Title
Intelligent Tutoring Systems for Scientific Inquiry Skills

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 8(0)

Authors
Shute, Valerie
Bonar, Jeffrey

Publication Date
1986

Peer reviewed

Intelligent Tutoring Systems for Scientific Inquiry Skills[1]

Valerie Shute and Jeffrey Bonar

Intelligent Tutoring Systems
Learning Research and Development Center
University of Pittsburgh

Abstract

We describe the initial prototypes of several intelligent tutoring systems designed to build students' scientific inquiry skills. These inquiry skills are taught in the context of acquiring knowledge of principles from a microworld that models a specific domain. We have implemented microworlds for microeconomics, electricity, and light refraction. All of the systems are highly interactive; students can pose questions, conduct experiments by manipulating domain specific factors, and record results. Using protocol studies of expert and non-expert learners using these microworlds we identify important inquiry strategies. We have represented these strategies formally, allowing the microworld to detect effective and ineffective inquiry strategies. We conclude with a description of a partially implemented "inquiry coach". This coach will be incorporated into the microworlds and teach the inquiry strategies in the context of the specific microworld domain knowledge.

1. Introduction

How can we help students develop better scientific investigative behaviors? In this paper we describe the key features of several intelligent, interrogatable microworlds under development. These microworlds are designed to support students in learning scientific inquiry skills, as well as in learning the particular domain modeled in the microworld. Our systems make online inquiry tools available to students and generally foster student interrogation. In particular, our systems coach the student in the skills of

[1]This work was supported by the Learning Research and Development Center, supported in part as a research center by funds from the National Institute of Education (NIE), Department of Education. The opinions expressed do not necessarily reflect the position or policy of NIE and no official endorsement should be inferred.

conducting systematic investigations.

We and others in our lab have developed laboratory microworlds for the scientific domains of microeconomics, light refraction, and electrical circuits.[2] In each microworld the student is given the goal of learning the regularities and rules of the domain from observation and discovery. Initially, the student manipulates the environment and observes the results. As new knowledge is acquired, students can form hypotheses and conduct experiments to test those hypotheses. As we describe in detail later, students use the microworlds either to explore, experiment, or do exercises.

Work in our microworld environments is safe and immediate, without the typical complications of laboratory activities in the real world. For example, the microworlds are designed to eliminate much of the extraneous information and phenomena which hinder the discovery of regularities. The microworlds and their underlying simulations allow for instant feedback about the effects of varying relevant variables.

In the process of doing or learning science an individual makes observations, generates hypotheses, tests principles and laws, and predicts from theories. We are concerned with taking widely differing individual exploration strategies, and coaching our students to be more systematic and effective. Students differ on many dimensions and microworlds must be flexible enough to accommodate these differences. For example, prior related knowledge and systematicity of interrogative skills represent two important characteristics underlying successful learning in a microworld environment. Below, we describe work in progress on a exploration skills intelligent tutor.

Our paper is organized as follows. First, we describe a prototype intelligent interrogatable discovery environment: the economics microworld called *Smithtown*. Second, we describe the inquiry tools provided in our microworlds. Students use an interactive notebook, tabular displays, and graphical displays. Students also can propose hypotheses and examine a history of their explorations. Third, we describe the specific inquiry behaviors (skillful and otherwise) found in our protocols of novice and expert learners using *Smithtown*. Fourth, we describe a partially implemented intelligent tutor for the inquiry skills described, focussing our discussion on the diagnostician and the coaching component. Finally, a plan for evaluating the effectiveness of this research approach is discussed.

---

[2] The microeconomics microworld was developed by Valerie Shute and Jamie Schultz. The refraction microworld was developed by Peter Reimann [Reimann, 1986]. The electrical circuits microworld was developed by Jeff Bonar, Joyce Ivill, Cindy Cosic, Leslie Wheeler, Gary Strohm, and Paul Resnick. All these systems have been developed using LOOPS on Xerox Interlisp-D workstations.

2. Overview of the System

Our microworld for microeconomics is a hypothetical town called *Smithtown*. When the student sits down at the computer, s/he is introduced to this simulated town.[3] A series of menus pop up for the student to select from (see Figure 2.1). The first menu contains the markets currently available in *Smithtown*. The student uses this menu to select a good or service to investigate. The second menu is the 'planning menu', containing all possible variables that either may be changed by the student, or that change as a function of something else. The student must state the variables they are interested in investigating. This assists the system in understanding and classifying student activities. Third, a menu with relevant economic indicators for *Smithtown* appears. These indicators include average income, population, weather, consumer preference index, and number of suppliers. Each of these indicators has a system supplied default value (e.g., population = 10,000). The current value for each of these indicators is shown on the screen.

After the student examines and/or modifies the indicators, s/he sees the 'prediction menu'. If desired, the student may use this menu to state the variables and relationships when predicting the outcome of an event. Finally, the student is presented with a 'Things to Do' menu where s/he can see the effect of market manipulations on price, quantity demanded, quantity supplied, surplus, or shortage. Using this information the student may:

1. Adjust the market price or have the computer make a price adjustment,

2. Use the inquiry tools to assist in the investigation (e.g., make a notebook entry of the market data), or

3. Select an experimental framework.

The experimental frameworks let the student manipulate the market in various systematic ways and observe the effects. For example, a student might want to generalize a concept across goods. This would allow him/her to see how widely a concept applies. The framework to accomplish this generalization is: *Change the good, keep the same independent variables.*

As students interact with new subject-matter situations, they compare their observations with their current beliefs and theories. Consequently, these beliefs may be rejected, accepted, modified, or replaced. In the course of this developing knowledge students ask questions, make predictions, make inferences, and generate hypotheses about why certain events occur with systematic regularity. In *Smithtown*, the results of this student interaction with the environment are immediate, dynamic, and recordable. We do this with a set of inquiry tools, described in detail in the following section.

