Evaluating Foundations: A Critique of Traditional Conceptions of Knowing and Learning in Intelligent Tutoring

Eckehard Doerry

CIS-TR-94-24
November 1994

Department of Computer and Information Science
University of Oregon

Evaluating Foundations: A Critique of Traditional Conceptions of Knowing and Learning in Intelligent Tutoring†

Eckehard Doerry
Dept. of Computer Science,
University of Oregon
Eugene, OR. 97403
eck@cs.uoregon.edu

"Somewhere there are people who understand [technology] and run it, but those are technologists, and they speak an inhuman language when describing what they do. It's all parts and relationships of unheard-of things that never make any sense no matter how often you hear about them."
— Robert Pirsig, *Zen and the Art of Motorcycle Maintenance*

1.0 Introduction

From the Renaissance to the Industrial Revolution to modern times, the rate at which scientific knowledge is added to the collective human store has increased exponentially. The knowledge and principles at the cutting edge of scientific inquiry in Newton's day are now the bread and butter of an average seventh grade curriculum. At the same time, mandatory education has led to an overburdening of the educational establishment; the student-teacher ratio has steadily deteriorated. It has gotten to the point where relatively few teachers are expected to teach increasing numbers of students more and more information in a limited time period.

Since the earliest days of computing, there have been those who have dreamed of addressing this dilemma by capitalizing on the speed and flawless memory of computers to create devices to store, organize and communicate human knowledge, augmenting or even replacing human-human instruction. Unfortunately, two decades of research in the area of Intelligent Tutoring Systems (ITS) has yielded mixed results at best. While tutoring systems — if they have been tested in the real world at all — have had some success in very simple procedural domains (Burton & Brown, 1979; Burton, 1982; Brown and VanLehn, 1980; Sleeman, 1982; Sleeman, 1984), generating convincing results in more demanding areas like basic physics (Hollans, Hutchins & Weitzman, 1984; Brown, Burton & DeKleer, 1982; Borning, 1981; Douglas & Lui , 1989), higher mathematics (Anderson

† This paper was originally presented in partial satisfaction of Area Exam in June 1992.

et al., 1985; Kimball, 1982; O'Shea, 1982) and abstract reasoning (Stevens, Collins & Goldin, 1982; Goldstein, 1982; Johnson & Soloway, 1985) has proven an elusive goal.

Instead of providing a review of existing tutoring systems and techniques in ITS[1], the goal of this paper is to approach the area from a broader analytical perspective in hope of gaining new insights and to suggest promising directions for future research. The overall focus of this paper is the nature, representation and development of students' domain knowledge, both in tutoring systems and in human problem solving. After surveying approaches to student modeling in existing tutoring systems, I explore the conceptions held by naive human problem solvers, and finally examine the philosophical foundations of both areas. Integrating observations from these three perspectives highlights directions for new research. In particular, I suggest that future work in ITS may achieve better results by:

1) shifting the traditional perspective to redefine the role of knowledge modeling in tutoring systems.
2) basing the design of the ITS, in particular the curricular component, on a deeper understanding of students' initial conceptions of a domain.

Section two begins with a brief review of grounding assumptions and the overall structure of tutoring systems before going on to survey approaches to representation of student knowledge in ITS. Section three steps back to review problem solving in humans, suggesting ways in which tutoring systems should be modified to better support the development of students' expertise. Section four reviews the philosophical assumptions inherent in tutoring systems and presents a plausible alternative foundation in order to promote a more open-minded and cautious approach to knowledge modeling in tutoring systems. The remaining sections close the paper, drawing conclusions and pointing out research challenges and limitations.

2.0 Intelligent Tutoring Systems

The line where rote number crunching ends and Artificial Intelligence begins has always been difficult to define. The subarea of intelligent tutoring systems is no exception; a distinction between mechanical computer aided instruction (CAI) and intelligent tutoring systems (ITS) is difficult to define succinctly. For the purposes of this paper, an intelligent tutoring system has the following characteristics:

[1] See Wenger (1987) for just such a comprehensive review. See also the collection of seminal papers editted by Sleeman and Brown (1982).

1) An explicit model of the domain or skills being taught, thereby allowing the system to define the notion of performance "expertise".

2) Dynamic diagnostic capabilities which allow the system to continually assess the state of the student's knowledge.

3) A notion of pedagogic expertise which ties the other elements together by allowing the system to dynamically modify the curriculum in response to student behavior. This reactive flexibility is the central defining feature of intelligent tutoring.

There are several foundational assumptions on which the entire tutoring systems enterprise in based. First, the system must have access to various types of knowledge about the domain, didactic methods, diagnostic techniques and so on. Thus, there is the assumption that this knowledge can be codified and represented in the machine. This leads to an even more drastic two-part assumption: Since the goal of tutoring is knowledge communication and since the system only "knows" what is symbolically represented in its memories, there is the assumption that a) the transference of the system's "expert" domain model is the goal of the tutoring process and b) that human problem solving (and learning) can be and is, in fact, based on such symbolic models. This assumption is clearly an instantiation of the Physical Symbol Systems Hypothesis (Newell, 1980) that is the foundation for most work in artificial intelligence. The following figure presents this hypothesis graphically.

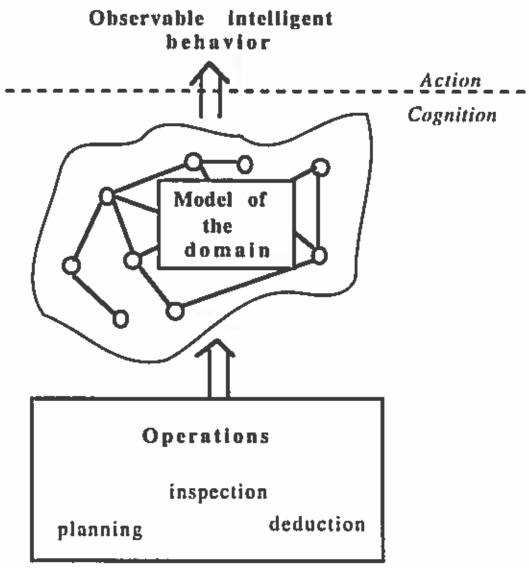

Figure 2.1: The Physical Symbol System Hypothesis

The essence of the Physical Symbol System (PSS) hypothesis is that all of human reasoning and behavior is based on an internal symbolic representation(s) of the world, a model, which is manipulated in various ways to arrive at conclusions and produce plans.

In this view, actions are a result of plan implementation. While the PSS hypothesis is widely accepted in the computer science community, there are skeptics elsewhere (Suchman, 1987; Winograd and Flores, 1986). In section 4, I present one such alternative view and consider its implications for ITS. For now, we will tacitly accept PSS in order to carry on with the discussion.

The foundation above serves as the basis for defining important terms used throughout this paper:

- A *model*, as applied to a tutoring system, is a symbolic representation of the world. Aspects of the world that will be modeled include domain knowledge, the student, and teaching strategies.
- A *mental model* refers to a human's internal representation of the world required by the PSS hypothesis. Again, it is this symbolic representation that is the basis of all reasoning and action.

With this terminology in hand, we can define the goal of an intelligent tutoring system more succinctly:

> The goal of an intelligent tutoring system is to accurately capture the mental model used by domain experts (in the system's "expert model") and to transfer that symbolic model to the student, thereby transforming her (by definition) into an expert as well.

Finally, I reserve the term *cognition* to refer to "the mental process that results in human behavior". In particular, I wish to disassociate the term with the model-based reasoning process dictated by the PSS hypothesis, using it, instead, as a neutral term.

2.1 Overview of Tutoring Systems

In the world of tutoring systems, there is no such thing as a "standard" approach — an incredible variety of perspectives, approaches and domains have been explored. However, most systems can be described in terms of an abstract framework proposed by Wenger (1987) and shown in Figure 2.2.

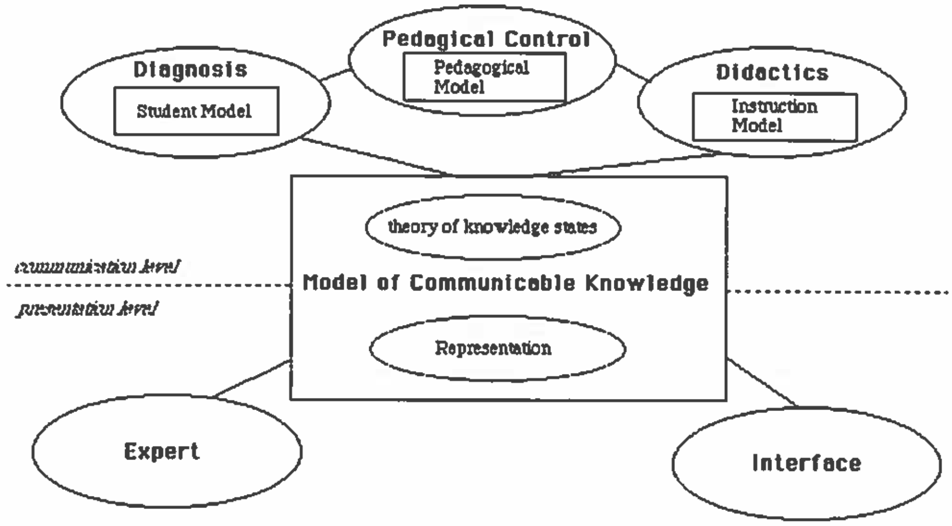

Figure 2.2: An analytic framework for tutoring systems (adapted from Wenger).

According to this framework[2], a tutoring system can be viewed as a collection of separate components, organized around some model of communicable knowledge, each of which performs one of the fundamental tasks in knowledge communication. Each of the components is briefly described below:

- The purpose of the student model is to represent the knowledge state of the student. Clearly, this component is central to the theme of this paper, since it is, in effect, a representation of the student's mental model of the domain. This component is the main focus of this paper and is discussed in detail later.
- The instructional model and didactic component together can be viewed as the explanatory engine for the ITS. That is, when the pedagogical model indicates that explanatory intervention is necessary, it calls on this component.
- The pedagogical component is concerned with higher level tutoring issues like when to intervene and decisions regarding the curriculum.
- The expert/task component is responsible for recording the context (e.g. what other problems have been attempted, environmental parameters, etc.) and acting as the system's problem-solving engine.
- The discourse model represents the interaction model on which the system's user interface is based.

[2] Young (1983) gives a similar framework in a paper oriented more towards user-interfaces. While this paper focuses primarily on the *user's* perspective, Young works to provide a broader view, covering user, system analyst and system designer. The similarity between models implies that we can generalize over the two areas, ITS and User Interfaces, to see both as instances of *knowledge communication systems*, which covers all systems aimed specifically at the transfer (versus the generation or manipulation) of information.

