The tutor has followed the protocol closely enough so that at least 3 uninterrupted predictions were made.

2. There was at least one prediction error in a sequence of uninterrupted predictions.

3. There was time available to tutor after a sequence of predictions were made.

Following is a categorization of the initial responses to the student in these 58 situations:

1. Open ended questions. In this category of response, the tutor prompts the student for an explanation of why predictions were made. The tutor generally has multiple intentions. The question may provide a hint and the student has the opportunity to discover an error. The tutor can often verify the most important error pattern. Sometimes a student difficulty is exposed (In most cases, CST's input understander will not be able to parse the student's response to this type of question). This type of response occurred 32 times. Sometimes the question is implied. An example from k33-tu-72:

T: Ok. Let's talk about your predictions. You said that SV increased. Explain your reasoning.

2. Questions about determinants. The tutor has decided on a strategy for remedying an error

pattern. The error pattern corresponds to a causal relationship. This type of response occurred 13 times. An example from K15-tu-59:

> T: Not bad ... But there are some mistakes we ought to discuss. You predicted that SV would increase. What are the determinants of SV?

3. Questions about perturbations. The tutor has reason to believe that information about the perturbation in question will help the student in correcting prediction errors. The tutor has, in a sense, inferred a student difficulty solely from predictions. The tutor has not tried to verify the existence of a student difficulty through dialogue. This type of response occurred twice. An example from K-48-tu-108:

> T: Ok, one minute. Tell me what the perturbation is for this patient.

4. Questions about the response phases. As with questions about perturbations, there may be a student difficulty inferred from predictions, but not confirmed through dialogue. This type of

response occurred twice. An example from K-30-tu-34:

> T: Good. Let's talk about your predictions. First, what does DR mean?

5. Questions about the neural variables. This is similar to questions about determinants except that the error pattern the tutor has focused on corresponds to the causal effects of the central nervous system. This type of response occurred three times. An example from K10-tu-29

> T: Ok. Let's take a look at some of your predictions. Take the last one first. Can you tell me how TPR is controlled?

While there are times that JAM and AAR tutor student difficulties in the transcripts I have analyzed, they do not deliberately look for them on a regular basis. A complete listing of the 58 initial responses can be found in Appendix C.

4.8 Student Difficulties in CST

Instances of the following two student difficulties are attached to every error pattern:

1. Slip. This is an inadvertent mistake.

2. Does Not Know Concept. This is always a possible cause of a sensitized error pattern.

Attached to every student difficulty is a history slot. It is yet to be determined how this history slot will be used; it is available for use when the instructional planner is implemented. The following is the list of substantive student difficulties (misconceptions), and their associated error patterns, that are currently implemented. This list is a compilation of common misunderstandings observed by JAM and AAR:

1. "Does not know definition of DR" student difficulty attached to the "Neural variables remain unchanged in DR phase (except for primary neural variables)" error pattern. Understanding the role of DR is one of the "rules of the game" referred to in Section 3.3.

2. "Preload confusion" student difficulty attached to the "CVP to SV direct causal relationship" error pattern. Preload is a term applied to several cardiovascular parameters related to filling of the heart.

3. "In/Out balance of the heart confusion" student difficulty attached to the "CO to CVP inverse

causal relationship" error pattern. The student thinks a change in the amount of blood exiting the heart has an immediate effect on the volume of blood entering the heart.

4. "Afterload confusion" student difficulty attached to the "MAP to SV inverse causal relationship" error pattern. Some students think that an increase in MAP causes an increase in SV.

5. "Does not know definition of RR" student difficulty attached to the "MAP in RR makes the opposite qualitative change from MAP in DR" error pattern. Again, this is one of the "rules of the game."

6. "Does not know that MAP is the regulated variable" student difficulty attached to the "MAP in RR makes the opposite qualitative change from MAP in DR" error pattern.

7. "Sympathetic/Parasympathetic confusion" student difficulty attached to the "MAP in DR phase to TPR in RR phase," and the "MAP in DR phase to HR in RR phase," and the "MAP in DR phase to IS in RR phase" error patterns.

8. "Does not know how to sum the DR and RR columns" student difficulty attached to the "SS qualitative change for a variable is the algebraic sum of the

qualitative changes for that variable in DR and RR" error pattern.

9. "Does not know the effects of clamping" student difficulty attached to the "all clamped variables remain unchanged in RR" error pattern.

10. "Causality/algebra confusion" student difficulty attached to the "MAP = TPR x CO," and the "CO = HR x SV" error patterns. The formula is written so that it appears to be an algebraic statement; it is not.

11. "Thinks pressure is fully compensated" student difficulty attached to the "MAP in SS makes the same qualitative change as MAP in DR" error pattern. The steady state is the period of time after the effects of the reflex has propagated throughout the CV system. It normally takes a much longer period of time for MAP to return to its pre-perturbation level. The reflex reverses the effect on MAP; it does not fully compensate for it.

4.9 Global and Local Assessments

I regularly sat in a room with JAM and AAR during keyboard to keyboard tutoring sessions. In one session, a student solved a problem after AAR provided a hint. AAR, as he was typing, said "I've got a sharp one here!" This comment, along with subsequent interviews, made me realize that JAM and AAR rely heavily on an overall, or global, assessment of the student. They use this assessment as a guide. If the global assessment is high, they provide hints in hopes that the student can solve a problem with little help. In doing so the student actively explores the problem solving process. If the global assessment is low, they provide more direct help.

The two primary components of the global assessment are the student's demonstrated mastery of the domain and the student's responses to questions and hints. Obviously, if the student can solve the CST problems then the tutor will have a high assessment of the student. Likewise, if a student responds well to questions and hints, the tutor is likely to maintain a positive assessment of the student. Responding well may mean the student correctly answered a question; it may simply mean that the student realized that the tutor offered a useful hint. A student that does not even recognize that a hint has been offered is likely to be given a low assessment in the tutor's mind.

In CST, the tutor has no global assessment at the start of a tutoring session. (This may not be completely true in traditional human tutoring because the tutor may know something about the student.) During a tutoring session a global assessment emerges. CST calculates a global assessment by using a formula that includes (1) the percentage of correct predictions made, (2) the percentage of surface layer relationships not violated by the student's predictions and (3) the percentage of correct responses to questions and hints. Specifically, a ratio is calculated by dividing the sum of:

- Total of correct predictions, plus
- Total of error pattern predicted correctly, plus
- Total of primary variable chosen correctly, plus
- Total of direction of primary variable predicted correctly, plus
- Total of correct responses to hints and questions, plus
- Number of error patterns where tutoring was terminated without an explanation.

by the sum of:

- Total of incorrect predictions, plus
- Total of class 1 error pattern sensitized, plus
- Total of primary variable chosen incorrectly, plus
- Total of times the direction of primary variable predicted correctly, plus
- Twice the total consecutive class 1 error pattern (Specifically, this is the number of consecutive Ws

in the prediction history slot of an error pattern.), plus

- Total of correct responses to hints and questions, plus
- Number of error patterns where tutoring was terminated with an explanation.

A number between -2 and +2 is generated from this ratio using the following formula:

```
if ratio < 0.8
                     then global assessment is -2
else, if ratio < 1.5
                     then global assessment is -1
else, if ratio < 1.9
                     then global assessment is  0
else, if ratio < 2.3
                     then global assessment is +1
else,
                          global assessment is +2
```

Subsequent interviews with JAM and AAR revealed that they also maintain a local assessment. This assessment is a measure of how well a student is progressing on a particular topic. In CST, the local assessment is initialized to an empty list at the start of a topic. The judger may add a C or W to the local assessment list after determining the correctness of a response.

The global assessment and local assessment are used to determine the choice of tactics. These concepts are discussed in Chapter V.

CHAPTER V

HINTING AND OTHER TUTORING TACTICS

Analysis of tutoring transcripts and interviews with JAM and AAR reveal a relationship between the tutor's perception of the student (the student model) and the tutor's inclination to encourage active learning. JAM and AAR are constantly guiding the student to discover and correct erroneous predictions. They regularly prompt the student with hints. The use of expository tactics is a last resort, after attempts at hinting have failed.

5.1 Tactics

I have identified eight tactics used by JAM and AAR: explanation, summarization, information gathering questions, use of two types of hints, a multi-turn dialogue where the tutor asks directed questions, and two types of acknowledgments: positive and negative. The analysis of these tactics is difficult because the tutor generally has multiple purposes for each of the utterances (Evens et al., 1993). For example, JAM and AAR generally ask an information gathering question after the student completes predictions for a column (see Dialogue 4). This type of question allows the student to quickly change an inadvertent error. The student's response may also provide the tutor with information about an underlying difficulty.

In fact, it may also be intended to provide a hint. It is, however, extremely difficult for the tutors to reconstruct all of their intentions while reading the transcripts of tutoring sessions.

Dialogue 4 from K47-tu-56:

> T: OK. That completes your DR predictions. Most of them are correct. However, I want to pursue IS with you. Can you tell me what you think that IS means?

Dialogue 5 is an example of a negative acknowledgment. It also further illustrates JAM and AAR's tendency to produce multiple intentions in a single utterance. The tutor is clearly providing a hint with the negative acknowledgment.

Dialogue 5 from K30-st-73:

> S: If stroke volume increases, that means that the rap will decrease
>
> T: You are correct, if we're talking about a situation in which SV or CO changed first. What if RAP changes first. How will that affect SV?