---

[3] Before actually interacting with the system, the student receives a short guidebook to Smithtown, outlining the purpose, set up, and terminology of the system.

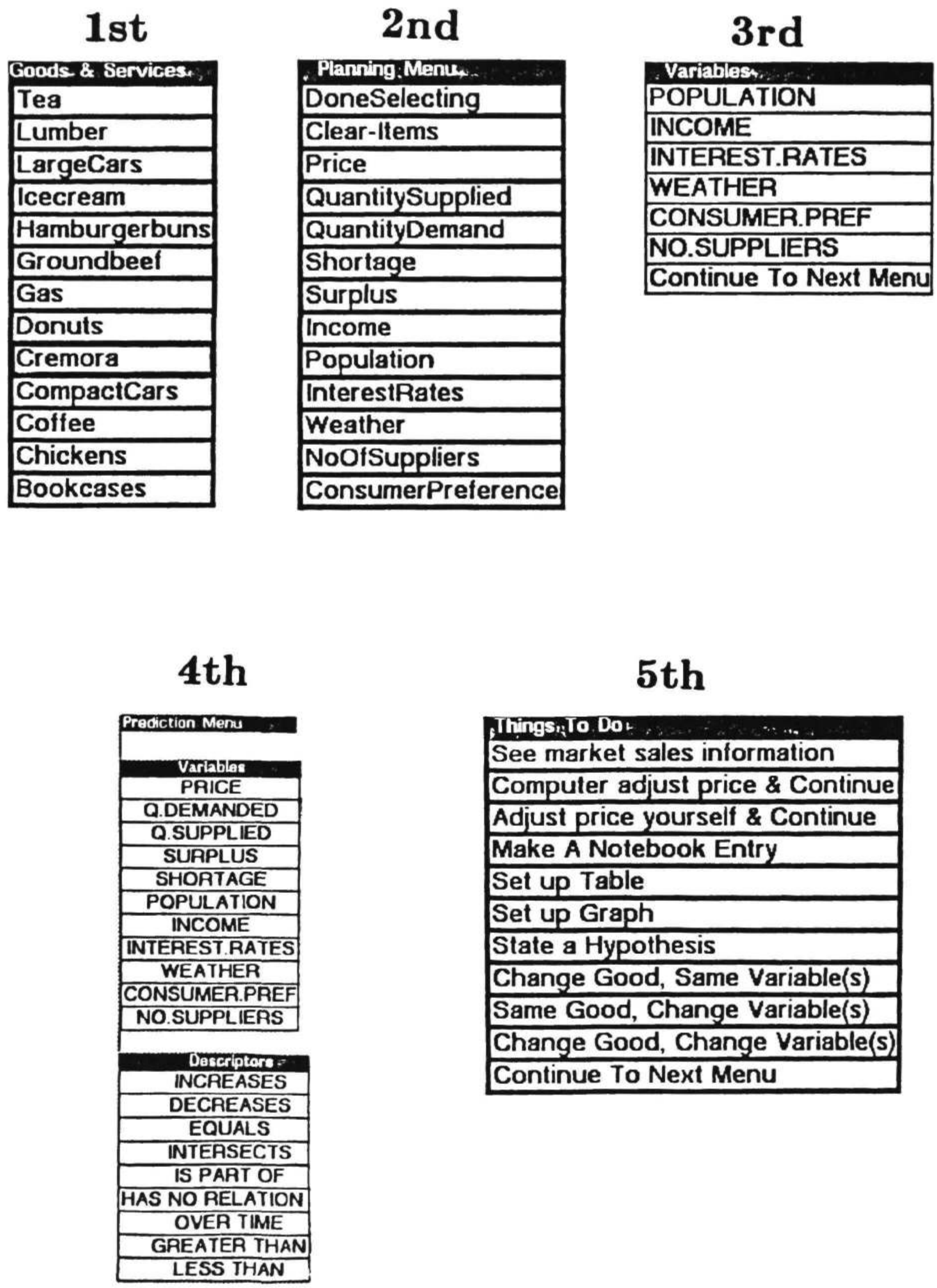

Figure 2-1: FLOW OF MENUS

3. Tools for Systematic Investigations

We have several online tools for scientific investigations in the microworld environments. These include: a *Notebook* for collecting data and observations, a *Table* to organize data from the notebook, a *Graph* utility to plot data, a *Hypothesis menu* to compose relationships among variables, and three *History windows* that allow the students to see a chronological listing of behaviors, data, and concepts learned so far. Each of these will be discussed in turn.

An example of the Notebook is shown in Figure 3.1. The students select variables to record, and current values are automatically put into the Notebook. Once they have collected data in the Notebook, they can elect to isolate some of the variables and put them together into a Table. Tables provide sorting tools for reordering the entries (see Figure 3.2) This is an important tool for simplifying and making sense of raw data in the Notebook. The Graph utility allows a student to plot data collected from their explorations/experiments. This provides an alternative way of viewing relations between variables. An example of a Graph is shown in Figure 3.3.

The Hypothesis menu (Figure 3.4) allows students to make inductions or generalizations of relationships from the data they have collected and organized. There are actually three interconnected menus comprising the Hypothesis menu. First, the *connector menu* includes the items: if, then, as, when, resulting in, and, the. Next, the *variable menu* contains the economic indicator variables used by the system: income, population, quantity demanded, demand, quantity supplied, supply, market price, surplus, shortage, and so on. Finally, the *descriptor menu* describes the types of change: decreases, increases, equals, intersects, is part of, has no relation to, is greater than, and is less than. As students choose words from these menus, the emerging statement appears in the Hypothesis Statement Window.