The model of communicable knowledge is the cornerstone of the entire framework, capturing the systems representation of domain knowledge. Whether represented as a semantic network, a set of low-level heuristics, or a set of more abstract issues, it is this knowledge that will be considered the distillation of "expertise" in the domain and the goal of the tutoring system to communicate to the student.

2.2 Student Modeling

If the central purpose of a tutoring system is to transfer some representation of expert domain knowledge to the student, then a primary subgoal must be to determine to what extent this transfer has occurred. The way in which most tutoring systems accomplish this is by *differential modeling*, by continuously comparing a model of the student's knowledge, as deduced from the student's problem-solving performance, with the system's representation of domain expertise. Thus, the primary purpose of the model of student knowledge in most systems is to serve as a resource for this comparative diagnostic process. In some systems, the student model is used for pedagogical and didactic purposes as well. For example, in both WUSOR-II and WUSOR-III (Goldstein, 1982), the system uses the student model to guide the selection of problems presented to the student. In WEST (Burton and Brown, 1979), the student model helps to drive didactic decisions like when and how often to intervene with hints and critique.

Though the basis of the diagnostic process, differential modeling, applies to most tutoring systems, there are a number of ways in which this approach can be implemented, each of which embodies a particular set of assumptions about domain knowledge, how the student's knowledge is modeled, and the nature of the learning process. In the next sections, I identify several distinct categories and discuss each briefly.

2.2.1 Nominal Modeling

Systems that adopt a nominal approach to modeling record only the success or failure of individual problem solving efforts. In other words, the comparison (differential modeling) between the system and student goes no further than performance; a correct answer indicates that the student has acquired expertise. For example, WHY (Stevens, Collins, Goldin, 1982) is a Socratic tutor in which the domain knowledge is represented by a set of hierarchically organized scripts. However, WHY[3] maintains no global student model over

[3] In all fairness, the focus in the WHY tutor is not on diagnosis and student modeling, but on investigating the pedagogic utility of the Socratic method.

the course of a dialog, and so its diagnostic ability is limited to evaluating individual responses.

The Integration Tutor (Kimball, 1982) implements a complex and intricate statistical bookkeeping scheme to model student knowledge in the domain of symbolic integration. Integration problems are solved by transforming them step-by-step from some initial state into the final solution state using a pre-defined set of integration techniques as transformation operators. There are several interesting features in this domain. First, there may be multiple "correct" solutions to a given problem, since there may be more than one sequence of operators[4] leading to a correct solution. Second, the criteria for choosing an appropriate integration technique at any point in the problem solution are not well-defined — in some cases, the only way to know a technique is inappropriate is to try it and fail. However, it is clear that human domain experts are generally able to efficiently discover an optimal solution (i.e. a shortest successful sequence of operators). This leads us to the defining feature of systems based on nominal modelling. Instead of attempting to articulate this problem solving expertise further by modelling actual problem solving reasoning, the Integration Tutor uses a complicated statistical scheme to track the probability that the student will use each integration technique (known as *approach probabilities)* when presented with an instance of any given class of symbolic expression. The system expert's own approach probabilities are based on human expert's solutions to all of the problems known to the system. Remediation and dynamic curriculum adjustment is then based on the comparison of approach probabilities between expert and student.

A similar approach to diagnosis is taken by ACE (Sleeman and Hendley, 1982), a system whose primary focus is the analysis of student's natural language articulation of the reasoning process they used in interpreting nuclear magnetic resonance (NMR) spectra. The student's analysis is compared to a hard-wired expert's analysis of the given NMR spectrum and appropriate comments are returned from some pre-defined set. The student modelling in this system is functionally identical to the WHY tutor in that no global student model or history of performance is maintained over the course of a session.

There are both advantages and disadvantages for the nominal approach. One advantage (compared to other approaches described below) is that the system makes no assumptions

4 The similarity of this problem structure to a classic heuristic search is obvious and interesting. However, in the Integration Tutor, search is not warranted since the set of problems to be worked is fixed and pre-defined.

about the "correct" mental model the student is to adopt — anything that produces the same behavior as the expert is acceptable. For example, in Integration Tutor (Kimball, 1982), there is no single correct way to solve an integration problem; any one of several approaches may lead to the solution. This characteristic is not so much a planned feature of these systems, as an inherent consequence of the fact that the models used to embody expertise in these systems are not useful to human problem solvers. A prime example of such a "black box" expert is SOPHIE-I (Brown, Burton and Bell, 1975; Brown, Burton and DeKleer, 1982), a system for tutoring in the domain of basic electronics. SOPHIE-I is based on numerical models of the circuits presented to the student which, while very efficient for simulation, are clearly not a basis for human analysis. This leads us directly to the disadvantages of the nominal approach, centered about the system's inability to communicate about its problem-solving behavior or, more importantly, to have any insight into the student's low-level problem solving behavior. For instance, in the Integration Tutor, the system solves the problem with a set of hard-wired heuristics which are essentially useless as a way of explaining the reasoning process which leads to a choice of integration approaches. Clancey (1983) relates similar observations in his experiences with GUIDON, a tutor based on the MYCIN (Buchanan and Shortliffe, 1983) medical diagnostic system.

In sum, nominal approaches to student modeling are useful in that they do not make any assumptions about problem-solving specifics— the actual reasoning performed during problem solving. This is especially useful for domains in which there is *more than one* correct solution. However, the failure to model problem solving at a deeper level precludes diagnosing and addressing problems in the student's internal reasoning process. That is, we must represent the student's mental model of the domain if we hope to repair it.

2.2.2 Overlay Models

Whereas nominal models function by simply keeping track of problem-solving success, overlay models attempt to model the student's low-level problem-solving process in order to provide more in-depth diagnostic and explanatory capability. To do this, the overlay approach requires that system expertise be broken into discrete chunks. Individual chunks are marked as "learned" based on inferences drawn from the student's observable problem-solving behavior. In this way, the student's knowledge is viewed as a subset of the expert's knowledge. The specifics of the knowledge representation vary from system to system. For instance, DEBUGGY (Burton, 1982), WUSOR-II (Goldstein, 1982), PROUST (Johnson and Soloway, 1985), and SOPHIE-III (Brown, Burton and DeKleer,

1982) capture domain knowledge as a set of heuristics[5] while other systems like SCHOLAR (Carbonell, 1970) use a semantic network representation. The nature and level of detail of problem solving knowledge varies as well. DEBUGGY's heuristics for place-value arithmetic are at a very low level, addressing primitive actions such as subtracting digits and borrowing during subtraction. In WEST (Burton and Brown, 1979) the heuristics, called issues, are at the more abstract level of mathematical operator combination and gaming strategy.

Overlay modeling operates as follows: a problem is presented, the system uses its "expert" representation to solve the problem, resulting in some sort of "problem-solving trace" detailing the expert's reasoning behavior. This trace serves as the main resource for both diagnosis and explanation. The system compares observable student behavior to the trace in order to determine missing or defective areas; remediation is aimed at modifying student behavior to match expert behavior.

While the overlay approach does allow the system to model and remediate low-level problem-solving behavior, it implicitly demands several underlying assumptions. Since student knowledge is viewed as a *subset* of the system's expert knowledge, there is the assumption that the student initially possesses an initial problem-solving perspective identical to the system's. For example, the central relationship to be communicated in CARDIOLAB (Douglas and Liu, 1989) was "pressure difference, modulated by resistance, causes flow". However, any discussion at this level necessarily assumes that students already have a robust conception of pressure, flow and resistance. In particular, the assumption is that entities like pressure and resistance even exist in the student's current model of the domain. If they do not, then clearly any explanation relying on these terms will be ineffective. A second difficulty arises if the domain allows multiple problem-solving paradigms and the student uses one that is different from the system's. The system will find fault with the student's solution simply because it is based on an alternative (but equally valid) model. For instance, the definition of the "best" move in the WEST game depends heavily on the game-playing strategy being used. WEST provides a unique solution to this dilemma by allowing the system to dynamically select one of several possible expert models representing various pre-defined strategies, to match observed student strategy. The notion of dynamically adjusting the system's perspective to match the student's is an important theme in this paper — it amounts to recognizing that the true mark

[5] It is worth pointing out once again that, as detailed earlier, the implicit claim here is that experts actually use these heuristics as a mental model and, indeed, that these heuristics define expertise.

of a gifted tutor is to be able to go beyond recognizing how a student is wrong to see how the student is "right", by discovering and shifting the problem solving perspective to the student's model of the domain. Finally, overlay modeling assumes that observed errors are strictly content related. However, procedural errors may be caused by *distortions* of correct knowledge instead of gaps in factual knowledge. The DEBUGGY system explores this limitation by augmenting expert knowledge with a set of pre-defined "buggy" heuristics which are distortions of correct heuristics, which the system can insert to hypothesize distortions in existing problem solving knowledge.

In sum, overlay modeling is based on representing the student's knowledge as some subset of the expert's knowledge. While this appears to work reasonably well in simple, factual or procedural domains, it is constrained by its rigid reliance on the expert model as the definition of expertise. Systems like WEST and DEBUGGY explore slight modifications to this paradigm; the next section presents a more radical modification.

2.2.3 Evolutionary models

Strictly speaking, evolutionary models are a special case of the overlay paradigm described above; expert knowledge is characterized as some set of facts or problem solving heuristics and student knowledge is modeled as a subset of that expert knowledge. However, evolutionary models deserve special emphasis because they extend the overlay paradigm in an effort to overcome its most serious shortcoming: the inflexibility associated with defining a single "version" of expertise, embodied in the system's expert model. As noted above, the various strategic perspectives available to the system in WEST are an effort to ameliorate rigidity in a context where there are multiple equally "correct" solutions. The evolutionary paradigm makes an even stronger statement by acknowledging that even in domains with a single "correct" model of expertise, that model is reached by evolution through a number of simpler models. That is, expertise develops over time, and may involve changes to the *structure* as well as the content of knowledge.

In WUSOR-III (Goldstein, 1982), for example, the system guides the learner through the specialization and refinement of problem-solving heuristics. The domain for this system is analytical reasoning, embodied in a game known as "Hunt the Wumpus", which takes place in a warren of interconnected caves. In the game, the student is given certain clues as to the position of the beast and various obstacles and must plot a course based on analysis of these clues. Domain knowledge in this system is based on a set of condition-action rules which are used to interpret the clues. For instance, upon moving into a cave, the student might be told "You hear a squeak". An appropriate rule to apply in this situation might be "If squeak, then add all neighbors of this cave to the set that potentially contain dangerous bats". What makes WUSOR-III interesting is that it represents not only this rule, but both simpler and more elaborate versions of the rule as well. The domain model in this system is a connected graph with condition-action rules at the nodes, and arcs that indicate how these rules relate to eachother. Possible relations cover not only evolution of individual rules (i.e. refinement, specialization, and generalization), but also may indicate analogical relationships between rules and the dependence of certain rules on each other and on pieces of basic game knowledge. For instance, the rules above assume that the student remembers that squeaks indicate that bats are in a neighboring cave. In this way, the domain model represents a continuum of knowledge, organized from very coarse to highly refined, instead of representing only the most refined expert knowledge. Consequently, though student modelling proceeds just as in a typical overlay model, by marking nodes as "learned", the overall state of student knowledge is characterized as a *frontier* that falls somewhere along the continuum from simplistic to expert knowledge. The advantage of this more elaborate representation is obvious: remediation can be in terms of the student's under-developed model instead of the expert's. More importantly, the focus of remediation is now less on the presentation of the expert's knowledge and more on promoting the elaboration of the student's model to the next stage of complexity.