Dialogue 6 from K9-tu-29:

> T1: When CO increases, it does so at the expense of the CBV (venous volume). During that period, CO exceeds venous return. When the venous return finally catches up, CO = VR. However VR never exceeds CO so the veins cannot be refilled without reversing the original process, i.e., reducing CO.

Dialogue 7 from K34-tu-74:

> T: Let me summarize. There are three neurally controlled CV effectors. CC, HR and TPR. If the stimulus that we apply to the system doesn't act directly on any of them none of them will change in the DR. Understand?

Both explanations and summaries require that the student passively absorb information. Summaries review what the tutor presumes the student knows, what has transpired in the dialogue, or what predictions were made. One common use of summaries is to conclude a topic under discussion. Explanations are, in a sense, presentations of complete segments from the domain knowledge base. Explanations and summaries are used as components in the construction of hints (see section 5.2); they are also used as tactics in their own right. An example of an explanation is found in Dialogue 6 and an example of a summary is found in Dialogue 7.

5.2 Hints

Tutoring sessions that allow mixed initiatives provide opportunities: (1) for students to be active and (2) for tutors to keep the students focused on the current lesson. Hinting is a tactic that encourages active thinking structured within guidelines dictated by the tutor. Hints function by activating otherwise inert knowledge thus allowing that knowledge to be used in the current task (Bransford et al., 1989). It is now recognized that students actively confronting information have better access to it and are better able to use it thereafter than students simply receiving it passively (Shuell, 1986).

Hints are a useful and common pedagogical tactic, particularly in one-on-one tutoring sessions. Hints can serve: (1) to activate otherwise inert knowledge making possible its recall, or (2) to stimulate the inferences required to complete a task using knowledge thought to be available to the student. Hints either explicitly convey information to the student or they point to information presumed to be available to the student. More specifically, I have previously defined a hint as (Hume et. al, 1993, p. 564):

> A rhetorical device that is intended to either: (1) provide the student with a piece of information that the tutor hopes will stimulate the student's recall of the facts needed to answer a question, or (2) provide a piece of information that can facilitate the student's making an inference that is needed to arrive at an answer to a question or the prediction of system behavior.

5.3 Different Types of Hints

I have identified ten different forms of hints used by JAM and AAR. Dialogue 8 contains an example of one kind of hint found in the transcripts of the tutoring sessions. First, the tutor asks a question [T1] and finds that the student can not provide an answer [S1]. In response to this, the tutor asks a question [T2] that actually provides information about a part of the physiological relationship not remembered, or not known, by the student. The student then correctly answers this question [S2]. Having been prompted by this exchange, the student is now able to answer the tutor's next question [T3]. The hint provided thus enables the student to reason to a correct answer to the originally stated question [S3].

An analysis of the transcripts suggests that hints can be categorized by the manner in which students are prompted with the information they need to proceed with their problem solving. Some hints directly convey information to the students (CI-Hints). Other hints point to pertinent information but do not explicitly convey information to the student (PT-Hints).

Dialogue 8 from K32-tu-104:

T1:	What's the relationship between CO and RAP?
S1:	I don't know.
T2:	When CO decreases, as it did in this case, what would that do to the volume of blood in the venous (central) compartment?
S2:	It would cause it to increase
T3:	And how would that affect RAP? What's the relationship between CO and RAP?
S3:	It would cause it to increase
T4:	Absolutely.

I originally categorized hints found in the transcripts of nine tutoring sessions (K30 through K38). Hints in each category can be conveyed in a number of identifiable forms. CI-Hints can be found in three general forms: (1) explanation followed by a question, (2) summary followed by a question and (3) some combination of explanation and summary followed by a question. In all forms, the question may be implied or explicitly stated.

What follows are definitions and examples of each category of CI-Hints:

A1. Summary followed by an explicit question. In this example, from k38-tu-144, the tutor reviews the discussion so far and follows up with a question.

> T: Right, so the first change that occurs in the system is that RAP is decreased. What do you want to predict next?

A2. Summary with a question left implied. In this example, from k32-st-229, the student makes an incorrect statement. The tutor provides a negative acknowledgment and then points out that a previous answer leads to a different conclusion. The tutor assumes the student will follow up appropriately even though there is no explicit follow up question.

> S: TPR INCREASED MORE THAN CO
>
> T: No. You said that both of them increased and they are the determinants of MAP. Can't be.

B1. Explanation followed by an explicit question. This example is from k33-tu-132.

> T: This drug acts on alpha adrenergic receptors. Do you know what cardiovascular structure has these receptors?

B2. Explanation with a question left implied. In this example from k34-st-155 the tutor, after an incorrect response, provides a complete explanation. This is very close to a PT-Hint. The reason it is classified as a CI-Hint is because of the context of the conversation. They have been talking about the effect of a drug on TPR.

> S: TPR would drop
>
> T: The drug is stimulatory to the contraction of vascular smooth muscle.

C1. Explanation and summary followed by an explicit question. This example is from k37-tu-56.

> T: Let me briefly deal with your second thought. Most of the blood that is displaced to the periphery is in the veins. While it is true that they will get bigger, the veins contribute so little to TPR that we can ignore this effect. So,

(continued from previous page)

your first line of thinking is correct - TPR in DR is unchanged. Do you want to now think about HR and CC?

C2. Explanation and summary with a question left implied. This example is from k33-tu-124.

T: Let's try again. MAP= COxTPR. CO doesn't change. TPR increases. MAP should increase by that logic. But you correctly said that the reflex doesn't completely correct MAP. So?

The form of PT-Hints can vary from a direct question ("What are the neurally controlled variables?") to a declarative statement ("You have not predicted RAP yet.") to an imperative statement ("Remember the definition of DR"). The tutor assumes that the student understands the current problem or question even if no explicit question is posed.

Following are examples of PT-Hints:

D. A statement. This example, from k38-tu-52, is interesting because the tutor prepares the student with the statement that hints are coming. These are

PT-hints because the student has to figure out why the statements are relevant.

> T: Let me give you two hints that might help you. First, think about the definition of DR. Second, think about afterload.

E. Answering a student with a question. In this example, from k34-st-139, the tutor responds with a question in lieu of directly answering a student question. Again, this requires a sequence of activity by the student. The student must understand the tutor's question, recognize the connection between the tutor's question and the original question and then infer or recall the answer to the original question.

> S: I think I'm confusing venous return with preload.
>
> T: What factor in the table represents preload?

F. Partial acknowledgment. The partial acknowledgment is part of the hint. Also, in this example from k31-st-121, the uppercase 'DIRECTLY' adds to the hint.

> S: With a lower HR, there is more time for filling, and a higher EDV, so the SV is increased:

(continued from previous page)

T: That's correct but certainly not the major thing that HR affects, directly. What parameter is DIRECTLY affected by HR?

G. Carrying out the implications of a response. The intent of the tutor is for the student to recognize and correct the error in logic. In this example, from k33-tu-122, the tutor is merely giving the student an opportunity to self discover a mistake.

T: Now look at your predictions: MAP D, TPR I, CO 0. Is this possible?

5.4 <u>Directed Line of Reasoning</u>

On further analysis I have identified another tactic that is very similar to hinting. A directed line of reasoning (DLR) is a multiturn dialogue segment in which the tutor prompts the student in a stepwise manner for information the student is presumed to have available. The sequence of tutor questions is designed to help the student reason about the problem at hand. The tutor generally does not, however, provide specific answers to the student until

Table 3. Distribution of Hints by Category and Form in Two Hour Tutoring Sessions

Category	Session by Number and Tutor									
	AAR 1-5					JAM 6-9				
	K30	K31	K32	K33	K34	K35	K36	K37	K38	Total
A1	2	3	2	0	4	1	1	3	1	17
A2	0	0	1	0	0	0	0	0	0	1
B1	10	6	4	2	5	7	9	4	6	53
B2	3	4	3	1	5	0	0	0	0	16
C1	0	1	2	1	1	0	1	1	0	7
C2	0	0	1	1	0	0	0	0	0	2
D	12	22	8	4	7	9	10	5	7	84
E	0	0	0	0	1	0	0	1	0	2
F	3	1	1	1	6	0	0	0	0	12
G	1	0	0	2	0	1	0	2	1	7
Totals:										
	31	37	22	12	29	18	21	16	15	201

Legend for Table 3:

CI-Hints:

- A1 Summary + question
- A2 Summary + implied question
- B1 Explanation + question
- B2 Explanation + implied question
- C1 Explanation/Summary combination + question
- C2 Explanation/Summary combination + implied question

PT-Hints:

- D Question, explicit or implied
- E Tutor replies to student's question with a question
- F Partial acknowledgment (positive or negative)
- G Summary or implications of incorrect student response

attempts at hinting or DLRs fail. Dialogue 9 contains an example of a DLR.

Dialogue 9 from K37-tu-212:

T1:	The afterload is defined as the pressure against which the ventricles must pump. Here we are talking about the MAP as the afterload for the left ventricle. In the DR period what happened to afterload?
S1:	It increased
T2:	Right, it increased because MAP increased. Then, what will the increased afterload alter?
S2:	SV
T3:	And how will SV change?
S3:	Decrease
T4:	Right. And what will follow from this change in SV?
S4:	CO will decrease
T5:	Right, so what happens to RAP in DR?
S5:	RAP will not change yet.