A pattern matcher analyzes key words from the input and checks whether this matches stored relationships for each targeted concept. If so, the system flags that concept as having been conditionally learned. Otherwise, the student is informed that the statement is not understood.

Three history windows are included in the system. As students continue to interact with the microworld, histories accumulate summarizing the various actions resulting from different explorations and experiments. This summary is maintained in the Student History Window. The Market Data Window keeps a record of all variables and associated values that the student has manipulated. Finally, there is the Goal History Window. This provides a chronological representation of what the student has successfully learned in terms of concepts targeted by the system.

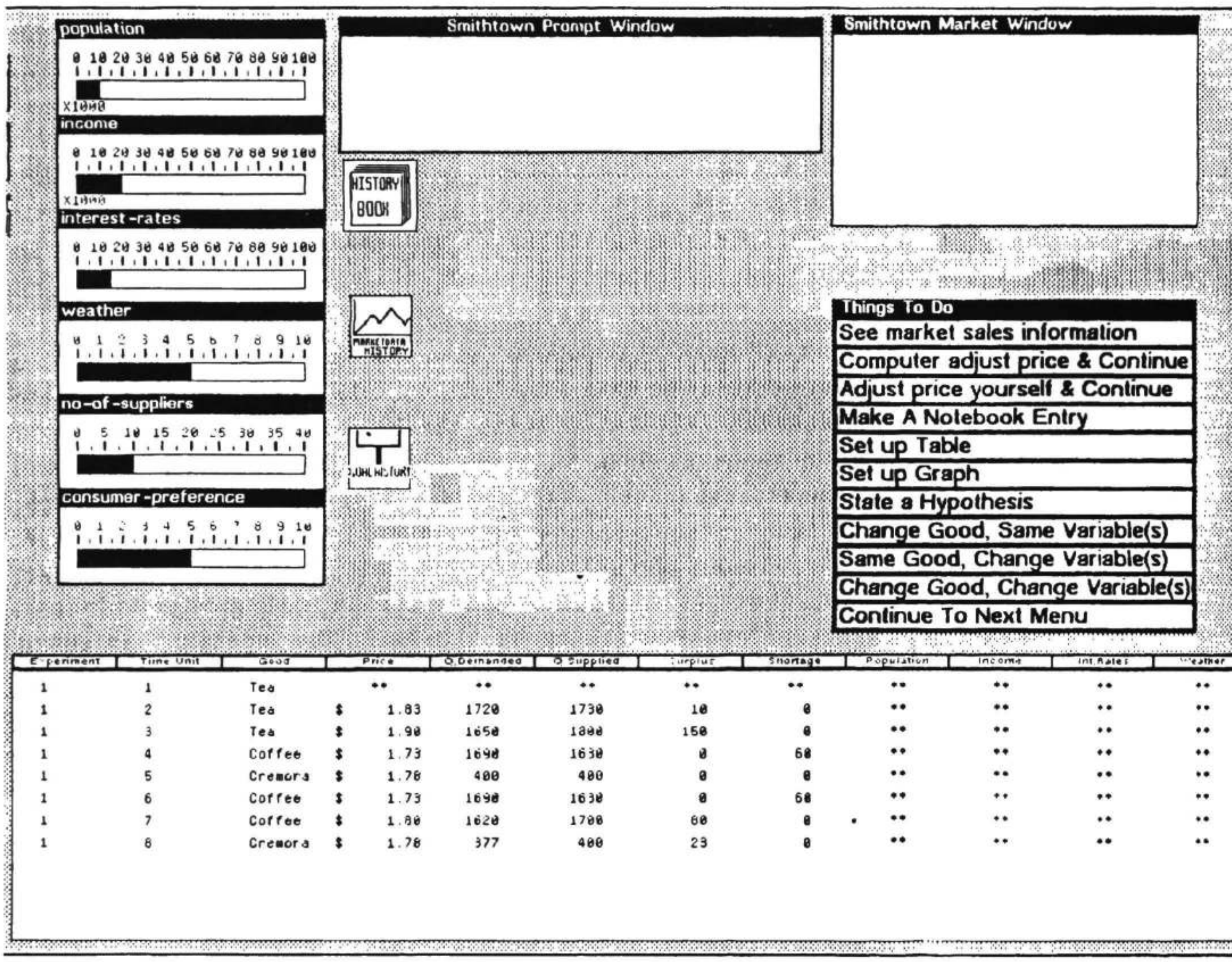

Figure 3-1: NOTEBOOK WITH DATA RECORDED

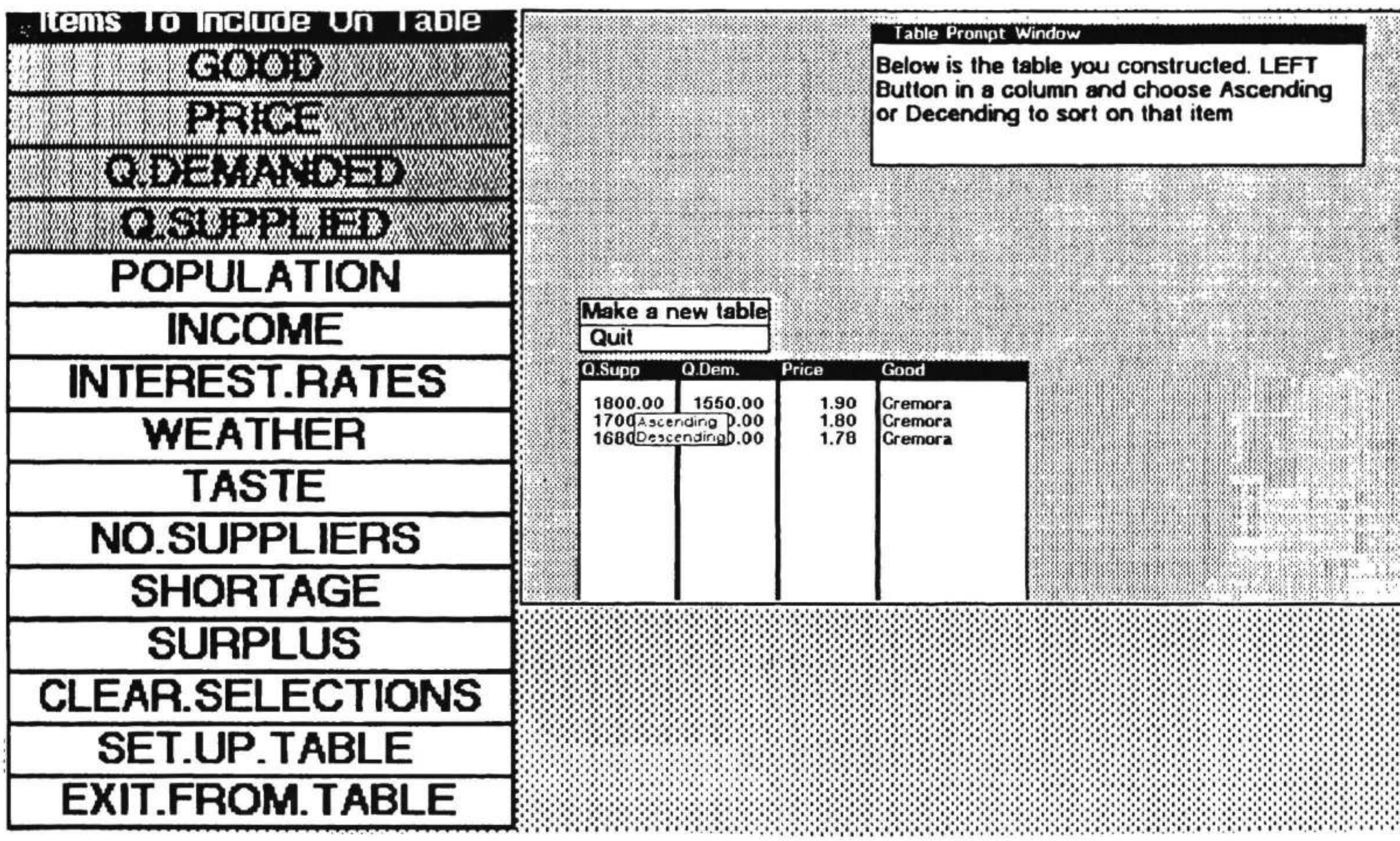

Figure 3-2: TABLE WITH FOUR VARIABLES REPRESENTED

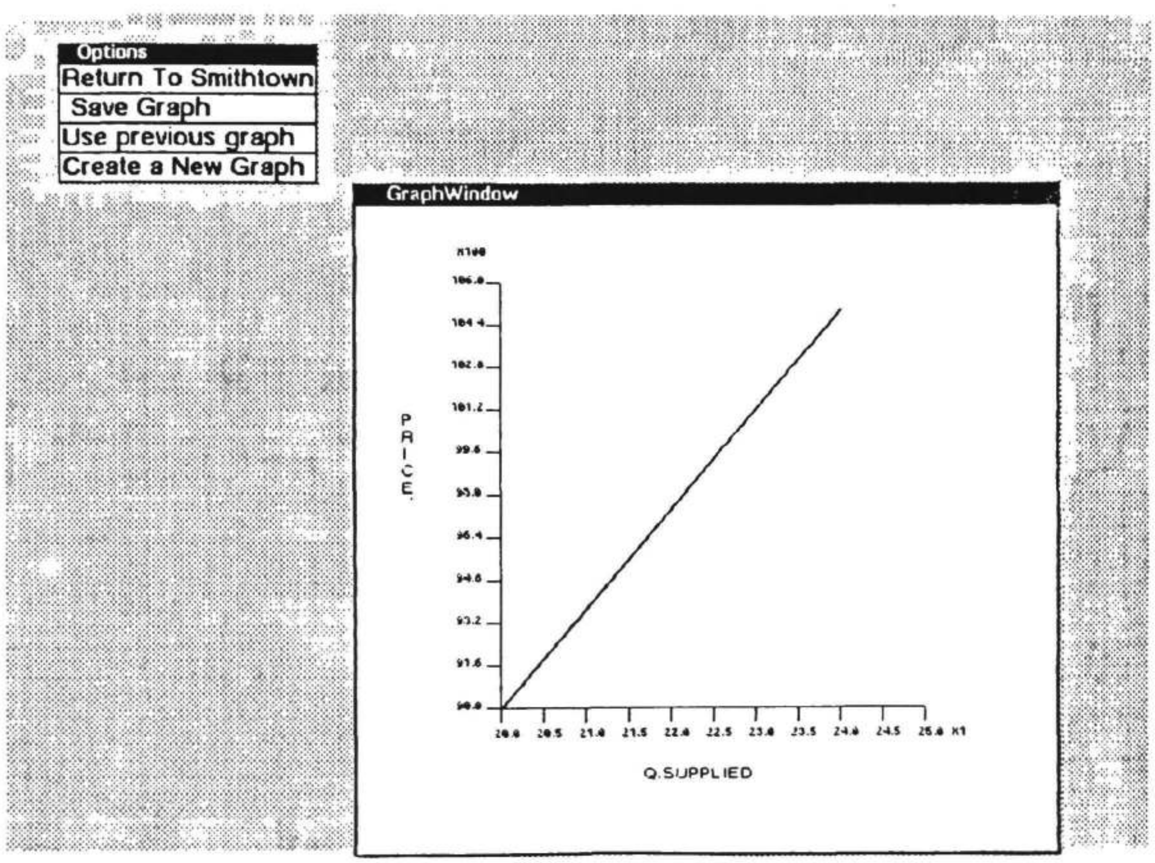

Figure 3-3: GRAPH WITH PRICE AND QUANTITY SUPPLIED PLOTTED