The QUEST (White and Frederiksen, 1986) project provides a more in-depth exploration into the evolution of domain knowledge in the context of electronic troubleshooting. White and Frederiksen based their system on a detailed analysis of the domain in which they characterize the dimensions along which domain models may differ. For instance, the *order* of the model reflects the order of the derivatives used in describing changes to model parameters. That is, a zero-order model deals only with binary conditions (e.g. the presence or absence of voltage), a first-order model addresses linear change, and second-order model handles rates of change. Other dimensions include *type* , which deals with scales of measurement (i.e. qualitative, proportional, and quantitative) and *degree* , which is concerned with the granularity[6] of the model. As in WUSOR-III, the domain expertise in QUEST is embodied in condition-action rules. However, instead of being connected in a single huge graph, as in WUSOR-III, the rules are organized along the dimensions

---

[6] This representation of model granularity can be viewed as a characterization of the basic problem solving perspective — the entities that exist in the model. This is extremely important. As we shall see in the next section, naive human conceptions may not explicate certain entities (e.g. friction) that exist in the expert's model.

defined above. Still, the overall effect is similar: learning is viewed as an evolving progression of mental models, moving from simplest to most complex.

A clear implication of evolutionary student modeling is that the evolution is aimed towards some final state, a single expert's model that represents the pinnacle of problem-solving expertise. This is obviously a carry-over from the work in overlay modeling described earlier in which there was a single "correct" model of system expertise to which student performance was compared. However, White and Frederiksen (1986) conclude their work on QUEST with the observation that true expertise lies not in a single model, but in the coexistence of several complementary models which fall along the above dimensions. In the next section, we will find ample evidence for this claim in that human experts (at least in the domain of motion physics) appear to use several different mental models.

In sum, the evolutionary approach to modeling student progress, though still technically an overlay model, represents a fundamentally different perspective on student modeling and, in fact, on tutoring in general. While the simple overlay paradigm defines a single model of expertise and works to transfer this model to the student, the evolutionary paradigm acknowledges and attempts to model the evolution of mental models and the dependence of new knowledge on old, by defining a succession of domain models and describing how they are inter-related. In this way, evolutionary models like WUSOR-III's genetic graph embody not just a theory of expertise, but also a *theory of curriculum*, which details how that expertise develops. This expanded view of modeling will play an important role later in this paper.

A special class of tutoring systems that should be mentioned are systems that have no student model or, for that matter, no active tutorial component at all. STEAMER (Hollans, Hutchins and Weitzman, 1984), a simulation of a steam propulsion plant, and THINGLAB (Borning, 1981), a simulation of basic motion physics are excellent examples of such passive simulation environments. According to our strict definition earlier, these systems are not Intelligent Tutoring Systems at all, since they have no active pedagogical component. Wenger (1987) has labeled such systems as *knowledge presentation* systems, distinguishing them from full-fledged tutoring systems which he calls *knowledge communication* systems. The underlying premise in these systems is that realistic experience with a domain, via a simulation, is just as valuable, if not more so, than overt tutoring. In actuality, most passive simulation environments are used within some curricular context (e.g. with a lab book) in order to organize and guide student exploration of the simulation.

2.3 Discussion

Figure 2.3 summarizes the discussion in this section.

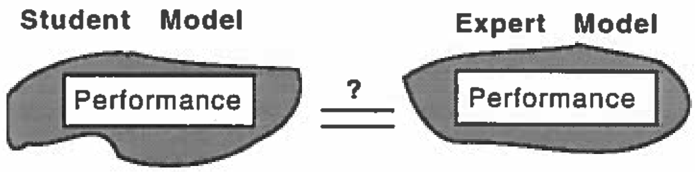

Figure 2.3a: Nominal modeling compares only problem-solving results.

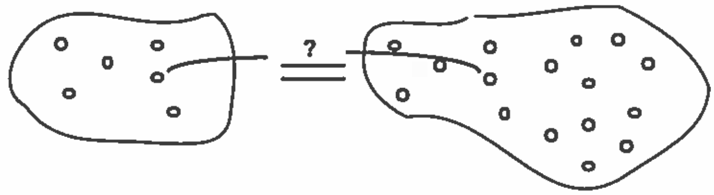

Figure 2.3b: Overlay modeling characterizes expertise as a collection of discrete skills or heuristics. Diagnosis determines which subskills students possess.

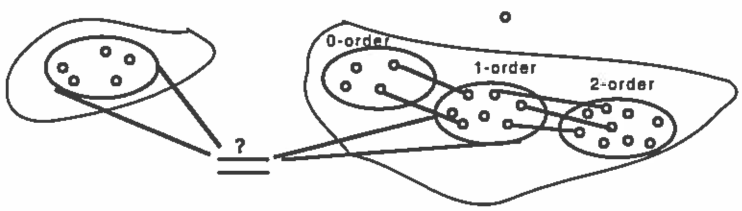

Figure 2.3c: Evolutionary modeling uses a succession of models.

Nominal models are the simplest sort, recording only the success or failure of problem-solving efforts and are typically used when the system's expertise is based on a model that is not psychologically plausible. It is hard to even consider these as student models, since they do not attempt to represent the student's problem-solving knowledge or process in any way. In overlay models, domain expertise is divided up into discrete units or heuristics, and the student's knowledge is modeled by assuming the student has acquired some subset of these heuristics. Problems arise in this paradigm due to the assumption that the student already shares the expert's background and perspective of the domain but has certain factual shortcomings. The evolutionary paradigm extends the overlay notion to accommodate evolving mental models of the domain.

Based on the analysis presented above, I emphasize the following observations.

- Current tutoring systems make a *tabula rasa* assumption. All of the student models discussed above start out "blank", with no representation of what the student may already know before tutoring begins. Taken literally, this amounts to saying that the student initially has no conceptions about the domain at all. Clearly this is not the case.
- Communication is based on the expert model. In systems that have explanatory capabilities, explanation assumes that students have already adopted the skeletal framework of the expert's model and need only to have the system elaborate on this framework. This is the foundational assumption of the overlay paradigm.
- Multiple views of expertise are useful. This is obviously true in domains like integral calculus or the WEST game, where there may be a number of valid approaches. Even in domains like electronics (White and Frederiksen, 1986) and abstract deductions (Goldstein, 1982), having a strong definition of expertise, it is useful to represent multiple evolving models of expertise.

While systems based on evolutionary modeling overcome some of the rigidity associated with the overlay paradigm, they may still fail to make the connection between the system's representation of expertise and the student's current model. That is, the hierarchical succession of models in such systems are based on the expert's view of the domain — there is no guarantee that the "first" model in the succession is anything like the student's initial model. Thus, we are left with several open questions:

1) Do students have naive mental models of the domain that can be formally described, diagnosed interactively, and used, say, as the foundation of evolutionary modeling?
2) Is the number of such conceptions across all students small enough to be tractable.
3) What is the nature of the process by which mental models evolve in humans?

The next section turns to an examination of work exploring mental models of basic physics held by novice problem-solvers in hopes of answering these questions.

3.0 Mental models in reasoning about physical systems

Mental models have proven to be useful tools for rationalizing human reasoning behavior in a large number of domains from open-sea navigation (Hutchins, 1983) to the thermodynamics (Williams, Hollan, and Stevens, 1983; Wiser and Carey, 1983) to motion physics (Clement, 1983; McCloskey, 1983; Larkin, 1983; Forbus, 1983; DiSessa, 1983). They have also helped to explain how people use analogy (Gentner and Gentner, 1983; Young, 1983; Greeno, 1983), an important conceptual "glue" between domains. To narrow the analytical focus, the discussion in this section is constrained to exploring the nature and development of human intuition in the domain of primitive motion physics. This domain is a particularly fruitful domain for several reasons:

- Humans have an intuitive grasp of the domain. They are able to make accurate predictions about the behavior of many everyday scenarios.
- It has been observed (DiSessa, 1983; McCloskey, 1983; Larkin, 1983) that people do very poorly in formal classroom physics. Also, tutoring systems in this domain have had very limited success.
- The domain knowledge, embodied in standard Newtonian principles, is well-defined and tractable.

From the standpoint of tutoring systems, the nature of the naive conceptions held by humans is the central issue of interest in this discussion. As pointed out in Section two, current systems take a *tabula rasa* approach to student modeling: they begin with no information about the student and try to build a student model based on the inferences drawn from the student's problem-solving performance. Most importantly, the student model is constructed from the expert's perspective, that is, by comparison to the expert's model. Clearly, a better approach would be to shift the modeling enterprise towards the student's perspective — which implies that the student's perspective must first be characterized. Another important goal of this section is to gain insight into the development of human mental models from naive beginnings to more advanced conceptions.

3.1 Naive Mental Models

As a classic example of naive misconception in motion physics, consider the following example (from Clement (1983)):

> A rocket is moving along sideways in deep space, with its engine off, from point A to point B. Its engine is fired at point B and left on for several seconds while the rocket travels to some point C. At point C, the engine is switched off. Draw the shape of the rocket's path, labeling points B and C.

The expert physicist's answer and the prevalent incorrect answer are shown in Fig 3.1.

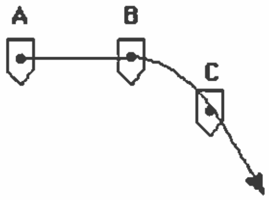

Figure 3.1a: The Expert's solution

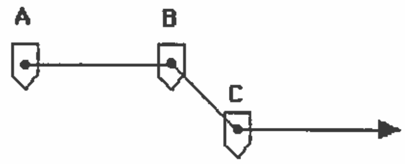

Figure 3.1b: A common Misconceived Solution

Clement uses this and other examples to argue that people have a "Motion implies a Force" preconception, which dictates that all motion implies a force that is causing that motion. To clarify how this assumption results in the incorrect answer shown Figure 3.1 above, consider Figure 3.2, which details the forces on the rocket at various points, according to the "motion implies force" misconception.