While analyzing transcripts, it is difficult to determine when a tutor decides to engage in a DLR. If a student does not respond well, the DLR does not materialize. In this case, the student provides short and correct answers to the tutor's prompts.

5.5 <u>Questions Raised by an Initial Study of Hinting</u>

I have initiated a study of the use of hints and other tactics by experienced tutors in the hopes of formulating a strategy for using hints in an ITS. Specifically, I have attempted to (1) define hints, (2) categorize them by form, intent of tutor, and content of information conveyed, and (3) observe when they are used and not used. Table 3 summarizes the total number of hints, in each category and format, over the first nine tutoring sessions I analyzed.

An analysis of the transcripts and discussions with the tutors suggests that the most important criterion for identifying an utterance by the tutor as a hint is the tutor's intention to assist a student in arriving at an answer without actually providing the answer. For example, a tutor may pose a question with the sole intention of gathering information about the student's cognitive state. On the other hand, an utterance of the same form may be intended to stimulate recall by the student or to provide enough information to allow the student to make an inference. The tutor's intention was the primary factor considered when identifying hints. The nature of information provided to the student was the primary factor considered when categorizing hints. Some hints supply some portion of the information needed to arrive at a desired conclusion (CI-Hints) without supplying the complete answer. Other hints only remind the student of certain information which, it is assumed, they know (PT-Hints). While it is obvious that hints occur when the student has made an error, it is less obvious how the tutor decides that hints are not working and stops using them.

There are a number of questions that naturally follow from this study. What are the rules used by tutors to generate hints? Then, having decided to hint, how are the content, category and form of hint chosen? The transcripts suggest that the student's cognitive state plays an important role in the generation of hints; a tutor must have some reason to believe that a particular hint may be a useful tactic. Evidence of student deficiencies in certain areas certainly help in the selection of domain knowledge that is presented in a hint. What are the other factors that contribute to the generation of a hint? What causes a tutor to use a PT-Hint as opposed to a CI-Hint? Why are some hints explicitly stated as questions? In what order are the decisions about content, category and form considered by the tutor?

These transcripts suggest that JAM and AAR form their hints quite differently. JAM almost always provides a CI-Hints with explicit follow up questions (categories A1, B1 and C1). Also, JAM used PT-Hints much less often than did AAR (an average of 9 per session versus 14 per session). The sample is, of course, small but this observation raises more questions about the use of hints. How much does personal style enter into the differences between experienced tutors?

Another important topic for research deals with the effectiveness of hints. For example, is one type of hint more successful in assisting students than another? Does the success of different types of hints vary in any systematic way with the cognitive state of the student (how much they know, how well they solve problems)? While it may be possible to count the number of hints that immediately produce a desired result, that statistic ignores important non-tangible considerations. Is the tutor providing hints that are too obvious and is the chain of reasoning too simple? On the other hand, can an obscure hint trigger a desired result after the tutoring session is completed? I have attempted to answer many of these questions.

I have concluded that studying the surface form of hints was impeding my search to identify important rules about hinting. It is clear from interviews with JAM and AAR that the tutor must first decide upon the topic to be tutored. Their rules for determining the topic to be tutored are almost identical; this has been verified by a study of the tutoring transcripts and interviews with JAM and AAR. The actual surface form of hints appear to be determined by personal preference. The most obvious example is that JAM rarely leaves questions implied; AAR often leaves questions implied (see Table 3).

After studying more transcripts, I totaled the usage of CI-Hints, PT-Hints and DLRs (Table 4). I included DLRs because hints and DLRs perform similar functions. They both prompt students; neither provides answers. By concentrating on the contents of hints and DLRs, rather than their surface form, I hoped to identify the rules that dictate when JAM and AAR use various tactics. Specifically, I wanted to identify the rules for hinting.

5.6 When to Hint in CST

The rules for hinting in CST are based on the behavior of JAM and AAR. Therefore, I have concluded that two conditions must be present for hinting to occur:

1. The student must have exhibited some deficiency or error. Of the 295 hints I have identified, only 5 were not preceded by some type of error.
2. The tutor must have some reason to believe that the student is likely to be able to respond positively to the hint.

There are two conditions when JAM and AAR cease to use hinting as a tactic. First, the student may consistently display a lack of background information or poor problem solving skills. In this case, the tutor will cease to use hinting in the tutoring session once that deficiency is identified. Second, if repeated hints (usually two) directed at a particular issue are not successful, hinting

Table 4. Frequency of Hints

Session #	Duration	Tutor	CI-Hints	PT-Hints	DLRs
K9	1 hour	AAR	1	1	1
K10	1 hour	AAR	4	3	1
K11	1 hour	AAR	7	5	
K12	1 hour	JAM	6	2	4
K13	1 hour	JAM	2	2	3
K14	1 hour	JAM	8	1	2
K15	1 hour	JAM	2	6	
K16	1 hour	JAM	3	5	
K30	2 hour	AAR	15	16	1
K31	2 hour	AAR	14	23	
K32	2 hour	AAR	13	9	
K33	2 hour	AAR	5	7	1
K34	2 hour	AAR	15	14	1
K35	2 hour	JAM	8	10	
K36	2 hour	JAM	11	10	1
K37	2 hour	JAM	8	8	
K38	2 hour	JAM	7	8	1
K47	2 hour	AAR	6	9	3
K48	2 hour	JAM	13	8	2
Totals:			148	147	21

will stop and the tutor will usually give an explanation. JAM and AAR have determined that repeated, unsuccessful hints are frustrating to the student and are likely to impede learning. JAM and AAR will abandon hinting as a tactic when the student simply does not have sufficient mastery of the domain material. An example of this is in session K9 (see Table 4); only two hints were provided in a one hour session.

There were 73 sets of predictions (columns) made during the 19 transcripts analyzed. Each column corresponds to a physiological phase (DR, RR or SS). Of the 73, there was at least one error in 62 of the columns. Of those 62, 57 times there was at least one hint provided in the ensuing dialogue. Of the five columns with prediction error(s) and no hint, twice time prohibited any meaningful dialogue and once the tutor initiated a DLR (similar to a hint). In essence, in only two out of 60 columns where error(s) occurred did JAM and AAR chose to solely provide explanations. JAM and AAR virtually always provide a hint when an error pattern has been identified.

5.7 <u>Student Responses to Hints</u>

I have classified student responses into seven categories:

A. Answer is incorrect. The following example is from K14-tu-77. The student provides the correct answer.

T: Well, you can start by thinking about the reflex that was activated and what it will seek to accomplish. What is the stimulus here that activates the reflex?

S: Hr i

B. Answer is partially correct. There may be a portion of the answer that is wrong and/or missing. The following example is from K14-tu-41. In this situation CO is correct, SV is wrong.

T: For a compliant structure (like a balloon filled with air) the pressure inside is a function of the compliance of the structure (how "stretchy" it is) and the volume it contains. What parameter in the predictions table relates to the volume that will be present in the central venous compartment?

S: CO and SV

C. Answer is completely correct. The following example is from K37-tu-38. The student attempted to answer the question; the answer is simply incorrect.

T: Just one hint, and then make a prediction.
Remember the definition of DR.

S: Decrease

D. Answer is not incorrect, but it does not address the tutor's intention. The student may have provided an answer that is essentially correct, but the answer was outside of the tutor's intended context. The following example is from K13-tu-73. In this example, the response in S1 is not incorrect but, as the tutor states in T2, it was not the intended answer.

T1: In order to predict SV you have to know what its determinants will do. You haven't predicted CC and are you sure you can predict RAP without having predicted CO?

S1: I thought that SV was a determinant of CO?

T2: You are right, it is. but, you know what the reflex is attempting to accomplish and can therefore determine what must happen to CO.

E. The student asks a question to clarify the tutor's intention. The following example is from K12-tu-91.

> T: There's practically no parasympathetic innervation of blood vessels (erectile tissue and a few other fun places). Most -- --almost all of the innervation to blood vessels is sympathetic and the primary effect is norepi acting on alpha receptors to cause vasoconstriction. Now what do you say about what the reflex does vis-a-vis TPR?
>
> S: I'm sorry I just got lost. Are you saying it is not vasodilation?

F. The student explicitly states that he/she does not know the intended answer. The following example is from K33-tu-132.

> T: This drug acts on alpha adrenergic receptors. Do you know what cardiovascular structure has these receptors?
>
> S: Not sure

G. The student does not recognize the tutor's intention. The following example is from K15-tu-73. In this case the tutor's intention was to get the student to focus on SV.

> T: Well, if CO 0 and HR 0 then SV would have to be 0 and you didn't predict that.
>
> S: True, I predicted CO i because i thought venous return might increase.

Using this categorization, I identified 153 strings of hints in the 19 transcripts I analyzed. A string of hints is a segment of tutoring dialogue where:

1. The tutor has determined a topic to focus on. This generally is an error pattern but occasionally is a student difficulty.
2. The tutor initiates the tutoring of a topic with a hint. It is rare for JAM and AAR not to hint; they adopt other tactics only when the global assessment of the student is low.
3. While tutoring that same topic, the tutor often provides more than one hint. Within this string of hints there may also be prompts from the tutor that do not contain a hint.
4. The tutor decides to move on to another topic or phase of the tutoring protocol, thereby

terminating the string of hints, because either:

a) The student has demonstrated an understanding of the topic by correctly answering hints and questions, or by changing predictions.

b) The tutor abandons hinting because the local and global assessment becomes too low. The topic is concluded with a summary or an explanation.