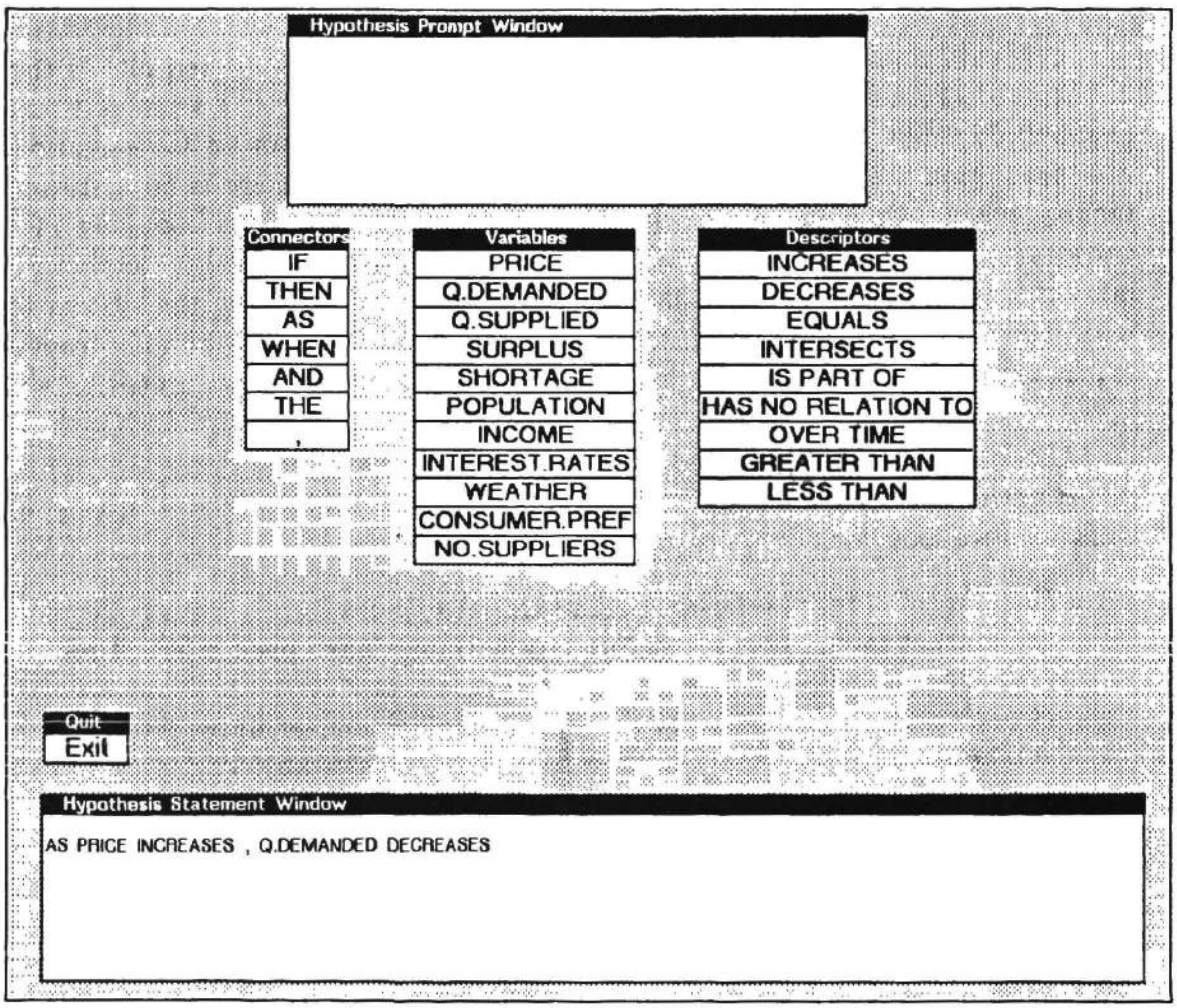

Figure 3-4: MENUS FOR HYPOTHESIS GENERATION

4. Investigative Behaviors

The scientific inquiry tools provide students with means for collecting, organizing, and understanding data from their investigations. Moreover, there are several types of systematic investigations recognized by the system. These include *explorations*: obtaining information from the microworld in order to refine and complete developing hypotheses about the microeconomic concepts; *experiments*: a series of student actions conducted to confirm or differentiate hypotheses; and *exercises*: tests on a previously confirmed hypothesis, perhaps to see the extent or limitations of its application. Experiments carry a specific prediction while explorations do not.

The distinction into the type of investigation being conducted allows the system to determine whether tutorial intervention will occur. That is, a student is considered in *exploration* mode if s/he has no prediction of outcomes from an investigation. Ignoring the 'prediction menu' would result in this classification. There is no assistance from the inquiry coach while in this mode, unless the student remains too long in it (e.g., 20 consecutive actions without making a prediction). If a student predicts an experimental outcome, we consider her/him in *experimental* mode and allow the system to intervene with the student. Finally, if the student is replicating the results from a previous experiment, s/he is classified as being in *exercise* mode and no intervention will occur.

Using the tools provided, there are various dimensions on which student investigative performance can be evaluated. The following is a listing of scientific behaviors we look for:

- *Baseline Data Collection*: For an initial exploration, is data <u>collected</u> from the market in equilibrium, before any variables have been altered?

- *Baseline Data Entry*: For an initial exploration, is data <u>entered</u> into the notebook from the market in equilibrium, before any variables have been altered?

- *Thorough Data Entry*: Does the student make a notebook entry every time a variable has been changed?