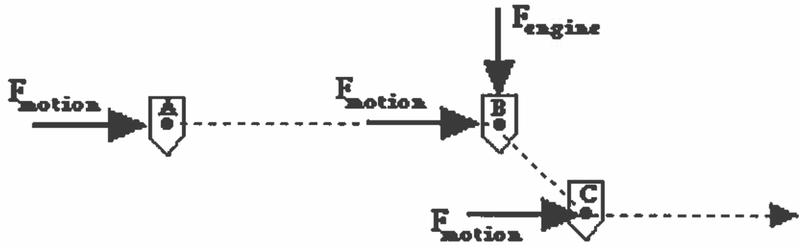

Figure 3.2: Forces on the rocket according to the "Force implies motion" conception.

At point A, there is assumed to be some force on the rocket, causing it to move through space from left to right; between point B and C, the forces of the engine and the sideways motive force combine, clearly dictating a linear diagonal path for the rocket; finally, at point C, the only force on the rocket is once again the sideways motive force, and so the path must necessarily go from left to right again. Specifically, the characteristics of this preconception are as follows:

1. Continuing motion, even at constant velocity, implies the existence of a force in the direction of motion.
2. The invented forces are especially common when motion occurs despite some obvious opposing force. For instance, in the case of an object traveling upward after being thrown into the air, Clement found that subjects often assumed some upward force on the object, which was being opposed by gravity.
3. The imagined forces may be gradually "overcome" or may "build up" to account for the object's change in velocity. Again, a good example in the object thrown into the air: as the imagined upward force is "dies out", the object slows and returns to earth.

McCloskey (1983) describes a number of studies very similar to those performed by Clement. McCloskey's aim is to learn more about the models that people develop through experience with moving objects. Thus, unlike the outer space example above, all of his experiments refer to mundane scenarios: thrown objects, pendulums, objects dropping from a height and so on. To explain the outcome of his experiments, McCloskey suggests that the students' mental model reflects a naive "Impetus theory" very similar to that espoused by early (6th century) philosophers, which essentially states that impelling a body in some direction fills or "charges" it with some a certain unseen quantity, known as impetus, which the object then uses to propel itself in the direction of the initial motion after the motivating force is gone. For instance, in throwing a ball, the thrower hand would "fill" the ball with impetus, which the ball would then draw on to continue its motion after leaving the hand.

The similarity of this work to that of Clement is obvious. Both researchers are clearly describing the same cognitive behavior, though they each posit slightly different mental models to rationalize it. Impetus theory creates an invisible quantity, impetus, as the source of the motive force implied by Clement's "Motion implies Force" model, instead of simply assuming the force exists.

A striking feature of McCloskey's experiments is that he insists on instructing his subjects to "ignore friction and air resistance" in their considerations. He then goes on to hypothesize his Impetus Theory based on the erroneous predictions that students come up with. However, he fails to note that the Impetus model observed in his subjects is not "incorrect" at all, **if** we assume that friction and air resistance are *not* ignored. In other words, I suggest that his subjects found it *impossible* to ignore friction and air resistance as instructed, because these abstract entities are not defined in their primitive mental models and thus have no meaning. That is, naive mental models are not sufficiently articulated to distinguish friction as a separate force. Given that people's naive mental models are derived through experience with a world in which friction is always a factor, Impetus Theory is both sensible and correct (for earth-bound scenarios).

The notion that naive theories of motion correspond directly to real world experience is nicely elaborated in DiSessa's (1983) work, in which he defines a small set of "phenomenological primitives" (P-prims), small scale mental models which explain his subjects' reasoning about a variety of domains. DiSessa's P-prims are, in fact, more general than Clement's and McCloskey's results in that they cover a wide variety of phenomena, including and extending beyond projectile motion. This is illustrated by the following experiment (from DiSessa, 1983):

> Think of vacuum cleaner, whose intake nozzle you hold in your hand. If you put your hand over the nozzle, will the pitch of the sound you hear from the motor go up or down?

Some subjects stated that the pitch would go up, since the added resistance would cause the motor to "work harder". Others felt the pitch would go down, because the interference (resistance) of the hand would slow the motor. In either case, DiSessa points out that subjects' reasoning can be explained using the same mental model, which he calls "Ohm's P-prim". Figure 3.3 illustrates the nature of this primitive.

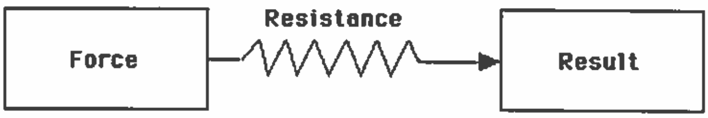

Figure 3.3 (adapted from DiSessa): The Ohm's P-prim

Ohm's P-prim assumes some force acting through a resistance to effect a result. The relation to Ohm's Law[7] is obvious and all of the familiar implications hold (qualitatively): increasing the force increases the result, increasing the resistance decreases the result and so on. Note that Ohm's P-prim subsumes both Clement's "Motion implies a force" model and McCloskey's Impetus Theory.

3.2 The Development of Mental Models

DiSessa goes on to hypothesize how P-prims are developed into a fully articulated and robust mental model of the domain. In this way, he goes beyond the issue of initial representation to address the more challenging question of how mental models are refined over time.

Essentially, the initial p-prims are modified in four ways during the development process:

1) Elimination. Certain p-prims may turn out to be incorrect, or more likely, be subsumed by other generalized p-prims.
2) Abstraction. P-prims that apply to very specific situations can be abstracted and generalized to apply to a broader range of scenarios.
3) Integration. This is basically another form of abstraction. Several p-prims covering specific situations can be integrated into a single more abstract p-prim.
4) Refinement. Based on new insight, a p-prim may be further articulated. A perfect example of this is the case of McCloskey's Impetus Theory. Recall that the theory was correct, but did not recognize resistance as a distinct separable entity. A refinement would separate friction from the general environment.

As a resource for these changes, we also may assume a pool of "textbook knowledge", which could be in the form of lectures, reading, or laboratory experience.

[7] It is important to note the causality implied in the P-prim however. Whereas Ohm's law is a bi-directional and causally neutral relationship, the force in Ohm's P-prim is clearly causing the result.

Larkin (1983) also focuses most of her work on the development of mental models from naive to expert. She begins by characterizing both kinds of knowledge. Novices have a simple model of the domain, which Larkin calls a "naive" representation, which represents only physical objects and immediately apparent influences. This meshes very nicely with the work on naive models described above. On the other hand, experts have what Larkin terms a "physical" representation, which captures the abstract features of a given problem, perhaps even failing to represent certain irrelevant physical details.

Finally, the issues raised by DiSessa and Larkin with respect to development of mental models lead to a broader perspective of curricular design. I suggest that two overall approaches can be identified, as summarized in the table below.

Learning by Example (Bottom-up)	**Laying Down the Law** (Top-down)
Theme: Generalize by exploration and from presented examples	**Theme:** Give the framework of abstract domain laws and try to piece them together and flesh them out.
Instances: ThingLab, CVCK, Steamer, traditional language instruction.	**Instances:** Traditional physics and engineering curricula.
Pros: Strong foundation in everyday experience. Closely connected to students' naive models. Examples are concrete — "relevant and memorable".	**Pros:** Highly directed and efficient. Fast and abstract.
Cons: Burden of generalization and learning is on the student. Potential lack of guidance can lead to mis-generalization.	**Cons:** Not connected to naive model and concrete experience; leads to lack of believability.

Table 3.1: Endpoints on the spectrum of pedagogical philosophy.

Metaphorically speaking, the bottom-up approach is like letting people observe a kitten at play and asking them to induce its internal structure while the top-down approach amounts to giving students a pile of bones muscle diagrams and asking them to deduce how this implies the living behavior of a playful kitten.

3.3 Summary

The common theme running through the work discussed above is that humans develop strong, consistent conceptions of reality as a result of perceiving and experiencing the world. These mental models directly correspond to subject's perception of real world behavior. Conclusions drawn from our discussion are as follows:

- The set of mental models describing all subjects appears to be quite small. That is, there is not a tremendous variation in the primitive models of motion physics. This is reasonable since we all experience roughly the same reality. However, while this observation is promising with respect to modeling student behavior in ITS, it must be accepted cautiously, since DiSessa's work focuses on a narrow domain and certainly does not account for cross-cultural variations in models.

- Initial models that appear to be "wrong" are not necessarily so. They are just overly general in that they do not explicitly represent abstract entities like friction and inertia, assuming, instead, that they are always present in the problem environment. For example, the Impetus theory held by McCloskey's subjects is essentially correct for a world in which friction is always present.

Considering this section in broader terms, some more general observations can be extracted as well:

- There is no such thing as a truly "correct" model. A model, by definition, is simply a framework for understanding and can never be a complete definition of the domain. Even in a "well-understood" domain like physics, there are phenomena outside of our current understanding (model). Thus, the definition of a "correct" model is simply "the one that currently rationalizes the most phenomena". In this way, what is considered the "correct" model can change over time, as evidenced by the evolving models of planetary motion, physics, and the atom. In some cases, like economics, what is the "correct" model is a matter of faith.
- Pedagogically speaking, incorrect models can be just as useful as correct ones. There are two ways in which this can occur. First, there may be limited scenarios in which a more tractable simplified model will do as well as the more complicated "correct" model. For example, Newtonian physics is perfectly adequate under the assumption that we are solving non-relativistic problems. Similarly, the Impetus model is "correct" for scenarios that include friction. A second use for incorrect models is as a didactic tool, to act as a counterpoint in presenting the correct model. Galileo (1967) provides a brilliant example of this technique in his discussion of planetary physics, using the Aristotelian (earth-centered) model as a pedagogical contrast in his argument for a Copernican (sun-centered) view of the cosmos.

Finally, it is worth summarizing this section specifically with respect to the tutorial enterprise. People develop an understanding of the physical world through direct experience and perception of it. This point is re-enforced again and again in the mental models literature: DiSessa's P-prims, McCloskey's "Impetus" theory, and Clement's observations. In all cases, the models posited to explain student behavior contain precisely those influences which are directly perceivable and no more. The elaborate abstractions developed in formal physics — friction coefficients, momentum, inertia and so on — are not present in these models. It is clear, then, that the traditional instructional approach to physics, as embodied in the "law-based" pedagogical strategy defined above, demands an enormous leap of faith for the beginning student[8]. Indeed, it implies a *denial* of the students direct perceptions and experiences in order to replace them with equations and abstract quantities. Only in a handful of cases are students able to understand how these abstractions relate to their everyday experience. In most cases, as indicated in the work cited above, students that have been through formal instruction attempt to incorporate the new terminology and concepts into their existing model. The result is what I call the Frankenstein syndrome: the combination of incongruous elements leading to total confusion.

However, the bottom-up "learning by example" approach is not without its pitfalls either. If students are simply given further instances of the world to observe, there is no guarantee they will refine their mental models at all.