A summary of the categorization scheme and a complete list of the 153 strings can be found in Appendix B. Dialogue 10 contains an example of a string of hints. Comments are to the right of the dialogue.

The string of hints in dialogue 9 is described by the following notation:

PT *A* PT *C*

where the italicized letters refer to the response categories (above and in Appendix B) and PT refer to a PT-Hint. In this case there is a PT-Hint, followed by an incorrect answer (category *A*), followed by a PT-Hint, followed by a completely correct response (category *C*).

There is always at least one response for every hint. After a response the tutor may provide another hint, but

Dialogue 10 from K34-tu-182:

Tutor/Student Dialogue:		Comments on Dialogue:
T: This is the DR. How will HR change?		This is a PT-Hint, the tutor alludes to the phase of the response without providing an explanation.
S: MAP changing affects BR changing, affecting HR		The student is incorrect (response category *A*).
T: In DR no reflex changes have occurred yet.		The tutor provides another PT-Hint. Again, it is up to the student to determine the relevance of this information.
S: So HR will not change.		The student is correct and the topic is terminated. (response category *C*)

sometimes the tutor follows up with an explicit question. One purpose for identifying these strings was to determine how long JAM and AAR are willing to provide hints while tutoring a particular topic. In other words, when do they stop hinting? In the above example, the string is terminated with a correct student answer. However, when the final student response is not correct and the tutor subsequently provides an explanation.

5.8 <u>When to Discontinue Hinting</u>

A tutor has a domain topic in mind when he first provides a hint. Again, this hint is virtually always error driven. A string begins with either a PT-Hint or a CI-Hint and is terminated when the tutor decides the topic is over. The student may have successfully answered questions and changed predictions; the tutor may have simply provided an explanation.

Of the 153 strings of hints, 34 times the initial response of the student was incorrect (category *A*; Appendix B). The tutor followed up with another hint 21 times (61%). In fact, JAM and AAR are likely to follow up with a second hint for any response that is not fully correct. There were 63 initial responses that were not fully correct (any category except *C*). Thirty eight times (60%) a follow up hint was provided. Not only are JAM and AAR predisposed to provide hints to remedy problems, they are also willing to use that tactic more than once to achieve a goal.

When is the hinting tactic abandoned? JAM and AAR's behavior suggest that two failed hints (hints that do not provide correct or partially correct responses) is the most common limit. Table 5 displays the number of hints in strings of hints for two different situations found in the 153 strings of hints. The middle column represents the number of hints while successive responses were simply incorrect (category *A*). For example, in the row designating 1 hint in a string of hints, the 20 means that there were 20 times that: (1) the tutor provided a hint, (2) the student's answer was incorrect and (3) the tutor did not provide another hint while tutoring that topic. The notation of such a string of hints looks like:

CI *A*

The right column in Table 5 represents the number of hints while successive responses were not fully correct (any category except *C*).

Table 5. Follow Up Hints

Number of Hints in a String of Hints (n)	Frequency of n Hints While Successive Responses Were Incorrect	Frequency of n Hints While Successive Responses Were Not Fully Correct
1	20	34
2	28	31
3	7	7
4	2	7
5	1	4
6	0	1
7	0	1

JAM and AAR generally will provide a small number of follow up hints when a student continues to have difficulty. The limit is generally one follow up hint (a total of two hints in the string) when the student provides an incorrect answer. The tutor is likely to provide even more follow up hints when the student provides partially correct responses. JAM and AAR say that they will continue to hint so long that there is evidence that hinting might aid the student.

5.9 Hinting and the Global and Local Assessments

After predictions are made for a column, the instructional planner will select an error pattern to remedy. CST's global assessment (see Section 4.9) will provide the instructional planner with an integer value from -2 (lowest global assessment) to +2. The instructional planner, based on the global assessment, will determine if a hint will be provided. Because of JAM and AAR's strong inclination to initially provide hints (see Section 5.6), a global assessment of -1 or greater will result in a hint being initially provided. After every hint and question, the local assessment will be updated. Again, the local assessment's potential values are from -2 to +2. If a problem has not been remedied after a hint, the instructional planner's decision regarding follow up hints will be based on (1) the global assessment, (2) the local assessment and (3) the length of the current dialogue. The threshold values for these parameters have not been determined.

5.10 Is Hinting Successful?

Much more work must be done before the effectiveness of hinting can be measured. It is generally assumed that hinting aids in learning and long term retention. An analysis of the 153 strings suggests that hinting is effective, at least over the short term. Table 6 shows the frequency of the final student response in the 153 strings. The categories of responses in Table 6 are described in Section 5.7 and in Appendix B. One hundred and nine times (71%) strings were terminated upon a completely correct response from the student.

Table 6. Terminating Responses in Strings of Hints

Category of Response		Frequency of a Final Response Terminating a String of Hints
A	(Incorrect)	27
B	(Partially Correct)	5
C	(Fully Correct)	109
D	(Did Not Address Tutor's Intention)	2
E	(Clarification Question)	5
F	(Does Not Know)	4
G	(Does Not Recognize Question)	1

There is data that suggests that CI-Hints and PT-Hints serve different purposes (see Section 5.11) but it does not appear that one tactic is more effective than the other. Of the 109 strings that were terminated with a fully correct response (see Table 6), 62 times the final hint was a CI-Hint versus 47 times for a PT-Hint. PT-Hints, however, produced more fully correct responses. Of the 295 hints identified (see Table 4), 151 times the immediate response was fully correct. PT-Hints produced an immediate fully correct response 79 times versus 72 times for CI-Hints. Determining the effectiveness of hints is difficult for two reasons. First, it is impossible to completely reconstruct the tutor's intentions and instructional goals. Second, hinting may have long term benefits more significant than the correctness of immediate responses.

5.11 **What Type of Hint to Provide?**

Table 7 shows the frequency of initial hints (PT-Hint versus CI-Hint) provided in a string of hints by JAM and AAR. JAM has a slight tendency to initially use a CI-Hint while AAR has the opposite tendency. This is probably accounted for by personal differences. For example, it has already been noted that JAM is less likely to leave questions implied.

A very interesting pattern emerges upon analysis of the second hint in a string. A CI-Hint, by definition,

Table 7. Initial Hints by Tutor

Tutor: / Hint Category:	JAM	AAR
CI-Hint	43	36
PT-Hint	30	43

provides explicit information. It appears JAM and AAR gauge the student's need for explicit information. If a student displays a higher degree of difficulty (e.g. an incorrect response), then the tutor is likely to provide a CI-Hint. On the other hand, a student that shows some progress (e.g. a partially correct answer) is likely to receive a PT-Hint. Table 8 shows the frequency of follow up hints (PT-Hint or CI-Hint) after a response from a hint. The data in Table 8 ignores cases where there is no follow up hint. Again, the description of response categories can be found in Appendix B.

Table 8. Frequency of Follow Up Hints

Response: / Hint Category:	A (Incorrect)	B (Partially Correct)	C (Fully Correct)
CI-Hint	36	6	17
PT-Hint	16	11	35

It should be noted that a completely correct response does not necessarily terminate a topic. The tutor may partition the topic into sub-topics. There were, however, 109 strings of hints terminated by a completely correct response (see Table 6).

5.12 Analysis of Strings of Hints

The following is a summary of rules about hinting based on the tutorial behavior of JAM and AAR during the 73 Predictions Table columns and 153 strings of hints:

1. Initially try hinting when errors are made. The exception is when the global assessment is very low.
2. If the global assessment is sufficiently high, try a second hint if the first hint is not successful.
3. Continue to provide hints on a topic as long as:
   a) The global and local assessment are sufficiently high, and
   b) The number of hints in a string is sufficiently low.
4. If a follow up hint is to be provided then:

a) Use a PT-Hint when the local assessment is high, and

b) Use a CI-Hint when the local assessment is low.

5.13 <u>Continuum from Passive to Active Learning</u>

Five of the tactics I described in Section 5.1 can be positioned along a continuum spanning passive to active learning. Explanations and summaries are tutorial tactics that require that the student store information. Explanations and summaries may vary in the degree of cognitive processing they require of the student; the usual intent of a summary is to integrate several pieces of information. However, the integration is initiated by the tutor and active learning on the part of the student is limited.

DLRs prompt the student for cognitive activity that lies near the middle of the passive to active continuum. The student must reflect upon the tutor's sequence of questions. Successive steps may place greater or lesser cognitive demands on the student. Individual steps may consist of CI-Hints or PT-Hints. While this tactic often concludes with a summary, it is presumed that the directed line of reasoning process intrinsically requires the student to engage in an active cognitive process that aids learning and retention.

CI-Hints guide the student into a greater degree of cognitive activity than DLR's. The isolated questions in a DLR sequence are generally very direct and prompt the student for information presumed to be available. CI-Hints are generally more complex and they often lead the student into making an inference.

PT-Hints provide the student with the opportunity to engage in the most active learning. They point to information presumed to be available to the student. The student must first retrieve this information, often by answering a question to themselves. They then must understand why that information is relevant to the issue at hand. This relevance is what stimulates recall and/or aids inferencing.