- *Relevant Data Collection and Entry*: Does the student enter only those variables that are being actively manipulated or change as a result of a manipulation (i.e., no superfluous information or incomplete recordings)?

- *Generalize a Concept Across All Goods*: Did the student try to generalize an economics principle as it holds for all goods? For example, having learned the law of demand operating in the donut market, does the student attempt to generalize to other goods?

- *Generalize a Concept Across Related Goods*: Did the student try to generalize an economics principle across specifically related goods, i.e., investigating substitute/complementary relationships or component-of relations?

- *Sufficiently Large Change to Variables*: Are the changes made to the variable(s) large enough to detect market effects if there are any?

- *Sufficiently Small Change to Variables*: Are the changes made to the variable(s) small enough to discriminate and refine subtle patterns in the data?

- *Number of Variables Changed*: Did the student change only one variable at a time for comparison and/or recording? This is related to what a student already knows; over time, a student can progressively handle more variables.

- *Isolating Variables in the Table*: Did the student put only a few variables into the table to reduce and make sense of the data?

- *Sorting on Relevant Variables*: Was the sorting option used on relevant variables in the table (e.g., if price was systematically varied, then it should be the sort key.)

- *Plotting Variables*: Did the student use the graph utility to plot potentially meaningful relationships between variables?

- *Saving Graphs*: Was the graph of a significant relationship saved for later comparisons?

- *Superimposing Graphs*: Was there an attempt made to superimpose two graphs to see relationships between functions, like supply and demand, or two demand curves in parallel?

- *Hypothesis Specification*: Were any hypotheses stated as a result of the observed systematicities in the data, either abstracted from the Table or from the Graph?

- *Complexity of Hypotheses*: Is there an increase over time in the chaining of variables when the student generates hypotheses from the menu? As knowledge increases, the number of variables strung together should go from two to more, in progressively more complex relationships.

- *Passively Monitoring the Market-- Computer Price Change*: Did the student allow the computer to make adjustments to the market price of a good and see the ensuing repercussions? For explorations, monitoring is probably the best strategy to follow.

- *Actively Manipulating the Market-- User Price Change*: Did the student more actively make adjustments to the market price of a good and see the market changes? For experiments, manipulating the price oneself is probably more effective.

- *Failure Driven Behaviors*: If a student conducts an experiment testing some specified belief/prediction, the results can lead to a number of different actions. If the experiment is confirmed, there may follow a generalization attempt, or an exercise, both of which test the limitations of the idea. If the experiment is disconfirmed, the student may do one of the following:

  - Re-do the experiment with parameter changes.

  - Ignore the results and go on to something new, or

  - Try a new hypothesis that fits the observed data.

5. An Intelligent Tutor for Inquiry Skills

We are currently implementing an intelligent tutor for the inquiry skills discussed above. Our tutor for *Smithtown* is organized to provide a range of guidance that can be gradually increased or decreased, depending on the characteristics of the learner's performance. At one end of the range is a purely discovery environment. As long as the student is progressing, the microworld will remain a discovery environment. "Progress" is defined as (a) demonstrating appropriate investigative behaviors, and (b) learning the domain concepts at a regular rate. At the other end of the range is a directive environment that explicitly assists the student in using an interrogative skill that is deemed problematic for him/her. For example, the system might instruct a student to enter data in the notebook from the market before altering any variables.

The system is composed of four main components: the knowledge base, the diagnostician, the student model, and the coach. The *knowledge base* includes the targeted elements to be learned, such as the "law of demand", from economics domain knowledge, and "generalization of a concept", from scientific skill knowledge. The *diagnostician* is a set of software *critics* that monitor the student's success in applying the inquiry behaviors and learning the domain concepts. The *student model* is the updated representation of the student's evolving knowledge base, and the *coach* instructs the student based on information provided to it by the diagnostician regarding deficient skills.

Tutor knowledge about the domain concepts is kept in the *knowledge base.* Using the knowledge base, the tutor models whether a student is progressing in the acquisition of domain knowledge. Although a student may be quite efficient in interrogating the environment and performing experiments, s/he may extract only a subset of the relevant concepts for instruction in the particular domain. To accomplish this assessment, the instructional domains for each microworld have been broken down into key concepts that are loosely organized in a bottom-up manner, from simpler to more complex ideas (see Figure 5.1) (see [Bonar, Cunningham, Schultz, 1986] for a description of this approach).

Although the microworld need not constrain the student to learn the concepts in any prescribed order, the conceptual hierarchy provides the basis for a model of student knowledge. Each concept is associated with a rule or relationship among economic variables. For example, the law of demand relates *price* and *quantity demanded* in an inverse relationship. When a student uses the hypothesis menu and generates a valid statement about the underlying variable relationships for a particular concept, then the system flags the concept as having been learned. This is called an *overlay student model* ([Carr and Goldstein, 1977]). The student is then provided with a congratulatory statement by the system, including the proper name of the concept. For instance, the system may respond to a student having just specified the law of demand (i.e., as price increases, quantity demanded decreases), with *"Congratulations, you have just discovered what economists refer to as the Law of Demand. Please note that the converse is also true. That is, as price decreases, quantity demanded increases. Now please conduct an experiment that illustrates this phenomenon."* This request for an experiment guarantees that the student possesses not only declarative knowledge about the concept in question, but also procedural knowledge about how to construct an experiment to demonstrate it.

The *diagnostician* evaluates a student's interrogation of the system. It determines whether a student is proceeding in a systematic, efficient manner. To do so, a comparison is made between the student's actual behavior and optimal behavior (much as was done with the WEST tutor [Burton and Brown, 1982]). For example, a student might consistently alter multiple variables at the beginning of each experiment. This obscures any conclusions that might be drawn from the resulting market conditions. This problem area would then be noted by comparing the student's behavior to the ideal behavior where only one variable is changed.