The solution is to support a directed progression, beginning specifically with the students' initial mental model and aimed towards a fully robust "expert's" model. This approach has been partially explored in systems like WUSOR-III (Goldstein, 1982) and QUEST (White and Frederiksen, 1986). However, these systems focus primarily on correct "expert's" representations of the domain. In particular, there is no explicit focus on eliciting initial student models of the domain and structuring the development around these models. An excellent illustration of the utility of such a focus can be seen in our work with the Cardiovascular Construction Kit, a tutoring system based on CardioLab (Douglas and Liu, 1989). In this system, students may construct various simple cardiovascular systems by piecing together various pre-defined components, running the simulator and observing system performance. While students may explore freely at any time, a lab book (embodying, essentially, a curriculum) was developed to guide exploration during initial sessions. Our intuition was to structure the curriculum starting with the simplest possible system, a simple loop, and then move on to more complicated systems involving valves and directed blood flow. The two systems are shown in figure 3.4.

[8] White and Frederiksen (1986) make a similar argument, criticizing traditional physics curricula for their lack of connection to students' "existing intuitions" about the physical world, as well as the emphasis on quantitative versus qualitative reasoning.

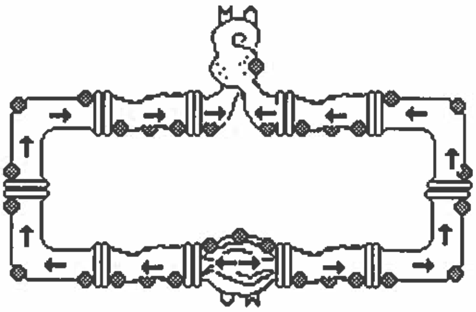

Figure 3.4a: Simplest construction; bi-directional blood flow.

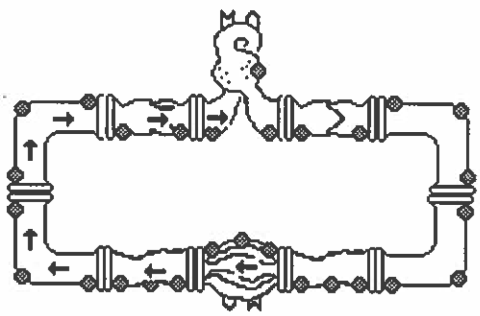

Figure 3.4b: Inserting valves to give uni-directional blood flow.

The result was mass confusion. Despite clear evidence to the contrary (e.g. the flow arrows shown in the figure), student's insisted that blood flow in the valveless system would be uni-directional and clockwise. After extensive protocol experiments, we discovered that students had an initial mental model that did not articulate valves or their function, and in which blood flow was inherently clockwise. Adjusting the curriculum to match these naive expectations, by presenting the valved construction first, then changing it to remove the valves later, dramatically reduced the earlier confusion. This is clear evidence that working to elicit and incorporate student's initial naive conceptions of a domain into the curriculum and student modeling components is important in ITS design.

4.0 The Epistemological Status of Mental Models

The work in both mental models and tutoring systems reviewed in the previous sections is implicitly based on the Physical Symbol System hypothesis: that humans have an internal representation (model) of the world and that intelligent behavior arises from inspecting, planning over, and manipulating that model. In this respect, the claim made by the ITS enterprise is especially strong since it defines expertise to be the possession of a symbolic model, in particular, the one posited as the "expert" model for the tutoring system. This view of knowledge and reasoning seems very natural to anyone (in a Western culture certainly) who has ever practiced introspection. A tradition of communicating, both internally (introspectively) and with others, about action in the real world makes it seem trivially natural to describe our reasoning behavior as a sequence of actions performed on our internal conception of the world.

4.1 The Descriptive View of Mental Models

There has been considerable criticism of the Symbol System hypothesis in recent years by those (Suchman, 1987; Winograd and Flores, 1986) who hold a radically different view of cognition. Though it is far beyond the scope of this paper to provide an in-depth analysis of this subject, it seems very appropriate, especially in light of the lukewarm success of tutoring systems to date, to briefly present this alternative perspective and to consider its impact on the design of tutoring systems.

Perhaps the best way to introduce this alternative perspective is by contrasting it with the Symbol Systems view. The following figure illustrates the difference graphically:

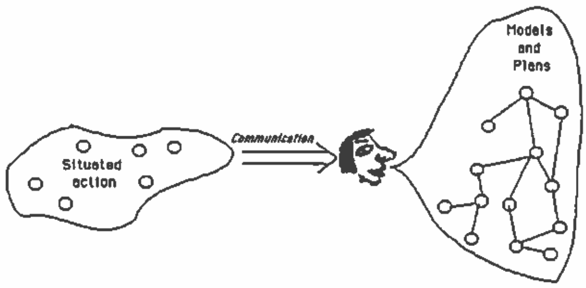

Figure 4.1: Plans and Models as a way for communicating about action.

The essence of the Physical Symbol Systems approach, which I shall call the *generative* view of mental models, is (refer back to figure 2.1) that cognition is based on the manipulation of some internal model of the world. The alternative we now consider, which I will call the *descriptive* view of mental models is, as the above figure indicates, radically different. Under this view, mental models are linguistic abstractions that arise from our particular world view, a view in which communicating about, predicting the behavior of and rationalizing our actions in the real world plays a central role. In other words, mental models do not exist as symbolic mental constructs which act as the generative basis for cognition. Instead, they come into being as a result of our need to rationalize and communicate about our actions and the behavior of the world. Our actions themselves occur in the embedded *situated* flow of everyday activity — they are not based on planning[9] and manipulation of a symbolic mental model. To see this more clearly, consider figure 4.2 below.

[9] We are concerned here with action at a very low level, where it actually occurs. Clearly, "planning" of some sort does play a part in our everyday lives, but at a relatively high level (e.g. "go to the store"), that could never serve as the basis for action.

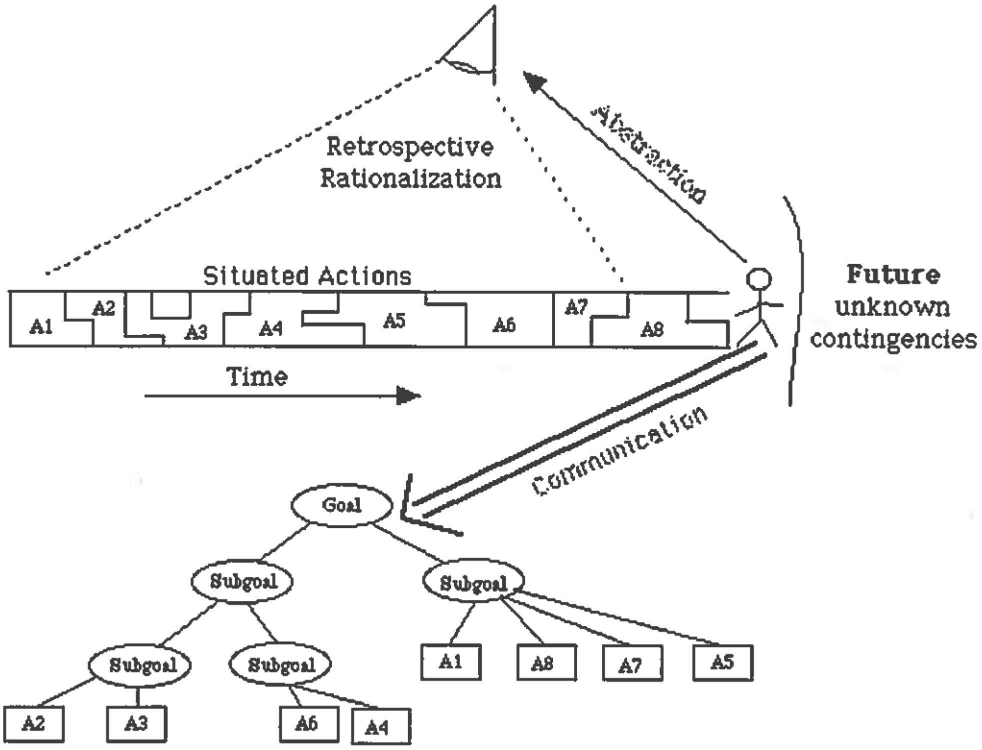

Figure 4.2: The Descriptive view of Plans and Mental Models.

As indicated in Figure 4.2, our everyday actions unfold as a direct means of dealing with contingencies as we encounter them. Only when we transport ourselves to a more abstract perspective in which we can consider and communicate about our action, do plans, goals, and mental models arise as linguistic tools for objectifying and rationalizing action.

There are several clear implications of the descriptive perspective:

1) If mental models and plans are **not** the generative basis for action, then it is clearly possible to act intelligently without any notion of mental model.
2) However, since mental models (and plans) are a linguistic tools for communicating about action, cultures that do not support those tools will find difficulty communicating about action with those that do.

In fact, precisely such evidence exists. Gladwin (1970), an anthropologist, reports the striking difference between the European and Trukese approach to open-sea navigation. The European navigator begins with a goal and a plan, in the form of charts, courses, and careful scientific readings. Any unexpected contingencies require careful recalculation to "repair" a plan that has gone astray. In contrast, the Trukese navigator was reported to start only with a goal and to adjust to changing conditions in an ad hoc fashion. In particular, there is no concept of a "course" or a "plan" for adhering to it. Hutchins (1983) elaborates on Gladwin's observations with an in-depth attempt to discover[10] the mental models used by Micronesian navigators. He is hampered throughout his work by the apparent inability of his subjects to describe their mental processing and conceptualization of the navigation task. In the end, Hutchins relies primarily on *his* observation and rationalization of the navigator's reasoning behavior, to hypothesize a world view and (partial) model for how open-sea navigation is accomplished in Micronesia. Though Hutchins, in his analysis, implicitly supports the view that his subjects do base their reasoning and subsequent actions on some (radically different) internal representation of the world, I believe his work represents some of the strongest possible support for Suchman's (1987) ideas (i.e. the descriptive model of cognition). In particular, it seems clear to me that Hutchins' subject *have* no notion of mental models and therefore lack the linguistic mechanisms for communicating about their actions as rationalized by a mental model. Even Hutchins almost admits it:

> "The tool box of the Western navigator contains scales and compass roses on charts, dividers, sextants and chronometers. These are all A/D and D/A converters. In our tradition, the operations of observation, computation, and interpretation are each a different sort of activity and and are accomplished serially [with respect to some model]. The Micronesian navigator's tool box is in his mind. There are no A/D or D/A converters because all of the computations are analogue. The interpretation of the result (bearing of the reference island, for example) is embedded in the computation (construction of the horizon image) which is itself embedded in the observation (time of day)."