The fact that JAM and AAR use PT-Hints as follow up hints after positive responses (see Table 8) demonstrates JAM and AAR's belief that PT-Hints promote the greatest degree of active learning. JAM and AAR gauge the capabilities of students and use tactics that promote an appropriate degree of active learning. When a student responds well to hints, he or she is likely to receive PT-Hints.

It is not possible to unequivocally place any particular tutorial utterance along the passive to active continuum for several reasons. A tutor's utterance may serve multiple purposes (Evens et al., 1993). The surface form of an utterance may not reveal the intention of the tutor; and the tutor's intention may not be correctly inferred by the student.

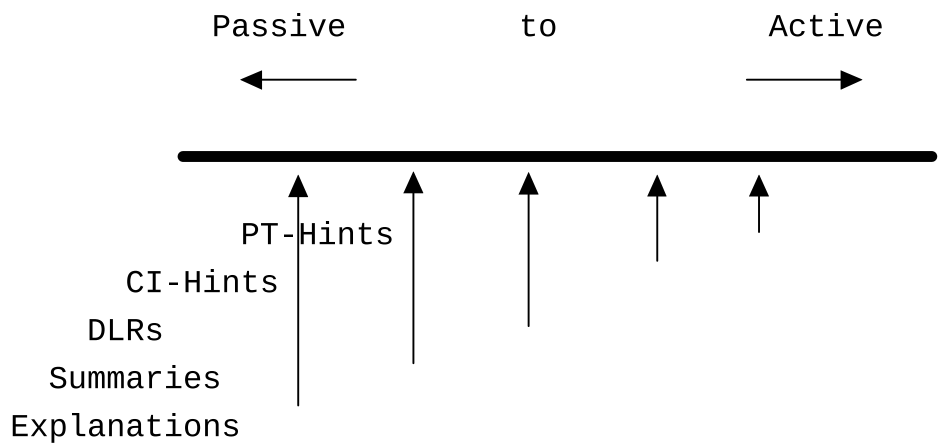

Figure 6. The Passive to Active Learning Continuum

There is, however, an passive to active learning continuum that is generally defined by the five tactics under discussion (See Figure 6). One thing is clear from my analysis: JAM and AAR encourage the student into as much active learning as possible. There is a definite tendency to choose a PT-Hint or a CI-Hint in order to remedy a problem. Also, DLRs are an alternate tactic that JAM and AAR believe promotes active learning.

CHAPTER VI

CONCLUSIONS

6.1 Summary

This thesis has described the design of a student modeller and an analysis of hinting by two expert tutors (JAM and AAR) in cardiovascular physiology. Their hinting behavior influences their modelling of the student; their model of the student influences their hinting behavior. The key elements of their behavior are:

1. JAM and AAR allow students to make several predictions before they engage in an interactive dialogue. This allows them to identify key patterns of errors (error patterns).

2. When an error pattern has been isolated, JAM and AAR generally tutor about the correct way to reason about the corresponding concept. They do not deliberately look for the underlying cause (student difficulty) of student errors. Occasionally the student difficulty is exposed and only then will they tutor about the student difficulty.

3. JAM and AAR use a very coarse grained scheme (student modeller) to evaluate students. They

form a global assessment of the student, a measure of the student's behavior throughout the tutoring session. They constantly form local assessments, an assessment of the student throughout tutoring on a given topic.

4. JAM and AAR attempt to use tactics that promote active learning. Specifically, they regularly use hints. The use of didactic tactics (i.e., explanation) is used when there is evidence that hinting will not be productive.

5. JAM and AAR rely heavily on their global and local assessments to determine when and how long to hint.

6. While the hints JAM and AAR use come in many surface forms, there are two significantly different categories of hints: convey information hints (CI-Hints) and point to information hints (PT-Hints).

7. While there are some differences in the patterns of hinting between JAM and AAR, there are many similarities. The differences are attributed to personal style and the similarities are attributed to the functions of CI-Hints and PT-Hints.

8. Most of the steps in student modelling discussed in this thesis have been implemented for CIRCSIM-Tutor (CST; v.3), an intelligent tutoring system (ITS). The features that are not implemented require a working input understander module and a discourse generation/text realization module.

6.2 <u>Significance</u>

The student model is the tutor's assessment of the student's cognitive state (VanLehn, 1988). After interviews with JAM and AAR, the tutors, and analysis of their human tutoring transcripts, I have concluded that the student model can and should provide answers to two important questions: *what* should be tutored and *how* should it be tutored? Modelling the student's responses to questions and hints is an essential part of providing these answers.

My research will impact the fields of intelligent tutoring and the broader field of Cognitive Science because:

1. This student modeller is designed to model the behavior of expert human tutors; student modellers in the literature are not based on an analysis of human tutors. The expert tutors act

according to the demonstrated ability of the student with regard to the domain material. Tutors gauge the student's ability to understand the implied messages that are attached to explicit statements. Specifically, tutors hint more when students respond well to hints; they hint less when students seem confused.

2. A systematic study of hints has been initiated. Hinting is commonly used by human tutors. ITSs have attempted to use hints. The current literature, however, provides no formal definition of hinting. A better understanding of when to hint and how to construct hints will lead to more effective computer tutoring.

3. The implementation of the student modeller provides platforms to conduct experiments. There has been little research performed on the effectiveness of student modelling and no research on the effectiveness of hinting. This student model will provide a tremendous amount of information that the instructional planner can choose to ignore or use. Which attributes of the student model should the instructional planner look at to enhance tutoring? Is the virtually limitless memory of a computer student model an asset?

6.3 Future Research

The design of student modelling for CST is based on the behavior of two expert tutors, JAM and AAR. Their behavior is consistent with the behavior described in tutoring literature. Specifically, they tend to (1) identify and tutor error patterns as opposed to student difficulties and (2) prompt students, with hints, to correct errors as opposed to didacticly providing explanations. Also, they are unable to remember (overlay model) all previous predictions and responses of students. They form an opinion (global assessment) of a student. This opinion may increase or decrease during tutoring.

JAM and AAR's tutoring has been proven to be effective (Khuwaja, 1994). What aspects of their behavior are responsible for this? What aspects of their behavior limit the effectiveness of tutoring. A fully implemented CST may help to provide some answers.

I have designed the student modeller so that any information that can be recorded is available to the instructional planner. The instructional planner decides what information to utilize. Future experiments can be designed by comparing the results of different instructional planners. Specifically, it would be interesting to see an instructional planner that relies on an elaborate overlay model as opposed to the very general estimate of a global assessment.

Much more work needs to be done on hinting. Specifically, CST needs to construct hints so that the resulting dialogue seems natural. At that point, experiments can be conducted to verify the effectiveness of hinting. In one experiment the instructional planner can provide hints by the rules outlined in this thesis. In another experiment the instructional planner can solely provide explanations.

The characteristics of CST tutoring are greatly determined by both the domain material (cardiovascular physiology) and the target audience (first year medical school students). How general is the design of my student modeller? How prevalent is hinting in other environments? Can this paradigm, in which local and global assessments determine when to hint, be applied in other environments? These are research topics I plan to pursue in the future.

APPENDIX A

INTERMEDIATE AND DEEP LEVELS OF THE CONCEPT MAP

APPENDIX B

STRINGS OF HINTS

Categories of responses within strings of hints:

A. Answer is incorrect.

B. Answer is partially correct. There may be a portion of the answer wrong or missing.

C. Answer is completely correct.

D. Answer is not incorrect, but it does not address the tutor's intention. The student may have provided an answer that is essentially correct, but the answer was outside of the tutor's intended context.

E. The student poses a question to clarify the tutor's intention.

F. The student explicitly states that he/she does not know the intended answer.

G. The student does not recognize the tutor's intention.

Categories of hints:

CI. CI-Hint.

PT. PT-Hint.

Following are the 153 strings of hints found in 10 tutoring transcripts. Responses are italicized.

