From: AAAI-92 Proceedings. Copyright ©1992, AAAI (www.aaai.org). All rights reserved.

Steps from Explanation Planning to Model Construction Dialogues

Daniel Suthers, Beverly Woolf, & Matthew Cornell

Department of Computer Science
University of Massachusetts
Amherst, Mass. 01003
{suthers, bev, cornell}@cs.umass.edu

Abstract

Human explanatory dialogue is an activity in which participants interactively construct explanatory models of the topic phenomenon. However, current explanation planning technology does not support such dialogue. In this paper we describe contributions in the areas of discourse planning architectures, heuristics for knowledge communication, and user interface design that take steps towards addressing this problem. First, our explanation planning architecture independently applies various constraints on the content and organization of explanation, avoiding the inflexibility and contextual assumptions of schematic discourse plans. Second, certain planning operators simulate a human explainer's efforts to choose and incrementally develop models of the topic phenomenon. Third, dialogue occurs in the medium of a "live information" interface designed to serve as the representational medium through which the activities of the machine and human are coupled. Collectively these contributions facilitate interactive model construction in human-machine dialogue.

The Discrepancy Between Cognitive Theory and AI Technology

This research addresses the automated generation of explanations in the context of interactive dialogues, particularly in educational settings. Explanation is viewed as a conversation in which agents articulate, elaborate on, transform or reject conceptual models of the topic phenomenon. This view is consistent with the recent literature on "situated cognition," which portrays knowledge and learning as partially embedded in the social and physical world [1, 6, 17]. Human dialogue is seen as a process of constructing and sharing representations that are meaningful because the dialogue participants perceive and act upon these representations [3].

However, current technology for automated explanation has yet to support the kind of interactive model construction we envision. Few explanation planners even use multiple models of their topics of explanation, let alone account for interactive model construction in their planning. (McCoy [9] and Paris [15] perform limited selection between multiple "perspectives" during explanation. In the most relevant work to date, White & Frederiksen [22] designed a curriculum of models and transformations between them. The curriculum is traversed in a manner sensitive to the student. However, the student and machine do not explicitly participate in model construction.) Many discourse planners are based on schematic plan operators that fix the possible explanation patterns, and so lack the prerequisite flexibility. Perhaps most crucially, until now explanation planning has neglected the key role of a shared representation of the models under construction which can be examined and manipulated by dialogue participants (whether human or machine).

Steps Towards Model Construction Dialogues

This paper describes small but collectively significant steps towards a technology that is more consistent with the view of dialogue just discussed. Specifically, we present three contributions in the areas of discourse planning architectures, heuristics for knowledge communication, and user interface design that function together to facilitate interactive model construction in human-machine dialogue.

First, our hybrid explanation planning (HEP) architecture separates various constraints that bear on the content and organization of explanations, and represents these constraints as various kinds of operators that dispose the explainer to respond to the current situation in appropriate ways [20, 21]. An architecture that factors the constraints on explanation so that they can operate independently does not suffer from the inflexibility and contextual assumptions of schematic discourse plans.

Second, certain explanation planning operators in our implemented explanation generator are sensitive to the dialogue history and user model to enable the machine to choose between explanatory models and to incrementally elaborate on them in response to user questions. Some operators reason about the tradeoff between the informativeness and the comprehensibility of candidate models for an explanation, enabling the planner to work coherently with knowledge bases that provide multiple models of each object or phenomenon. Other operators select model elaborations and justifications based on the dialogue history, providing a first step towards