Each scientific behavior is translated into a *critic* consisting of specific student actions or conditions to be met. A match is attempted and the student model is updated accordingly. For example, the behavior: "Sufficiently Large Change to Variables" addresses the questions: *Are the changes made to the variables sufficiently large enough to detect any market effects if there are any?*

The critic, translated into English, is shown in Figure 5-2.

In conjunction with each critic are actions where the student might go astray. These define the *buggy versions*. In Critic 6, the asterisk shows the line where students

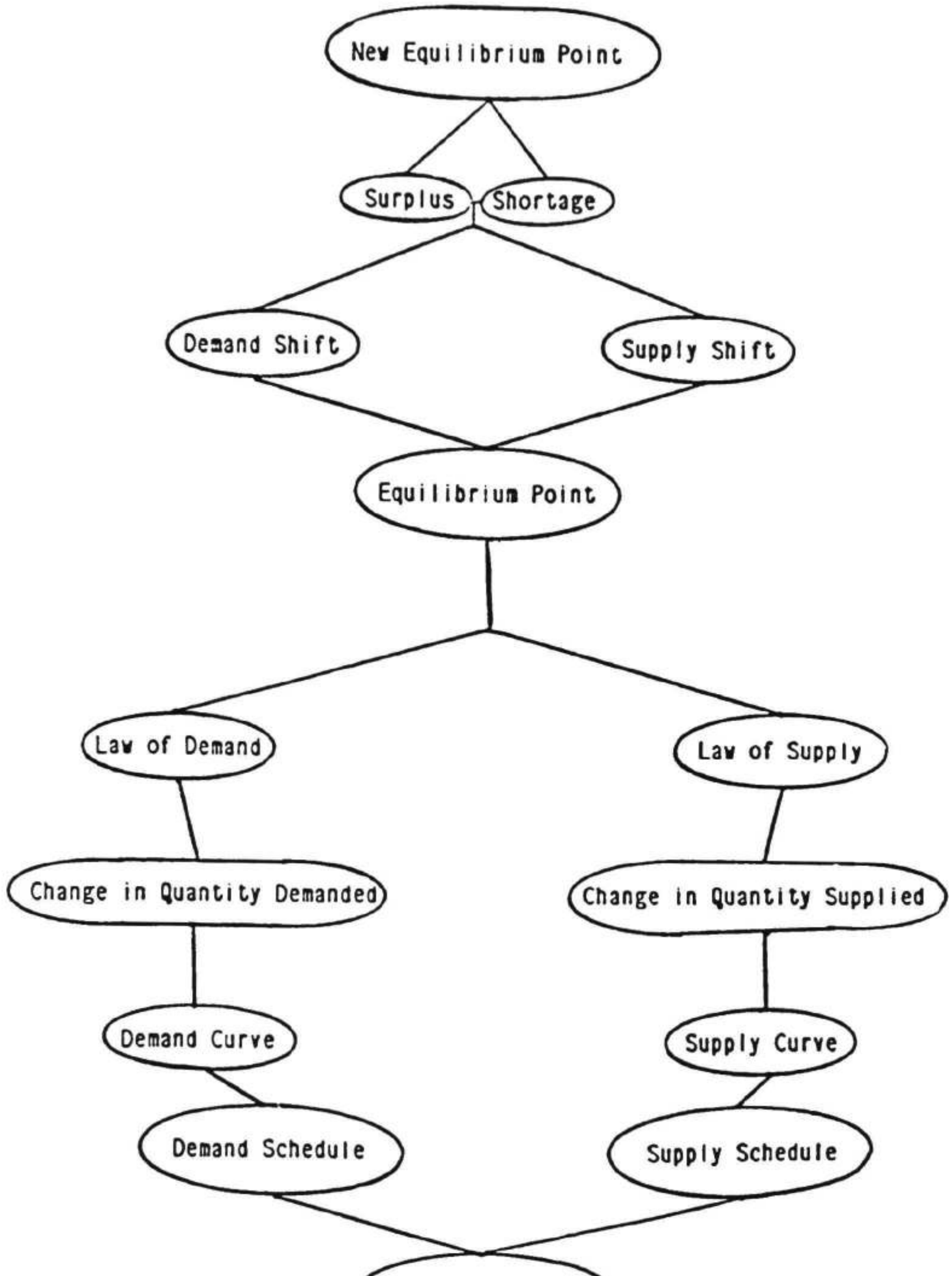

Figure 5-1: HIERARCHY OF SUPPLY AND DEMAND CONCEPTS

Critic 6

IF: For any experiment, AND
for any time number, AND
for any good selected, AND
from variable-menu there is a list of variables changed, AND
* the degree of change is at least 10% of the total range
(maximum minus the minimum values)

THEN: The behavior for 'sufficiently large change to variables' has been demonstrated.

ELSE: Alert the coach for helping the student, depending on the value of the Critic6-counter, which provides information on the number of times the student has erred on this skill.

Figure 5-2: A Critic For "SUFFICIENTLY LARGE CHANGE TO VARIABLES"

Explanation 1 (vague):

If you make a change, you should try to stack the deck. In other words, do something regarding the size of the change so that you'll be able to actually see any effects in the market place if there really are any.

Explanation 2 (analog mapping):

If you entertained the notion that having a lot of hats made a person more self-confident, you would probably compare the self-confidence of a group of people with no hats (or a small number of them) to those having many hats, where "many" was some large enough number to see if your hypothesis was correct. Apply the same logic here.

Explanation 3 (explicit):

You increased the <variable name> by only <amount>. Double it now and check out how the market changes for the values of quantity demanded and quantity supplied.