In transcript K9

PT *B*

CI *C*

In transcript K10

PT *C* CI *A* CI *A*

PT *A* CI *F*

CI *C* PT *A*

In transcript K11

PT *C* CI *C*

CI *C*

CI *C*

CI *C* PT *C* CI *A* CI *A*

PT *A* CI *C*

PT *A*

In transcript K12

CI *F*

CI *C*

CI *A*

CI *C* CI *C* PT *C*

CI *E*

PT *C*

In transcript K13

CI *C*

PT *A*

CI *D* PT *C*

In transcript K14

CI *A* CI *B*

CI *C* CI *C*

CI *C*

PT *A* CI *E* PT *A* CI *C*

CI *C*

In transcript K15

PT *C*

PT *C*

PT *C*

CI *C*

CI *C C D* PT *C*

PT *G*

In transcript K16

PT *F* CI *C* PT *E*

CI *C*

PT *A C* PT *B*

CI *D* PT *A*

In transcript K30

CI *B* PT *B* PT *C*

CI *A* CI *C*

CI *C*

PT *C*

PT *C*

CI *A* CI *C*

PT *C*

PT *A* CI *E* CI *A* CI *E A* CI *C* PT *D D C*

PT *B*

PT *B* PT *E* PT *A A* CI *C* PT *B* CI *C*

PT *C*

PT *C*

CI *C*

CI *C*

PT *C*

CI *C*

PT *C*

In transcript K31

PT *A* CI *F*

CI *A* PT *A* PT *A*

PT *C*

PT *C*

CI *B* PT *A* PT *D* CI *C* PT *C*

CI *B* CI *A* CI *A* CI *D* PT *C* PT *C*

CI *B* CI *C*

PT *B* CI *B* PT *G* PT *C* PT *A* PT *C*
PT *E* PT *B* PT *C* PT *C*
PT *C*
PT *A*
CI *B* PT *A* CI *C*
CI *G* *C*
PT *C*

In transcript K32

CI *C*
CI *C*
CI *G* *A* PT *C*
PT *C*
CI *C*
CI *C*
PT *A* *F* PT *C* PT *C*
CI *E* *C* *C*
PT *D* *C*
PT *C*
PT *A* CI *C* CI *A* CI *B* CI *C*
CI *A* CI *C*
PT *F* CI *C*

In transcript K33

PT *C*
PT *C* PT *A* CI *C*
CI *F* PT *E* CI *C* CI *A* PT *B* *C*
PT *C* PT *F*
CI *C*

In transcript K34

CI *G* CI *A*
CI *C*
PT *E* *D* PT *C* CI *A* CI *C* PT *B* *C* PT *A* PT *E* CI *C* PT *C*
CI *A*
PT *C* PT *C*
CI *D* PT *C* CI *A* CI *C* CI *C*
CI *C*
PT *B*

PT *C* PT *C* CI *C*
CI *C*
CI *C*
PT *A* PT *C*
In transcript K35
PT *C*
PT *A* CI *A*
CI *A*
PT *C*
PT *A* *E* CI *E* CI *A* CI *B* PT *A* PT *A* CI *C*
PT *C* *E* PT *A* PT *A* CI *C*
PT *E*
CI *C*
In transcript K36
CI *A*
CI *C*
CI *E* *C*
CI *D* *C*
CI *C*
CI *A*
PT *C* PT *C* PT *C* PT *C*
CI *A*
CI *C* PT *C* PT *C* PT *C*
PT *C* PT *C* PT *C*
CI *A*
CI *A* CI *C*
In transcript K37
PT *C*
PT *E* PT *D*
PT *D* *D* CI *C*
CI *C*
CI *D* *A*
PT *C*
PT *A* CI *C*
CI *C*
PT *C*

CI *A* PT *C*
CI *C*
CI *C*
In transcript K38
PT *C* PT *A* CI *A*
CI *A* CI *C*
PT *B* PT *A* CI *C* PT *G* PT *E*
CI *A*
CI *C* PT *C* CI *A*
PT *A*
In transcript K47
CI *C*
CI *E* *A*
PT *C*
PT *D* *C*
PT *C* PT *A*
PT *C*
PT *C* CI *C* CI *B* PT *A* PT *A*
CI *D* PT *C* CI *C*
In transcript K48
PT *D*
CI *B* CI *C*
CI *A* CI *A* CI *C* CI *C*
PT *C* *C* CI *C*
CI *C*
PT *A*
CI *E*
CI *C*
PT *C*
CI *C*
CI *C* PT *A* PT *C* PT *A* PT *A* CI *A*

APPENDIX C

INITIAL TUTOR RESPONSES AFTER PREDICTIONS

The following are unedited utterances of JAM and AAR that follow a sequence of student predictions. The categories of these responses are discussed in Section 5.7.

====================

Open ended questions

====================

K11-tu-25-1: Well, let's re think this.

K11-tu-25-2: HR is a component of one of the other variables in the predictions table.

K11-tu-25-3: That variable is not CC.

K12-tu-75-1: Yes, again.

K12-tu-75-2: And in this regard.

K12-tu-75-3: It is MAP that is regulated by the BAROceptor reflex.

K12-tu-75-4: That's why it's called that.

K12-tu-75-5: Now let's get back to TPR.

K12-tu-75-6: Do you have any new thoughts about it?

K16-tu-33-1: Pretty good job!

K16-tu-33-2: But tell me why you predicted that cc would increase.

K30-tu-98-1: OK.

K30-tu-98-2: Consider CC (contractility).

K30-tu-98-3: You say that it will be decreased in the SS.

K30-tu-98-4: How did you come to that conclusion?

K30-tu-228-1: You said that the drug was still acting.

K30-tu-228-2: How can MAP be restored, then?

K30-tu-146-1: Ok.

K30-tu-146-2: Let's talk.

K30-tu-146-3: Explain why you think that RAP increased.

K32-tu-46-1: OK.

K32-tu-46-2: Now let's review some of your predictions.

K32-tu-46-3: You said that RAP (right atrial pressure) would decrease.

K32-tu-46-4: Why did you think that?

K32-tu-102-1: That finishes the RR predictions.

K32-tu-102-2: Let's talk about some of them.

K32-tu-102-3: I want t o start with the same one we talked about in DR, RAP.

K32-tu-102-4: Why did you predict that RAP D?

K32-tu-148-1: Pretty good.

K32-tu-148-2: I just have one thing I want to taklk about.

K32-tu-148-3: That's your pre diction MAP 0.

K32-tu-148-4: Why did you say that?

K32-tu-228-1: Correct.

K32-tu-228-2: And you have predicted CO I and TPR I.

K32-tu-228-3: So how can you say MAP D?

K33-tu-72-1: Ok.

K33-tu-72-2: Let's talk about your predictions.

K33-tu-72-3: You said that SV increased.

K33-tu-72-4: Expla in your reasoning.

K33-tu-164-1: Remember, the CO is a measure of the rate at which blood is being taken fro m the central venous compartment.

K33-tu-228-1: Why did you say SV D?

K34-tu-164-1: I want to back up for a mon\ment.

K34-tu-164-2: Can you explain why you said CC I?

K35-tu-26-1: Ok, you have made all the required predictions.

K35-tu-26-2: Can you explain how you ar rived at your predictions for tpr, hr, and cc?

K36-tu-62-1: Remember that we said that the drug (an agonist) occupies ALL of the alpha receptors.

K36-tu-62-2: Do you want to rethink your prediction?

K36-tu-148-1: Why did you predict cc last?

K37-tu-140-1: Notice that you predicted (for ss) that co and tpr would be increased (abov e their value before the centrifuge) but that map would be 0.

K37-tu-174-1: Remember the definition of dr and see what other variables might be easily predicted.

K37-tu-194-1: Ok, let's talk about these predictions.

K37-tu-194-2: How did you make your prediction a bout rap?

K37-tu-288-1: Do you want to look at your predictions for sv again?

K38-tu-188-1: Why?

K38-tu-222-1: You have predicted that map will be increased.

K38-tu-222-2: What does this mean to you?

K38-tu-222-3: That is, relative to the value of map BEFORE the centrifuge, what will be the value of map in the ss (centrifuge still going and the cv system in ss)?

K47-tu-56-1: OK.

K47-tu-56-2: That completes your DR predictions.

K47-tu-56-3: Most of them are correct.

K47-tu-56-4: However, I want to persue IS with you.

K47-tu-56-5: Can you tell me what you think that IS means?

K47-tu-94-1: I think that we have a problem here.

K47-tu-94-2: You are to predict the changes from DR that are caused by the baroceptor reflex.

K47-tu-94-3: The reflex can't affect HR, but that doesn't mean that it can't affect other variables.

K47-tu-94-4: Is that the way you understand it?

K47-tu-136-1: Good then we can begin discussing your predictions.

K47-tu-136-2: After you predicted HR 0 (correct) you predicted CO D (also correct).

K47-tu-136-3: How did you get the CO D prediction.

K47-tu-136-4: Explain.

K47-tu-182-1: OK then let's talk about some of your predictions.

K47-tu-182-2: Starting with MAP.

K47-tu-182-3: You predicted that it I in DR and that the reflex Ded it in RR.

K47-tu-182-4: And concluded that it would be lower in SS than before the pacemaker broke.

K47-tu-182-5: How did you decide that?

K47-tu-246-1: That's completes your DR predictions.

K47-tu-246-2: Let's talk about your CVP prediction.

K47-tu-246-3: How did you get it?

K47-tu-304-1: That concludes your RR predictions.

K47-tu-304-2: It's 3:00 can you continue for a bit longer?

K47-st-305-1: Definitely, I need the help!!

K47-tu-306-1: OK.

K47-tu-306-2: Then please look at your predictions for CO, TPR and MAP and tell me what you think.

K48-tu-44-1: Give me a minute to organize m,y thionking.

K48-tu-44-2: Ok.

K48-tu-44-3: You predicted that TPR would increase.

K48-tu-44-4: Can you explain how you arrived at that prediction?

K48-tu-218-1: Look over your predictions carefully and tell me if you want to change anything.

============================

Questions about determinants

============================

K9-tu-23-1: Well, I was waiting for you to enter each variable separately.

K9-tu-23-2: But this is OK.

K9-tu-23-3: Your first 3 entries were OK.

K9-tu-23-4: However, do you know of a relationship between RAP and CO?

K13-tu-29-1: What is the most important determinant of sv?

K13-tu-65-1: How can you predict the change in rap before you predict what happens to its determinat?

K14-tu-23-1: How can you predict a change in rap without having predicted what will happen to its determinant?

K15-tu-59-1: Not bad ...