The following points summarize our observations:

- The Trukese navigators were unable to communicate their approach to navigation and, indeed, were reportedly confused by the very notion of doing so. This supports the claim that models and plans about the world are linguistic abstractions, associated with our particular world view, in which every object, place, action, or even mental entity is conceptualized as a mental artifact with certain operations that we can perform on it with a predetermined outcome. The Trukese world view, in contrast, centers on the perspective of being "immersed" in a world which more or less "happens" to them. There is little notion of planning and prediction. As a result, they have not developed linguistic abstractions (i.e. plans, mental models) for discussing such things and, indeed, can not even conceive of them.

[10] This notion of "discovery" emphasizes that an important underlying assumption to all work in mental models is one of *rationality*. That is, that seemingly random behavior is, in fact, not random at all and would make perfect sense if only we could discover an appropriate mental model.

• Despite the apparent lack of mental models, Micronesian navigation is successful. This reenforces the position that mental models are useful for communicating to others our rationale for action in the world, but not crucial to cognition.

4.2 Discussion

The purpose of this section is not to argue for a particular philosophical foundation for cognition. Clearly, determining the true epistemological status of the mental model — whether the Physical Symbol System Hypothesis is true or not — and its relationship to human knowledge is a complex philosophical question that may (and perhaps can) never be resolved. Certainly it is beyond my meager intellectual means. Instead, my goal in this section has been to present a reasonable alternative in an effort to shake the complacent view of symbolic models as not only tools for describing cognition, but as the very embodiment of knowledge in humans and the basis for all intelligent action. In particular, reconsidering PSS as the founding assumption for tutoring systems leads to several extremely stimulating questions with respect to Intelligent Tutoring Systems:

If Physical Symbol Systems are not the embodiment of knowledge and the basis for human reasoning and understanding, then what is? I suggest that this point is entirely academic. In fact, since PSS is the only way, to date, of conceiving of and communicating about knowledge and cognition, we have no other way of describing it. More strongly, the whole point of the descriptive theory of cognition is that ***any*** model we come up with for cognition will be just that — a model, a linguistic abstraction useful for communicating *about* cognition, but never its generative basis.

If we don't have a PSS as the embodiment of knowledge in a tutoring system, then just what is it we're trying to transfer to the student?
This is clearly a problematic issue for tutoring systems. Indeed, I think Suchman would argue that this issue is fatal for ITS. However, as an optimistic young researcher, I believe a solution lies in the literature cited earlier with respect to the development of naive conceptions of reality. In particular, I focus on several observations to be drawn earlier in the paper from the work by DiSessa, McCloskey, Clements and others:

1) Humans have robust, consistent intuitions about the behavior the physical world and the artifacts in it.
2) These conceptions, though perhaps "wrong" with respect to scientific models of behavior, accurately reflect the students' experiences in the world. For example, both the Impetus model observed by McCLoskey and the "Motion implies Force" conception

discussed by Clement are "correct" if we consider that they are based on experience with a world in which friction can never be ignored.

The implication is that (regardless of what knowledge *is*) students acquire correct conceptions of reality through direct experience. The purpose, then, of ITS must be to present such formative experiences. Obviously, this is a clear mandate for systems based on graphical simulation and may help to explain the fundamental attraction of this approach observed in the real world.

Are symbolic models useless then? Absolutely not. The computer, as a physical symbol system, must base its "reasoning" on such models. I suggest that, instead of serving as the embodiment of knowledge and the substance that we seek to transfer to the student, symbolic models be used as mechanisms for rationalizing and describing human behavior. In other words, I propose a shift of perspective, in which symbolic models are moved from their central role as the "complete" characterization of expertise and the basis for all explanation, to a pedagogic role in which they are used as a way of communicating and rationalizing student behavior. The computational implications of this approach are further explored in section 6.

5.0 Summary

The last point in the previous section provides the final foundational brick in my effort to modify our model of tutoring systems and the role of symbolic representations of student knowledge in them. Before going on I summarize the following points, taken from throughout this work:

- Current approaches to student modeling tend to be either weak (superficial) approaches, or center on characterizing the difference between the student's knowledge and the system's "expert" knowledge. This differential modeling approach is problematic several ways. First, it assumes that the student's model is some subset of the expert's. While this allows for differences in content, it will fail if the student's model is entirely different in structure. Second, it relies on the Physical Symbol System Hypothesis in the strongest way, assuming the expert's symbolic model to be the knowledge to be communicated to the student.
- Current approaches take a *tabula rasa* view of student's initial knowledge. That is, the student model is initially "blank" and does not acknowledge students' naive preconceptions about the domain. This leads to a lack of connection to students' real world experience and, consequently, a disjoint learning experience in which students are force-fed the expert's model.

• Research on humans indicates that (a) they do have robust intuitions about the real world, (b) these intuitions are firmly grounded in real experience and (c) can be extremely persistent. This is, to me, strong evidence that they should not be ignored in ITS design.

• People's conceptions of reality can evolve over time, becoming more complex and elaborated as experiences accrue. Thus, tutoring systems should focus on promoting the evolution of knowledge from its current state to some final goal state instead of simply presenting the goal state. Clearly, this involves determining the initial state as well as describing the evolutionary process.

• That the PSS Hypothesis is not necessarily true and that a more neutral approach is to rely on symbolic models for communication about cognition, not as its generative basis.

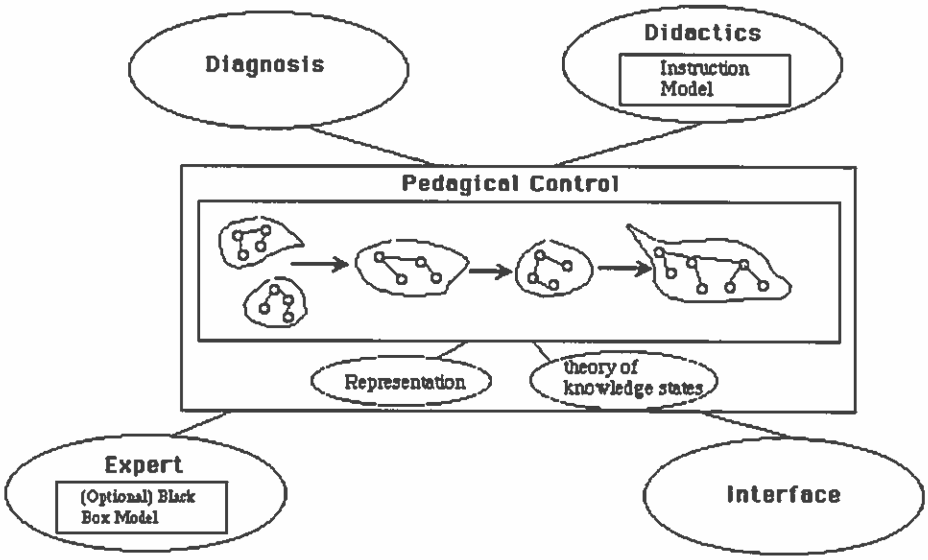

Figure 5.1: A modified framework for ITS design.

Based on these observations, I suggest modifying the abstract framework for ITS design proposed in Wenger as indicated in figure 5.1.

The following aspects of the figure should be highlighted:

• The student model, not the expert model, now plays the central role in the tutoring system. Diagnosis is now a matter of deciding which model in the system's curriculum best describes current student behavior. This is somewhat similar to BUGGY (Burton, 1982) in which the system attempts to rationalize student behavior by exploring deviations in expert behavior.

• Symbolic models are used to describe a curriculum, beginning with known student conceptions of the domain, and focusing on the development of those conceptions to include abstractions and techniques found in experts. At the very least, this implies performing studies like those described in section 3 to characterize students' naive conceptions.

Since the curriculum is based on student's naive conceptions, the incompleteness and erroneous aspects of those conceptions will be reflected in some of the models in the pedagogical module. This can be viewed as a combination of the evolutionary modeling techniques seen in QUEST (White and Frederiksen, 1986) and WUSOR-III (Goldstein, 1982) and the explicit modeling of erroneous procedures seen in BUGGY (Burton, 1982).

- Symbolic models are used for the system's internal reasoning and descriptive purposes only. In particular, explanation must be reconsidered. In current systems, any explanation is necessarily based on the system's "expert" model, with the underlying assumption that the goal is to transfer this particular symbolic model to the student. In backing away from this approach, which very strongly relies on the Physical Symbol Systems Hypothesis, I suggest a more passive mode of "explanation" centered on presenting the student with new problems that highlight inconsistencies rather than attempting to explicate them directly. In other words, the model proposed above implies a pedagogical approach constructed around guided discovery rather than more direct mixed-initiative tutoring.

6.0 Discussion of Future Work

Since this paper is an analytic survey and not a presentation of new research, I've purposefully avoided detailed conjectures about what a "better" ITS would look like. However, after critiquing the landmark classics in ITS history and suggesting shortcomings in their design, it seems appropriate to put it all together by sketching out an ITS that embodies the ideas brought forth in this paper. The following paragraphs briefly elaborate on each component depicted in Figure 5.1, discussing its features and pointing to challenges that will need to be addressed. We begin with a brief overview of system operation to set the stage for the discussion.

Early on, we established that the distinguishing characteristic of Intelligent Tutoring (versus Computer Aided Instruction) is the ability to dynamically assess the student's knowledge state and to adjust the curriculum to address any detected deficiencies or misconceptions. Though the approach taken is slightly shifted, the system I am proposing falls well within this definition. Briefly, the operation of the system, annotated with the components (refer to figure 5.1) involved, is as follows:

1) **Present a problem** (Interface module, Expert module). While the system does represent a pre-defined catalogue of common student (mis)conceptions (Pedagogical module), it must elicit input in order to decide *which* conception best describes the student's current knowledge state. The first such problem-solving task might be pre-defined; subsequent ones are determined as described below.
2) **Analyze problem-solving behavior.** Based on input recorded during the problem-solving process, the system must decide (Diagnostic module, Expert module) which of several symbolic characterizations (Pedagogical module) of commonly-held conceptions best describes student behavior.
3) **Present an appropriate new problem.** Based on its assessment of the student's current knowledge state, the system must choose (Didactic module) another problem which a) is designed to highlight inconsistencies in that model and cause the student to adopt a more advanced conception and b) is designed to provide the system with further diagnostic information.

With this basic operational outline of our tutoring system in mind, we turn to a more detailed discussion of the various components.

6.1 The Pedagogical Module.

This module is the central component of the ITS. All of the other components rely on and are designed around[11] this component. The pedagogical module actually serves two purposes: since we use it to characterize the student's current knowledge state, it is the student model; since the evolving sequence of models represent a theory of curriculum, it acts as a pedagogical model. The pivotal feature of this module is that it is designed from the learner's perspective, not the expert's. For instance, all models in the curriculum, especially at the naive end of the spectrum, are based on conceptions discovered (as described in section 3) in human learners including under-elaborated and "incorrect" models. The overall goal here is to shift from trying to characterize student performance in terms of its deviations from "expert" performance to representing some set of known student conceptions as a basis for rationalizing problem-solving behavior.