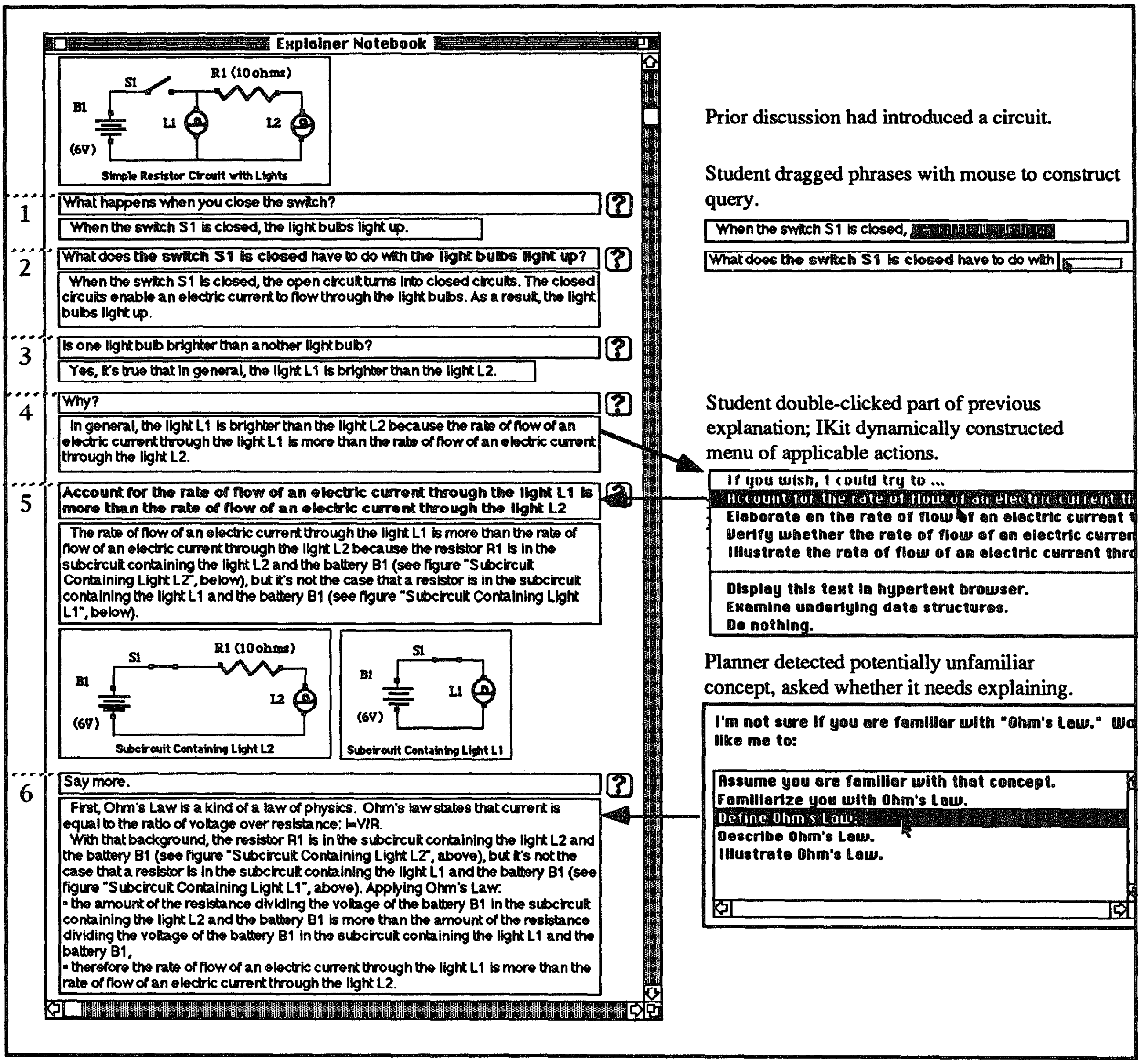

Figure 1: A Question Answering Dialogue with PEG

automating model transformations like those hand-crafted by White & Frederiksen [22].

Third, dialogue occurs in the medium of a "live information" object-oriented direct-manipulation interface, called Information Kit or IKit [4]. The information is "live" in the sense that each conceptually significant segment of displayed information retains the full functionality of the underlying data structure from which it was generated. Thus, displayed information can be interpreted by the machine (as well as humans) in other contexts. This is a step towards shared "representations" [3], as the interface representations are available for perception and manipulation by both human and computer dialogue participants.

The remainder of the paper first describes an example dialogue with our implemented explanation system, to provide a concrete example for subsequent discussion. Then the next three sections describe the contributions made by the HEP architecture, our model selection and presentation operators, and the IKit "live information" interface. Finally, we discuss ways in which this work fails to live up to our own visions.

A Dialogue

Here we provide an example of a question answering dialogue with an implemented explanation system that consists of the HEP architecture plus planning operators plus the IKit interface. We are using the domain of basic electricity and electrical networks.

In our example, the circuit at the top of the "Explainer Notebook" in Figure 1 had just been introduced. "Exchanges" consisting of a question and a response are numbered for convenience. In general, the student's questions were parsed by a natural language interpreter when the "?" icon was clicked. The responses were planned using internal representations (predicate calculus indexed in a frame-like concept hierarchy) before translation into natural language by a template-based generator.