Figure 5-3: LEVELS OF EXPLICITNESS FOR REMEDIATING CRITIC 6 might err.[3]

The results of the diagnosis are passed to the coach. If there is sufficient evidence of inappropriate student strategies or floundering, the coach will interrupt the student. This coach does not discuss the general inquiry skill directly. Instead, it addresses the issue in the context of the student's current investigation in the microworld. Intervention will occur after the coach receives several corroborations of deficient performance and if the student is in *experiment mode*. "Deficiency" is defined as either not using a behavior when it was appropriate, or using a buggy version of that behavior. A "batting average" is computed for each behavior consisting of the number of times the skill was used divided by the number of times it should have been used. If that number is less than some threshold (such as $< 50\%$), this will prompt a response by the coach. Also, if there are several deficient student behaviors, the tutor can address each, one at a time. This involves a hierarchy of coaching where, for example, the behavior for *Baseline Data Collection* takes precedence over *Save a Graph*, and so on.

To encourage an independence of thinking, the tutorial assistance will progress in terms of explicitness in a 3-part range: (1) Vague initially, with abstract directions and examples for the student, (2) Clearer, with analogies provided to the student, and (3) Very clear, with explicit instructions for conducting a particular experiment. For the example given above, the buggy version of this critic would address the size of the change to the variable. Figure 5-3 shows the three responses.

The preliminary principles calling for interrupting a student's activity are:

- Student weakness must be apparent. What is an "apparent" level will initially be an arbitrary threshhold ratio, to be refined after experimentation.
- Feedback should not be always critical, but also appreciative of any good aspect of a student's behavior.
- Any instruction or guidance should be spaced. That is, at least a gap of two or three events between interruptions.

6. Evaluation of the System

To evaluate whether the system is effective in enhancing students' learning scientific inquiry skills, we will need to look at transfer effects. That is, a student going through one microworld environment learning microeconomic principles should consequently transfer their interrogative skills in the investigation of a new domain, say, geometric optics.

---

[3] Performance data is kept as a proportion of actual use of behavior versus when the behavior should have been used.

Studies are planned to assess the degree to which students apply inductive reasoning skills within and across subject-matter domains. In the context of the various microworlds, a series of near and far transfer situations are being designed, drawing on recent literature on analogical reasoning, cognitive mapping, and rule assessment procedures (e.g., [Cheng & Holyoak, 1986, Gentner, D., 1983, Gick and Holyoak, 1983, Siegler and Klahr, 1982]). In addition, recent information will be considered from artificial intelligence research on learning systems that attempt to improve their own performance in inferential reasoning and problem solving ([Michalski, Carbonell and Mitchell, 1983]).

We are in the process of running subjects (college freshmen) on the system. None of the subjects will have any formal economics training. The major focus will be on how this group interacts with the system, analyzed on two levels: (1) Whether students become more facile across experimental sessions[4] in using the scientific tools/behaviors in the microworld, and (2) If the students actually learned any of the targeted microeconomic concepts. The first question will be answered by protocol analyses of subjects' justifications of actions, as well as detailed histories of student actions kept by the microworld. The second question will be answered from an analysis of the difference scores between pre- and post-tests.[5] Performance with regard to the solution of the complex scenarios will allow us to see individual differences in knowledge representation for economic phenomena. This will compare economic system understanding before and after interaction with the microworld.

7. Concluding Remarks

Research on scientific inquiry learning will contribute to the teaching of subject-matter content with interrogation and inferential skills. Our microworlds provide simulated experimental environments and allow for student interrogation of scientific phenomena by making observations, organizing the data obtained, formulating explanatory hypotheses, and testing experimental predictions. The microworlds incorporate a set of tools which support a student's inquiry process by making these scientific processes visible and manipulatable. Our microworlds allow student patterns of scientific reasoning to be monitored, assessed, and coached. This research should give us information for new methods in science instruction and contribute to knowledge of how to teach higher-order problem-solving skills.

---

[4]Note: the design involves four experimental sessions with the tutor, two hours each session. These occur every other day.

[5]Two types of objective tests have been developed for this purpose: a) multiple choice and short answer test of economic concepts, and b) complex scenarios requiring solution of 'what would happen if...' questions.

8. Acknowledgements

The authors wish to thank Jamie Schultz for his excellent work in developing the *Smithtown* system. Paul Resnick developed an earlier version of the system. Robert Glaser, Peter Reimann, and Kalyani Raghavan have contributed to the research reported here.

9. Bibliography

Bonar, Jeffrey G., Cunningham, Robert, and Schultz, Jamie. (1986, September). An Object Oriented Architecture for Intelligent Tutoring. In *Proceedings of the ACM Conference on Object Oriented Programming Systems, Languages, and Applications.*

Burton, Richard R., and Brown, John Seely. (1982). An Investigation of Computer Coaching for Informal Learning Activities. In Sleeman, Derek and Brown, John Seely (Eds.), *Intelligent Tutoring Systems* (pp. 79-98). London: Academic Press.

Carr, Brian and Goldstein, Ira P. (1977). *Overlays: A Theory of Modelling for Computer Aided Instruction* (AI Memo 406). MIT AI Lab.

Cheng, P.W., & Holyoak, K. J. (1986). Pragmatic reasoning schemas. *Cognitive Psychology*, , .

Gentner, D. (1983). A theoretical framework for analogy. *Cognitive Psychology, 7*, 155-170.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1983). *Machine learning: An artificial intelligence approach.* Palo Alto, CA.: Tioga.

Reimann, Peter. (1986). *REFRACT: A Microworld For Geometrical Optics.* Unpublished manuscript. Unpublished Technical Report.

Siegler, R. S., & Klahr, D. (1982). When do children learn? The relationship between existing knowledge and the acquisition of new knowledge. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 2) (pp. 121-211). Hillsdale, NJ: Erlbaum.