K15-tu-59-2: But there are some mistakes we ought to discuss.

K15-tu-59-3: you predicted that sv would increase.

K15-tu-59-4: What are the determinants of sv?

K16-tu-59-1: What do you need to know first in order to be able to predict what rap will do?

K30-tu-64-1: Ok.

K30-tu-64-2: Now lets talk about yuour predictions.

K30-tu-64-3: First, what are the determinan ts of RAP?

K31-tu-54-1: OK.

K31-tu-54-2: Let's look at some of your predictions.

K31-tu-54-3: Can you tell me how HR is con trolled physiologically?

K31-tu-152-1: What's the relationship between CO and MAP?

K33-tu-118-2: Let's take a look at your last prediction.

K33-tu-118-3: MAP D.

K33-tu-118-4: What are the determina nts of MAP?

K35-tu-42-1: OK, let's talk about these predictions and the questions you seem to have a bout them.

K35-tu-42-2: You predicted that sv would increase because cc increased.

K35-tu-42-3: Wha t is the most potent determiner of sv?

K35-tu-98-1: What variable is the determiner of rap?

K38-tu-114-1: No.

K38-tu-114-2: Sv is determined by the extent of filling (which is determined by rap) and by cardiac contractility.

K38-tu-114-3: Rap is the more potent of the two and you h aven't predicted it yet.

K38-tu-114-4: What do you want to predict now?

=============================

Questions about perturbations

=============================

K37-tu-256-1: Great job!

K37-tu-256-2: Only one problem.

K37-tu-256-3: The description of this experiment said that the drug (the alpha agonist) occupied ALL of the receptors.

K37-tu-256-4: What effect w ill that have on the rr response of tpr?

K48-tu-108-1: Ok, one minute.

K48-tu-108-2: Tell me what the perturbation is for this patient.

====================================

Questions about the responses phases

====================================

K30-tu-34-1: Good.

K30-tu-34-2: Let's talk about your predictions.

K30-tu-34-3: Firs, what does DR mean?

K30-tu-180-1: Ok.

K30-tu-180-2: What prediction did you make about the change in MAP in DR?

K33-tu-188-1: OK.

K33-tu-188-2: I meant to tell you that the drug we used was so strong that there's n o way to change TPR refexly.

K33-tu-188-3: So let's make it 0 in any case.

K33-tu-188-4: Could you st ate the function of the reflex in words?

K34-tu-48-1: That completes Hang on a minute.

K34-tu-48-2: Sorry I had to answer the phone.

K34-tu-48-3: That co mpletes your predictions of the direct response.

K34-tu-48-4: Can you tell me first wha t you think that we mean by the DR?

K36-tu-76-1: What reflex are we attempting to think through?

K37-tu-42-1: Ok, let's see what you have predicted.

K37-tu-42-2: First, though, can you tell me what DR means.

K48-tu-162-1: One minute...I'm not sure you understand how to think about the steady state (SS)>

K48-tu-162-2: Let's look at your predicts for hr.

K48-tu-162-3: You said that in DR hr was up and you said that it can't change so in RR it was no change.

K48-tu-162-4: In the new SS what must be the change (relative to the state before the perturbation occurred)?

K48-tu-298-2: I think you're still having trouble with SS.

K48-tu-298-3: You said tpr would decrease in DR (the drug effect), won't change in RR (it can't) and predicted that it would be unchanged in SS.

K48-tu-298-4: Remember,compare the value in SS to the v alue BEFORE the change in the system.

====================================

Questions about the neural variables

====================================

K10-tu-29-1: Ok.

K10-tu-29-2: Let's take a look at some of your predictions.

K10-tu-29-3: Take the last one first.

K10-tu-29-4: Can you tell me how TPR is controlled?

K12-tu-33-1: By what mechanism will it increase?

K31-tu-54-1: OK.

K31-tu-54-2: Let's look at some of your predictions.

K31-tu-54-3: Can you tell me how HR is con trolled physiologically?

BIBLIOGRAPHY

Berne, R. M., & Levy, M. N. (1993). *Physiology, 3rd Edition*. St. Louis: C. V. Mosby.

Bloom, B. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-on-one tutoring. *Educational Researcher* 13: 4-16.

Bransford, J., Vye, N., Adams, L. & Perfetto, G. (1989). Learning skills and the acquisition of knowledge. In A. Lesgold and R. Glaser (Eds.), *Foundations for a Psychology of Education.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, J. & Burton, R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 155-191.

Burns, H. & Capps, C. (1988). Foundations of intelligent tutoring systems: an introduction. In M. C. Polson & J. Jeffrey Richardson (Eds.), *Foundations of Intelligent Tutoring Systems* (pp. 1-19). Hillsdale, NJ: Lawrence Erlbaum Associates.

Burton, R., & Brown, J. (1976). A tutoring and student modelling paradigm for gaming environments. *SIGSCE Bulletin* 8:236-246.

Burton, R., & Brown, J. (1982). An investigation of computer coaching for informal learning activities. In D. Sleeman & J. S. Brown (Eds.), *Intelligent Tutoring Systems* (pp. 79-98). London: Academic Press.

Carr, B. & Goldstein, I. (1977). Overlays: a theory of modeling for computer-aided instruction. *AI Lab Memo 406 (Logo Memo 40).* Massachusetts Institute of Technology, Cambridge, MA.

Carbonell, J.R. (1970). AI in CAI: an artificial intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine*, 11, 190-202.

Chang, R., Evens, M., Rovick, A., & Michael, J. (1992). Surface generation in a tutorial dialogue based on analysis of human tutoring sessions. *Proceedings of the Fifth Annual IEEE Symposium on Computer-Based Medical Systems* (pp. 554-561). Durham, NC.

Chi, M., Bassok, M., Lewis, M., Reimann, P. & Glaser, R. (1989) Self explanations: how students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.

Clancey, W. (1986a). Qualitative student models. *Annual Review of Computer Science*, 1, 381-450.

Clancey, W. (1986b). Intelligent tutoring systems: a tutorial survey. *STAN-CS-87-1174*, Stanford University, Palo Alto, CA.

Clancey, W. (1987). *Knowledge-Based Tutoring: The GUIDON Program*. Cambridge, MA: MIT Press.

Clancey, W. (1991). The frame of reference problem in the design of intelligent machines. In K. VanLehn and A. Newell (Eds.), *Architectures for Intelligence: The 22nd Carnegie Symposiumon Cognition.* Hillsdale, NJ: Erlbaum.

Collins, A., & Stevens, A. (1982). Goals and strategies for inquiry teachers. In R. Glaser (Ed.), *Advances in Instructional Psychology II*. (pp. 65-119). Hillsdale, NJ: Erlbaum.

Evens, M., Spitkovsky, J., Boyle, P., Michael, J., & Rovick, A. (1993). Synthesizing tutorial dialogues. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp.137-142). Boulder, CO.

Florio-Ruane, S. (1987). Sociolinguistics for educational researchers. *American Educational Research Journal,* 185-197.

Fox, B. (1993a). *The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fox, B. (1993b). Correction in tutoring. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp.121-125). Boulder, CO.

Graesser, A. C. (1993). Dialogue patterns and feedback mechanisms during naturalistic tutoring. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 126-130). Boulder, CO.

Graesser, A., McMahen, C. & Johnson, B. (1991). Tests of some mechanisms that trigger questions. *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 564-569). Chicago, IL.

Green, J. (1983). Research on teaching as a linguistic process: A state of the art. In E. Gordon (Ed.), *Review of Research in Education* (Vol. 10, pp. 152-252). Washington, DC: American Educational Research Association.

Hume, G. (1992). A dynamic student model in a cardiovascular intelligent tutoring system. *Proceedings of the Fifth Annual IEEE Symposium on Computer-Based Medical Systems* (Durham, NC) (pp. 370-377) Los Alamitos, CA: IEEE Computer Society Press.

Hume, G. & Evens, M. (1992). Student modeling and the classification of errors in a cardiovascular intelligent tutoring system. *Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society Conference* (pp. 52-56). Utica, IL.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1993). The use of hints as a tutorial tactic. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 563-568). Boulder, CO.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1995a). Controlling active learning: how tutors decide when to generate hints. *Proceedings of the 8th Florida Artificial Intelligence Research Symposium* (pp. 157-161). Melbourne Beach, FL.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1995b). Hinting as a tactic in one-on-one tutoring. To appear in the *Journal of Learning Sciences*.

Kamsteeg, P. & Bierman, D. (1985). A method for the study of teachers' thinking in coaching situations. *Paper presented at the International Study Association on Teacher Thinking*. Tilburg, Netherlands.

Katz, S., Lesgold, A., Eggan, G., & Gordin, M. (1993). Modeling the student in sherlock II. To appear in the *Journal of Artificial Intelligence*.

Keene, S. (1989). *Object-Oriented Programming in Common Lisp.* Reading, MA: Addison-Wesley.

Khuwaja, R. (1994). A model of tutoring: facilitating knowledge integration using multiple models of the domain. *Unpublished Ph.D. dissertation*. Illinois Institute of Technology.

Khuwaja, R., Evens, M., Rovick, A., & Michael, J. A. (1992). Knowledge representation for an intelligent tutoring system based on a multilevel causal model. In C. Frasson, G. Gauthier, & G. I. McCalla (Eds.), *Intelligent Tutoring Systems: Proceedings of the Second International Conference, ITS '92* (pp. 217-224). Montreal, Canada.