To construct the second question, the student mouse-selected and dragged down two phrases from the explainer's first response (shown in boldface in the second question). Since "live information" remains attached to its internal representations, the machine need not reparse these phrases. Question 5 in Figure 1 was asked by selecting and double clicking part of response 4, from which the explainer dynamically constructed a menu of questions applicable to that response (shown to the right of exchange 5). The question that the student selected was echoed in the query box in exchange 5. Whenever the student mouse-selects an interface object, an internal focus of attention data structure is updated by bringing the selected context into the foreground. Focus of attention is used to identify the referents of phrases such as "the switch" (question 1), and to estimate what models are currently under active consideration by the human participants. While planning the response in exchange 6, the planner noticed that a potentially unfamiliar concept, namely Ohm's Law, was to be used, and asked whether the student wanted a background explanation for it (dialogue to the right of exchange 6). The student asked for a definition, which was provided as the first part of response 6.

This example provides an instance of a model construction dialogue. For example, in exchange 2 the student requested further information on the relationship between the switch closing and the light bulbs lighting. The planner determined that structural and causal relationships were the appropriate relationships for associating these two events, and chose the association that elaborated on the model already in focus in exchange 1 (other associations were available). In exchange 3, the student proposed her own addition to this model, which was verified by the explainer and subsequently elaborated on in exchanges 4-6 in response to follow-up questions. A manipulable representation of dialogue was employed for deictic reference in exchanges 2 and 5, and popup menus were generated in a manner sensitive to the context in which they were invoked.

We now discuss the HEP architecture, the model selection operators, and the IKit interface in more detail.

The HEP Architecture

Human explanation is flexible yet constrained behavior. It is flexible because people construct explanations interactively to meet the needs of the participants. It is constrained because good explanation conforms to the requirements of effective knowledge communication. Flexible yet constrained behavior can be achieved by allowing the various constraints on the content and organization of an explanation to function independently, rather than combining them into schematic operators. Observing that these constraints imply a variety of planning subtasks having distinct information processing requirements, we have designed a hybrid explanation planning architecture, HEP, that matches mechanisms to the requirements of these subtasks, facilitating explicit representation of planning knowledge for each task. No schematic operators are used: all structural relationships are constructed based on reasoning about the functionality of components of the explanation, rather than being prespecified in abstract plans.

Figure 2 illustrates how the architecture matches explanation planning subtasks to mechanisms. The arrows in Figure 2 indicate how the events generated by operators in one task group may result in the scheduling of operators of another task group. (The user model provides information to the other task groups.) Operators from several task groups can be executing simultaneously, coordinated by an agenda mechanism and a common workspace, in a manner similar to [14]. In the following section, we describe some operators that use this architecture to engage in model construction dialogues. The reader is referred to [20, 21] for further detail on the architecture.

The HEP architecture provides a clear correspondence between operators and sources of constraints. Spreading activation propagates assumptions about concept familiarity from concepts a student uses to related concepts, and a simulation model predicts inferences a student might make from an explanation (task 1 in Figure 2). Goal refinement operators specify what kinds of knowledge count as relevant answers to a question (task 2). Preferences select an informative and comprehensible model in a manner sensitive to the user model and dialogue history (task 3). Plan critics specify ways in which an explanation may have to be augmented to facilitate the student's appreciation, comprehension and retention of the explanation (task 4). A graph traversal derives the order in which the parts of the explanation are presented. Constraints on this traversal ensure that the explanation is organized to enhance the communicative functionality of the parts being ordered (task 5). The inferential task for user modeling is currently not implemented: see [23] for one approach. For brevity we omit discussion of the Associative aspect of user modeling: see [21]. The remaining tasks are discussed briefly below.