There are a number of challenging issues to be addressed in designing this module:

Representation. As always, the issue of how to symbolically represent reality is of central importance. Since this paper deals with the high level design of tutoring

[11] I emphasize again that the conceptual structure proposed here may not even exist in implementation. That is, Pedagogical, Diagnostic and Didactic components may be intertwined so that aspects of all three exist in a single data structure, rule, or procedure.

systems, none of the well-known problems in knowledge representation (c.f. Brachman and Levesque, 1985) are addressed. However, several comments are appropriate. Our goal is to represent student's conceptions of motion physics. Unlike systems like WHY (Stevens, Collins and Goldin, 1982) and SCHOLAR (Carbonell, 1970), which focus on factual knowledge, representing such conceptions implies capturing procedural as well as factual knowledge. Thus, it seems obvious that some combination of semantic net-like formalisms, which are well-suited for representing inter-related facts, and condition-action rules, which capture dynamic procedural knowledge, is required. Finally, our modified approach to tutoring systems does yield one promising implication. As discussed earlier, it is no longer our purpose to transfer the system's symbolic model to the student — such models are used for the system's internal reasoning processes only. Thus, representational details may no longer be quite so crucial. In other words, whatever works for the system is acceptable. Consider, for instance, Clancey's experience with the MYCIN, NEOMYCIN and GUIDON projects (Clancey, 1983). The knowledge representation used in MYCIN was found to be quite adequate for solving problems in its domain. Only when Clancey attempts to use this representation as the basis for explicitly communicating that knowledge to medical students does it fail. Since our system will emphasize accurate graphical simulation (something not applicable in MYCIN's domain) over direct explanation, there is reason to hope that representational details will not be as crucial.

Completeness. Even if we can identify the conceptions held by students and formulate an adequate symbolic representation for them, how can we ever be sure that we've captured *all* possible conceptions that students may have? That is, can we guarantee that our set of models is complete? It seems clear that the answer in no. Despite work by DiSessa and others discussed in section 3, which indicates that the set of naive conceptions is quite small (at least within a specific cultural group), there is no reason to believe that there will not be exceptions. We elaborate on the implications of this observation in our discussion of the diagnostic module below.

Evolutionary relationships. Figure 5.1 implies that the various conceptions are somehow linked, evolving from naive to fully elaborated. But how does one decide how various models are related. Developing such a detailed curricular model represents a major challenge.

6.2 The Diagnostic Module.

The goal of the diagnostic module is to determine the current knowledge state of the learner. In general, this is an extremely difficult task. Unlike a human tutor, who may resort to natural language interaction to gain further information, the only input typically available to the computer-based diagnostic process is the student's observable behavior. From this behavior, which is the end result of the student's internal reasoning, the diagnostic module must posit a hypothetical knowledge state (i.e. a mental model) for the student, which accounts for the observed behavior. To begin to understand the computational implications of this task, consider that it is essentially a machine learning problem. If we view the set of possible mental models as a state space (albeit a very large one), then diagnosis can be recast as an incremental classification problem (Quinlan, 1986). In fact, if we consider the tutorial process outlined at the start of this section as a sort of "dialogue", in which each exchange (i.e. new problem posed) nets the system further information, then the relationship of the diagnostic module to incremental learning systems such as those described by Sammut (1985) and Winston (1975) is obvious.

To give a detailed analysis of the complexity of hypothesis formation in tutoring systems is beyond the scope of this paper. Suffice it to say that, even in the presence of strong simplifying assumptions regarding the nature of the concepts[12] to be learned, the classification problem is computationally intense. In his experiments with the DEBUGGY system (Burton, 1982), which attempts to deduce simple procedural models to rationalize student problem solving behavior for subtraction problems, Burton notes that, even in this simple domain, the computational complexity is so great as to hamper the implementation of DEBUGGY as an on-line interactive system. In addition, this work highlights the information-poor nature of the diagnostic task — there are often multiple hypotheses that explain the observed behavior.

In sum, diagnosis presents a deep and profoundly difficult challenge. However, there is some hope that the approach I've outlined may help reduce the complexity somewhat. I see three ways in which this might happen:

1) **State space reduction.** By working to identify models that characterize student's naive and evolving conceptions (in the form of the curricular theory), the search space is

[12] With respect to tutoring systems, the "concepts" to be learned are mental models. Almost certainly, mental models can be presumed to be more complex than the simple concepts that are the focus of most work in machine learning. Consequently, the complexity of the task can be expected to be orders of magnitude higher.

much reduced. As described in section 3, the set of naive mental models that cover novice problem-solving behavior appear to be quite small. Essentially, the existence of some pre-defined set of "common" naive models changes diagnosis from a hypothesis *formation* to a model *recognition* problem.

2) **Best-first search.** There will almost certainly be cases where the pre-defined set of models fails to rationalize student behavior. At this point, we are faced an open-ended search. However, I suggest that exploring the search space "close to" the pre-defined models will prove to be a powerful search heuristic. In other words, the pre-determined conceptions can act as a starting point in the search for a viable model during diagnosis.

3) **Close is good enough.** Since we do not propose to rely on direct remediation, we will never need to use the model as the basis for an explanation to the student. The primary use of the student model is now to serve as a resource for the didactic module, enabling it to choose the next problem to present.

The last point highlights a pivotal issue: how accurate do we expect diagnosis to be? Obviously, this question is intimately related to the question of how detailed our representation of mental models must be. Is it enough to simply represent models as a set of beliefs about motion physics, or is a full runnable model required? The answer to these questions hinges on the informational requirements of the didactic module.

6.3 The Didactic Module.

Given that we have some assessment of the student's knowledge state and a curricular theory, the job of the didactic model is to choose a new problem or scenario to present to the student designed to "move" the student from the current model to the next, more elaborate model in the curriculum. There are several assumptions underlying the success of this module. First, there is an assumption that it is possible to define some "theory of instruction" which, given the student's current model, can generate new problems that highlight inconsistencies in the student's current model, and encourage evolution of that model. The second assumption is more pedagogical in nature. By avoiding direct explanations and, instead, responding to inconsistencies by presenting a new problem, we are taking what amounts to a Socratic approach. That is, we guide the interaction allowing students to (hopefully) discover their own errors. Anyone that has ever read Plato will understand the potential for frustration in this approach. It is my hope that other aspects of the system will provide the motivation to overcome this difficulty. For instance, casting the tutorial in a gaming context might prove beneficial.

6.4 The Expert Module.

Though we have removed the system's representation of expertise from its central role in the ITS, it still plays an important role. Clearly, the system must still be able to generate correct behavior in the form of an accurate simulation. However, since the emphasis in our system is on passive tutoring by demonstration (i.e. Guided Discovery), there will never be a need for the system to explain its reasoning process. Thus, the system may employ a "black box" expert for maximal efficiency.

6.5 The Interface Module.

A major theme of this paper has been that humans form strong persistent conceptions of motion physics based on direct experience with the physical world. Thus, a central feature of our approach includes a shift of emphasis from direct explanation to realistic experience, from mixed initiative tutoring to guided discovery. Clearly, this shift moves the interface module into a position of crucial importance. In other words, if we intend to directly explain less, we must show more. In their work with STEAMER, Hollan et al. (1984) advance the notion of *conceptual fidelity*, arguing that a simulation should be designed to reflect the abstractions used by expert reasoners, and not physical reality. However, as pointed out earlier, the assumption underlying this sort of claim is that the learner already shares the same conceptual framework as an expert. Thus, a simulation that directly represents abstractions like pressure, implicitly assumes that pressure is something the student understands. While this is perhaps a reasonable assumption for Hollan et al., I feel it is a dangerous one to make for naive subjects. Our work with the CVCK (Douglas and Liu, 1989) emphasizes the pitfalls of conceptual fidelity:

- Subjects were generally not able reason successfully based on given values for pressure, flow, and resistance. We conclude that it is not reasonable to take these abstractions as tutorial primitives.
- Our success improved significantly after we provided visual cues, in the form of small arrows indicating flow direction, in addition to the measured flow values. Subjects often tried to perform all reasoning based only the flow arrows.
- *Physical Fidelity* (i.e. realism) is important. There are two issues here. First, naive learners lack the ability to distinguish between important and non-important features of the simulation. For instance, some subjects became distressed when components did not (visually) connect perfectly, thinking it would cause a "leak" in the system. That is, they interpreted the simulation literally instead of seeing it as an abstraction. More subtly, realism is also important to establish *faith* that the simulation is, in fact, accurate and believable. Students expect to see the ventricle pulse, valves opening and closing,

and blood flowing. Note that it is not important (nor tractable) to be absolutely realistic, only that student's expectations are satisfied.

These observations highlight the importance of creating a realistic learning environment. In my view, both physical and conceptual fidelity must be supported — students must be presented with a realistic simulation, but also be encouraged to think abstractly.

7.0 Conclusion

Intelligent Tutoring Systems represent one of the most challenging incarnations of Artificial Intelligence. The goal of such systems is not only to perform complex reasoning tasks, but to transfer their problem-solving expertise to humans. Due to profound differences in the reasoning process, I believe it is this crucial communication component which is simply not being achieved in current systems, especially those with real-world (as opposed to abstract) domains. In this paper, I have identified two ways in which communication may be impeded.

First, successful communication requires that communicating parties share a similar context, a common background that serves as a foundational resource in communication. It is this background that serves as the interpretive perspective for the interaction. Clearly, a computer can never capture what it is to be human, but perhaps we can work to discover and represent a focused slice of problem-solving context — what the student's domain conceptions are upon entering the tutoring session and how they evolve. Thus, a central theme of this paper has been to suggest a shift from an expert-oriented to a learner-oriented perspective in order to establish a shared communicative context.

The second theme of this paper is somewhat more philosophical in nature, touching on the nature of knowledge and the consequent implications for tutoring systems. Since they are symbolic devices, anything a computer "knows" must be represented as a symbolic structure; "cognition" in a computer amounts to symbol manipulation. Consequently, if the machine works to explain its cognitive processes, its symbolic manipulations to a human, and the expects the human to gain a similar expertise from this, then a monumental underlying assumption is that expertise, including human expertise, can, in fact, be reduced to symbolic processing. This is the claim made by the Physical Symbol System Hypothesis. I feel there is sufficient evidence to cast a shadow of doubt on this hypothesis, as outlined in Section 4 of this paper. Therefore, I suggest that a more cautious approach is to go back to the basics: realistic experience. Focusing on realistic presentation of experience through simulation avoids the entire issue of interpretive frameworks and the nature of cognition.