Khuwaja, R. A., Rovick, A. A., Michael, J. A., & Evens, M. W. (1995). A tale of three tutoring protocols: the implications for intelligent tutoring systems. E. A. Yfantis (Ed.), *Intelligent Systems, Third Golden West International Conference* (pp. 109-118).Kluwer Academic Publishers. Dordrecht, Netherlands.

Kim, N., Evens, M., Michael, J., & Rovick, A. (1989). CIRCSIM-TUTOR: An intelligent tutoring system for circulatory physiology. In H. Maurer (Ed), *Computer Assisted Learning, Proceedings of the International Conference on Computer-Assisted Learning* (pp. 254-266). Dallas, TX. Berlin: Springer-Verlag.

Lepper, M., Woolverton, M., Mumme, D., & Gurtner, J. (1992). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S. Lajoie & S. Derry, *Computers as Cognitive Tools* (pp. 75-105). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesgold, A., Lajoie, S., Bunzo, M., & Eggan, G. (1992). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J. Larkin & R. Chabay (Eds.), *Computer Assisted Instruction and Intelligent Tutoring Systems: Shared Issues and Complementary Approaches* (pp. 201-238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesgold, A., Eggan, G., Katz S. & Rao, G. (1993a). Possibilities for assessment using computer-based apprenticeship environments. To appear in Regian, W., & Shute, V. (Eds.), *Cognitive Approaches to Automated Instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lesgold, A. Katz S., Greenberg, L., Hughes, E. & Eggan, G., (1993b). To appear in Dijkstra, Krammer and van Merrienboer (Eds.) *Instructional-Design Models in Computer-Based Learning Environments.* Heidelberg: Springer-Verlag.

Li, J., Seu, J., Evens, M., Michael, J., & Rovick, A. (1992). Computer dialogue system (CDS): a system for capturing computer-mediated dialogue. *Behavior Research Methods, Instruments*, & Computers, 24, 535-540.

Littman, D., Pinto, J., & Soloway, E. (1990). The knowledge required for tutorial planning: an empirical analysis. *Interactive Learning Environments*, 1, 124-151.

London, B. & Clancey W. (1982). Plan recognition strategies in student modeling: prediction and description. *AAAI-82 (pp. 335-338)*. Pittsburgh, PA.

Mark, M. & Greer, J. (1991). The VCR tutor: evaluating instructional effectiveness. *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 564-569). Chicago, IL.

McArthur, D., Stasz, C., & Zmuidzinas, M. (1990). Tutoring techniques in algebra. *Cognition and Instruction*, 7, 197-244.

Merrill, D., Reiser, B., Ranney, M., & Trafton, J. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of Learning Sciences*, 2, 277-305.

Michael, J., Rovick, A., & Evens, M. (1993). Analysis of tutoring sessions as an approach to uncovering the rules for generating a tutorial dialogue. *Paper presented at the National Tutoring Conference*, East Stroudsburg University, East Stroudsburg, PA.

Michael, J., Rovick, A., Evens, M., Shim, L., Woo, C., & Kim, N. (1992). The uses of multiple student inputs in modeling and lesson planning in CAI and ICAI programs. In I. Tomek (Ed.), *Computer Assisted Learning: Proceedings of the Fourth International Conference, ICCAL* '92 (pp. 441-452). Wolfeville, NS, Canada.

Ohlsson, S. (1987). Some principles of intelligent tutoring. In R. W. Lawler & M. Yazdani (Eds.), *Artificial Intelligence and Education, Vol. 1, Learning Environments and Tutoring systems* (pp. 203-237). Norwood, NJ: Ablex.

Park, O., Perez, R. & Seidel, R. (1987). Intelligent CAI: old wine in new bottles, or a new vintage? In G. Kearsley (Ed.) *Artificial Instruction & Instruction Applications and Methods*. (pp. 11-46) Reading, MA: Addison Wesley.

Person, N., Graesser, A., Magliano, J. & Kreuz, R. (1993). Inferring what the student knows in one-to-one tutoring: the role of student questions and answers. *Unpublished paper*.

Putnam, R. (1987). Structuring and adjusting content for students: a study of live and simulated tutoring of addition. *American Educational Research Journal,* 13-48.

Reiser, B. (1989). Pedagogical strategies for human and computer tutoring. *Proceedings of the American Educational Research Association*. San Francisco, CA.

Reiser, B., Friedmann, P., Gevins, J., Kimberg, D., Ranney, M. & Romero, A. (1988). A graphical programming language interface for an Intelligent LISP tutor. *Proceedings of CHI'88, Conference on Human Factors in Human Systems* (pp 39-44) New York: ACM.

Reiser, B., Ranney, M., Lovett, M., & Kimberg, D. (1989). Facilitating students' reasoning with causal explanations and visual representations. In D. Bierman, J. Breuker, & J. Sandberg (Eds.), *Artificial Intelligence and Education: Proceedings of the Fourth International Conference on AI and Education* (pp.228-235). Springfield, VA: IOS.

Reiser, B., Kimberg, D., Lovett, M., & Ranney, M. (1992). Knowledge representation and explanation in GIL, an intelligent tutor for programming. In J. Larkin, & R. Chaby (Eds.), *Computer Assisted Instruction and Intelligent Tutoring Systems: Shared Goals and Complementary Approaches*. (pp.111-149). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rovick, A., & Michael, J. (1986). CIRCSIM: An IBM PC computer teaching exercise on blood pressure regulation. *Proceedings of the XXX International Union of Physiological Sciences Congress* (p. 318). Vancouver, Canada.

Rovick, A., & Michael, J. (1992). The predictions table: a tool for assessing students' knowledge. *American Journal of Physiology*, 263 (Advances in Physiology Education, 8), S33-S36.

Sanders, G., Evens, M., Hume, G., Michael, J., Rovick, A. (1992). An analysis of how students take the initiative in keyboard-to-keyboard tutorial dialogues in a fixed domain. *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 1086-1091). Bloomington, IN.

Seu, J., Chang, R., Li. J., Evens, M., Michael, J. & Rovick, A. (1991a). Language differences in face-to-face and keyboard-to-keyboard tutoring sessions. *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 576-580). Chicago, IL.

Seu, J., Evens, M., Michael, J., & Rovick, A. (1991b). Understanding ill-formed input to an intelligent tutoring system in an LFG framework. *Proceedings of the Third Midwest Artificial Intelligence and Cognitive Science Society Conference* (pp. 36-40). Carbondale, IL.

Shim, L. (1991). Student modeling for an intelligent tutoring system: based on the analysis of human tutoring sessions. *Unpublished Ph.D. dissertation*. Illinois Institute of Technology.

Shim, L., Evens, M., Michael, J., & Rovick, A. (1991). Effective cognitive modeling in an intelligent tutoring system for cardiovascular physiology. *Proceedings of the Fourth Annual IEEE Symposium on Computer-Based Medical Systems* (pp. 338-345). Baltimore, MD.

Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research, 56*:411-436.

Spitkovsky, J. & Evens, M. (1993) Negative acknowledgments in natural language tutoring systems. *Proceedings of the 5th Midwest Artificial Intelligence and Cognitive Science Society Conference (pp. 41-45)*. Chesterton, IN.

Stansfield, J., Carr, B., & Goldstein, I. (1976) WUMPUS advisor I: a first implementation of a program that tutors logical and probabilistic reasoning skills. *Memo 381., AI Laboratory,* Massachusetts Institute of Technology, Cambridge, MA.

Stevens, A. L., Collins, A., and Goldin, S. (1982). Misconceptions in students' understanding. In D. Sleeman & J. S. Brown (Eds.), *Intelligent Tutoring Systems* (pp.13-24). London: Academic Press.

VanLehn, K. (1988). Student modeling. In M. C. Polson & J. Jeffrey Richardson (Eds.), *Foundations of Intelligent Tutoring Systems* (pp. 55-78). Hillsdale, NJ: Lawrence Erlbaum Associates.

VanLehn, K. & Jones, R. (1993). What mediates the self-explanation effect? knowledge gaps, schemas or analogies? *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp.1034-1039). Boulder, CO.

Weiser, A. (1974). Deliberate ambiguity. *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*. (pp. 723-731). Chicago, IL.

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems*. Morgan Kaufmann. Los Altos, CA.

Wilkens, D., Clancey, W., Buchanan, B. (1987). Using and evaluating differential modeling in intelligent tutoring and apprentice learning systems. In J. Psotka, D. Massey, & S. Mutter (Eds.), *Intelligent Tutoring Systems: Lessons Learned*. Hillsdale, NJ: Lawrence Erlbaum Publishers.

Woo, C. W., Evens, M., Michael, J., & Rovick, A. (1991). Dynamic instructional planning for an intelligent physiology tutoring system. *Proceedings of the Fourth Annual IEEE Symposium on Computer-Based Medical Systems* (pp. 226-233). Baltimore, MD.

Woolf, B. (1992). Building knowledge based tutors. In I. Tomek (Ed.), *Computer assisted learning: Proceedings of the Fourth International Conference, ICCAL '92* (pp. 46-60). Wolfville, NS, Canada.

Yob, G. (1975) Hunt the WUMPUS. *Creative Computing. 51-54.*