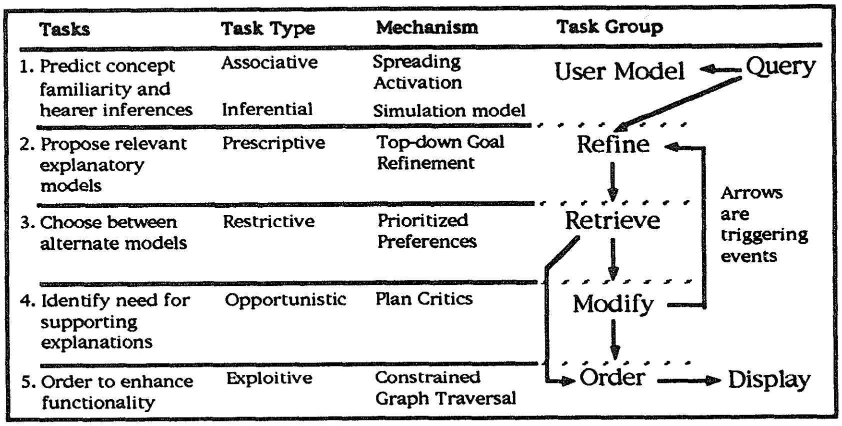

Figure 2: Matching Tasks to Appropriate Mechanisms in HEP

Proposing Relevant Models (Task 2). A *top-down goal refinement* planner [2, 8, 11] is an appropriate mechanism for refining an explanation goal to the communicative acts that achieve it. Such a planner is given a goal and a collection of refinement operators that prescribe what kind of information can satisfy different kinds of goals. Top-down planning records the goal behind each utterance, enabling the planner to recover from failure to communicate and to respond coherently to follow-up questions [11].

Choosing Between Proposed Models (Task 3). Top-down goal refinement in HEP can result in identification of multiple models that compete as relevant responses to a question. Model selection is crucial to a student's comprehension of an explanation and to its appropriateness for his or her purposes [9, 10, 15, 22]. There are a number of criteria bearing on this choice, some of which may be in conflict. For example, the model selected should provide the kind of information requested yet avoid unfamiliar concepts, and it should provide new information yet minimize the amount of new information that needs to be processed. Some of these criteria require examination of the dialogue history and a dynamically changing student model. Choice between alternatives based on conflicting and context-sensitive criteria is a *restrictive* planning task [5], which HEP accomplishes using model choice *preferences*, to be described in the next section.

Identifying the Need for Supporting Explanations (Task 4). The planner may need to include supporting explanations for pedagogical reasons, such as to motivate forthcoming material with an example [16], provide background information that the student will need to understand a subsequent explanation, point out correct inferences and prevent false ones [23], or introduce other topics of interest [7, 16]. These augmentations have been modeled in some planners as preconditions on top-down plan operators [2, 8, 11]. While this approach can plan supporting explanations that can be predicted beforehand, it is limited in a fundamental manner. The need for supporting explanations depends on what the explainer plans to say in the primary explanation and on what the explainer believes the hearer will understand and appreciate without further explanation. Only after a goal has been refined into the models used to satisfy it will one be able to identify the concepts and propositions in the model that require supporting explanations. This is especially true in an explanation system that has multiple models for each topic phenomenon, any one of which could be chosen for a given explanation. In HEP, we use *data-driven plan critics* for supporting explanations. These critics examine the evolving plan and post new goals in a context-sensitive manner. Data-driven operators are appropriate because this is an opportunistic task.

Ordering for Coherence (Task 5). An explanation can involve material from several models, each of which can contain multiple propositions and examples. This material must be ordered for purposes of coherent presentation. A non-arbitrary ordering decision can be made only on the basis of some relevant relationship between the units being ordered. (Examples are given below). HEP is designed to exploit these relationships to enhance the intended communicative functionality of the explanation. Plan operators install ordering *constraints* in response to the appearance of the relevant relationships. Then a *graph traversal* of the plan is conducted that orders propositions in a manner respecting the declared constraints.

Model Construction Operators

In this section, we consider how heuristics for cooperative model construction are embodied in plan operators that propose relevant types of models, choose between alternate models within a type, identify where supporting explanations are needed, and order the selected material to be coherent.

Proposing Relevant Models. Given a question, only certain models will make sense as the basis for a response. Goal refinement operators (see task 2 in Figure 2) are used to propose relevant models, by refining explanatory goals into specifications that, when matched to the knowledge base, access relevant portions of candidate models called "views" [18, 19]. For example, explanation 2 in Figure 1 is based on a specification to find *sequences of propositions* that associate "the switch S1 is closed" with "the light bulbs light up." Explanation 5 in Figure 1 is based on a specification to find *conditions related to the proposition* contrasting the current through the lights by causal and conditional relationships.

Choosing Between Proposed Models. The domain theory can provide alternate models on which to base an explanation. We have identified a number of "preferences" (see task 3 in Figure 2) that choose from among these models in a manner resolving the tradeoff between informativeness and comprehensibility with respect to previous queries and explanations. Preferences are predicates that take two models as arguments and return "<", "=", or ">" to indicate which is preferred, if either. Some of the preferences we have implemented are:

Say-Something-New: Ensures that an explanation is informative in the context of previous dialogue by preferring models that contain some new proposition.