In sum, the purpose of this area paper is not to indict the philosophical foundations of AI, or even to suggest that previous efforts in ITS have been misguided. Instead, my goal has been to step back from the focus on the systems and methods of ITS and consider the overall notion of knowledge communication in the hope that this broader perspective will provide some insight towards overcoming the ineffectiveness of tutoring systems in all but the simplest domains.

Whether or not my analysis is accurate and will prove useful in designing "new and improved" tutoring systems remains to be seen. Accordingly, I close the paper by summarizing some of the issues and assumptions raised by my research agenda:

- **Characterizing naive conceptions.** Given that we can (following DiSessa, Larkin, etc.) rationalize subject's problem solving behavior using some set of mental models and use them in a tutoring system, will this approach improve tutoring effectiveness? How stable are these naive models across cultures?
- **Representation of conceptions.** How can we represent student's naive conceptions symbolically? What sort of representation will be "adequate" for the system's internal didactic reasoning? Do we need the representation to be runnable models?
- **Development of models.** Can we define a theory of curriculum, a partially ordered succession of models, from simplest to fully elaborated in the domain of motion physics? What about other domain?
- **Didactics.** Will the Socratic-style "Guided Discovery" approach prove effective? Will it be too free-form and frustrating for potentially unmotivated novices? Is it absolutely necessary to provide some explicit explanation? Is it possible, given some characterization of the student's current model, to choose a new problem scenario which will cause the student to elaborate that model successfully?
- **Diagnosis.** Can the shift from a hypothesis formation to a model recognition process reduce the combinatorial complexity of this task. Will it ever be fast enough to be interactive?

Clearly, there is ample room for further research in the directions I have pointed out. Any one of the above issues will almost certainly provide the grist for one or more dissertations. I look forward to contributing to this research effort.

Bibliography†

ANDERSON, J. R., CONRAD, F. G. & CORBETT, A. T. (1989). Skill Acquisition and the LISP Tutor. *Cognitive Science*, Vol. 13, No. 4, 467-507.

ANDERSON, J. R., BOYLE, C. F. & YOST, G. (1985). The Geometry Tutor. Proceedings of the International Joint Conference on Artificial Intelligence, Los Angeles, Ca.

BOBROW, D. G. (. (1984). Qualitive Reasoning about Physical Systems. Cambridge, Mass.: MIT Press.

BONAR, J. G. (1985). Preprogramming: a major source of misconceptions in novice programmers. *Human-Computer Interaction*, 1, (2): 133-161.

BORNING, A. H. (1979). *ThingLab -- A Constraint-Oriented Simulation Laboratory*. Stanford University, Palo Alto, CA.

BORNING, A. H. (1981). The Programming Language Aspects of ThingLab, A Constraint-Oriented Simulation Laboratory. *ACM Transactions on Programming Languages and Systems*, 3, (4): 353-387.

BRACHMAN, R. J. & LEVESQUE, H. J. (1985). Readings in Knowledge Representation. San Mateo, Ca.: Morgan Kaufman Publishers, Inc.

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

BROWN, J. S. & DEKLEER, J. (1983). Assumptions and Ambiguities in Mechanistic Mental Models. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

BROWN, J. S. & VAN LEHN, K. (1980). Repair Theory: a generative theory of bugs in procedural skills. *Cognitive Science*, 4, 379-426.

BROWN, J. S., BURTON, R. R. & BELL, A. G. (1975). SOPHIE: a step towards a reactive learning environment. *Int. Jrnl Man-Machine Studies*, 7, 675-696.

BROWN, J. S.,BURTON, R. R. & DE KLEER, J. (1982). Pedagogical, natural language and knowledge engineering techniques in SOPHIE I, II, and III. In SLEEMAN, D. H. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

BUCHANAN, B. G. & SHORTLIFFE, E. H. (1983). Rule-Based Expert Systems: The MYCIN Experiment. Reading, Mass.: Addison-Wesley.

BURTON, R. R. & BROWN, J. S. (1979). An investigation of computer coaching for informal learning activities. *Int Jrnl of Man-Machine Studies*, 11, 5-24.

BURTON, R. R. (1982). Diagnosing bugs in a simple procedural skill. In SLEEMAN, D. H. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

CARBONELL, J. R. (1970). Ai in CAI: an artificial intelligence approach to computer assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11, (4):190-202.

CHI, M. T., BASSOK, M., LEWIS, M. W., REIMAN, P. & GLASER, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, Vol. 13, No. 2, 145-183.

CLANCEY, W. J. (1983). The epistemology of a rule-based expert system. In BUCHANAN, B. G. & SHORTLIFFE, E. H., Eds, *Rule-based Expert System: The MYCIN Experiment*. Reading, Mass.: Addison-Wesley.

CLEMENT, J. (1983). A Conceptual Model Discussed by Galileo and Used Intuitively by Physics Students. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

† **Bibliographic Note:** A number of the papers listed here appear in compiled collections of papers. Whenever possible, the location of the originally published work is cited here.

DISESSA, A. A. (1983). Phenomenology and Evolution of Intuition. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

DOUGLAS, S. & Liu, Z. Y. (1989). Generating Causal Explanation from a Cardio-Vascular Simulation. *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, Detroit, MI.

FARRELL, R. G., ANDERSON, J. R. & REISER, B. J. (1984). Interactive student modeling in a computer-based LISP tutor. *Proceedings of the Sixth Cognitive Science Society Conference.*, Boulder, Colorado.

FORBUS, K. D. (1983). Qualitative Reasoning about Space and Motion. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

GALILEI, G. (1967). Dialogue Concerning Two Chief World Systems - Ptolemaic and Copernican. Berkeley: University of California Press.

GENESERETH, M. R. (1982). The role of plans in intelligent treaching systems. In SLEEMAN, D. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

GENTNER, D. & GENTNER, D. R. (1983). Flowing waters and teeming crowds: mental models of electricity. In GENTNER, D. & STEVENS, A. L., Eds, *Mentals Models*. Hillsdale, New Jersey: Lawrence Erlbaum Assoc.

GENTNER, D. R. (1979). Toward an intelligent computer tutor. In O'NEIL, H. E., Eds, *Procedures for instructional systems development*. New York: Academic Press.

GLADWIN, T. (1970). East is a big bird. Cambridge: Harvard University Press.

GOETHE, J. W. V. (1840). Theory of Colours. Cambridge, Mass.: The M.I.T. Press.

GOLDSTEIN, I. P. (1982). The genetic graph: a representation for the evolution of procedural knowledge. In SLEEMAN, D. H. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

GREENO, J. G. (1983). Conceptual Entities. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

HOLLAN, J. D. & HUTCHINS, E. L. (1984). Reservations about qualitative models. *Proceedings of the Sixth Cognitive Science Society Conference*, Boulder, Colorado.

HOLLAN, J. D., HUTCHINS, E. L. & WEITZMAN, L. (1984). STEAMER: an interactive inspectable simulation-based training system. *AI Magazine*, 5, (2): 15-27.

HUTCHINS, E. (1983). Understanding Micronesian Navigation. In GENTNER, D. & STEVENS, A. L., Eds, *Mentals Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

JOHNSON, W. L. & SOLOWAY, E. M. (1985). PROUST: an automatic debugger for Pascal programs. *Byte*, 10, (4): 179-190.

KIMBALL, R. (1982). A self-improving tutor for symbolic integration. In SLEEMAN, D. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

LARKIN, J. H. (1983). The role of problem representation in physics. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, New Jersey: Lawrence Erlbaum Assoc.

MATZ, M. (1982). Towards a process model for high school algebra. In SLEEMAN, D. H. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

MCCLOSKEY, M. (1983). Naive Theories of Motion. In GENTNER, D. & STEVENS, A. L., Eds, Mental Models. Hillsdale, NJ: Lawrence Erlbaum Associates.

MILLER, M. L. (1979). A structured planning and debugging environment for elementary programming. *Int Jrnl of Man-Machine Studies*, 1, 79-95.

NEWELL, A. (1980). Physical Symbol Systems. *Cognitive Science*, 4, 135-183.

NWANA, H. S. (1990). Intelligent Tutoring Systems: an overview. *Artificial Intelligence Review*, Vol. 4, No. 4, 251-259.

O'SHEA, T. (1982). A self-improving quadratic tutor. In SLEEMAN, D. H. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

QUINLAN, J. R. (1986). Induction of Decision Trees. *Machine Learning*, 1, 81-106.

REISER, B. J., ANDERSON, J. R. & FARRELL, R. G. (1985). Dynamic student modelling in an intelligent tutor for LISP programming. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Los Angeles.

SLEEMAN, D. H. & HENDLEY, R. J. (1982). ACE: a system which analyses complex explanations. In SLEEMAN, D. H. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

SLEEMAN, D. H. & SMITH, M. J. (1981). Modelling students' problem solving. *Artificial Intelligence*, 16, 171-187.

SLEEMAN, D. H. (1982). Assessing aspects of competence in basic algebra. In SLEEMAN, D. H. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

SLEEMAN, D. H. (1984). An attempt to understand students' understanding of basic algebra. *Cognitive Science*, 8, (4): 387-412.

SLEEMAN, D. H. (1984). Misgeneralization: an explanation of observed mal-rules. *Proceedings of the Sixth Cognitive Science Society Conference*, Boulder, Colorado.

SOLOWAY, E. M. & JOHNSON, W. L. (1984). Remembrance of blunders past: a retrospective on the development of PROUST. *Proceedings of the Sixth Cognitive Science Society Conference*, Boulder, Colorado.

STEVENS, A. L.,COLLINS, A. & GOLDIN, S. (1982). Misconceptions in students' understanding. In SLEEMAN, D. H. & BROWN, J. S., Eds, *Intelligent Tutoring Systems*. London: Academic Press.

SUCHMAN, L. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge, England: Cambridge University Press.

VAN LEHN, K., BALL, W. & KOWALSKI, B. (1989). Non-LIFO Execution of Cognitive Procedures. *Cognitive Science*, Vol. 13, No. 3, 415-467.

WENGER, E. (1987). Artificial Intelligence and Tutoring Systems. Los Altos, CA: Morgan Kaufmann Publishers.

WHITE, B. Y. & FREDERIKSEN, J. R. (1986). Intelligent Tutoring Systems based upon qualitative model evolution. *Proceedings of the National Conference on Artificial Intelligence*, Philadelphia.

WILLIAMS, M. D.,HOLLAN, J. D. & STEVENS, A. L. (1983). Human reasoning about a simple physical system. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, New Jersey: Lawrence Erlbaum Associates

WINOGRAD, T. & FLORES, F. (1986). Understanding Computers and Cognition. Norwood, NJ: Ablex Corporation.

WISER, M. & CAREY, S. (1983). When Heat and Temperature were One. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

YOUNG, R. M. (1983). Surrogates and Mappings: Two Kinds of Conceptual Models for Interactive Devices. In GENTNER, D. & STEVENS, A. L., Eds, *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.