Minimize-Unfamiliar-Concepts: Promotes comprehensibility by counting the number of potentially unfamiliar concepts (according to the associative user model) in each proposed model and preferring the model with the minimal count.

Elaborate-Focal-Models: Helps the questioner relate new knowledge to the model active in his or her working memory by preferring models that elaborate on a model in the focus of attention.

Maximize-Active-Concepts: Keeps dialogue on the same conceptual basis for continuity by preferring models that contain proportionally more concepts that have been used in recent dialogue.

Minimize-New-Propositions: Reduces the amount of new information to be assimilated by counting the number of propositions that have not been stated previously and preferring the model with the minimal count.

Given a set of models, a preference is applied by identifying the subset of models that are most preferred under that preference, restricting subsequent consideration to those models. Preferences are prioritized (the above list is the default ordering) and applied in priority order until a single model remains or until there are no more preferences to apply, in which case the final choice is random.

To illustrate application of these preferences, consider explanation 5 and the second paragraph of explanation 6 in Figure 1. Both of these explanations were based on models that goal refinement identified as relevant for question 5. Applying the preferences, both say something new, but the model underlying explanation 5 contains fewer unfamiliar concepts, so was preferred at that point in the dialogue. Then, goal refinement for question 6 resulted in an identical model specification as for question 5. However, at this point in the dialogue, the preference to say something new will remove the model expressed in explanation 5 from consideration. The model in explanation 6 elaborates on that in explanation 5, so was preferred by the elaboration preference.

Identifying the Need for Supporting Explanations. After refinement and retrieval identified the primary model on which to base explanation 6 of Figure 1, a "background" plan critic identified Ohm's Law as a potentially unfamiliar concept, and verified this with the user. An auxiliary goal was then posted to define Ohm's Law. Other augmentations could be added. For example, clarifications to point out intended inferences and prevent false ones can be made based on reasoning about the inferences a student might make [23].

Ordering for Coherence. Our ordering operators are based on the theoretical assumption that the relevant relationships to exploit in ordering an explanation are those forming gradients along which the student can incrementally incorporate communicated concepts and propositions into his or her current model. We assume that reconstruction is facilitated by attaching new knowledge structures to existing ones, in particular by explaining concepts before they are used to express other concepts, by introducing concepts when they bear some conceptual relationship to other concepts that have already been introduced, and by doing so where possible with propositions expressing such relationships. For example, explanation 2 is ordered to follow the associative path between the topic propositions. In explanation 6, the background definition was placed before the model it supports. [20, 21] discuss these and other ordering operators in detail.

Live Information Display

The above dialogue illustrated how information displayed in the IKit interface can be mouse-selected to obtain a menu of questions or to reference the object in follow-up questions. This functionality reduces the demands on natural language interpreters [12]. Other relevant functionality is also available. For example, the user can incorporate display elements into her own documents (by mouse-dragging across windows), where they will retain their underlying semantic representation. This enables a student to construct a "live" multimedia notebook including the models discussed in a session with the explainer. While retaining the link between information's conceptual structure and its appearance, IKit also distinguishes the two. Appearance is determined by the kind of "displayer" within which the information object is presented. The same information can be displayed under alternate views. For example, a user can mouse-drag a unit of text into a graphic displayer and obtain a picture of the underlying information (provided one is available). Also, once we add *simulation displayers* to IKit, the student will be able to drag a model into such a displayer to obtain a simulation of its behavior. Thus, users can refine the information's appearance after it has been presented. This is desirable because programs cannot always make optimal media choices.

When the switch S1 is closed, the open circuit turns into closed circuits. The closed circuits enable an electric current to flow through the light bulbs. As a result, the light bulbs light up.

In general, the light L1 is brighter than the light L2 because the rate of flow of the electric current through the light L1 is more than the rate of flow of the electric current through the light L2.

First, Ohm's Law is a kind of a law of physics. Ohm's law states that current is equal to the ratio of voltage over resistance: I=V/R.

So, the resistor R1 is in the subcircuit containing the light L2 and the battery B1 (see figure "Subcircuit Containing Light L2", above), but it's not the case that the resistor is in the subcircuit containing the light L1 and the battery B1 (see figure "Subcircuit Containing Light L1", above). Applying Ohm's Law:

- the amount of the resistance dividing the voltage of the battery B1 in the subcircuit containing the light L2 and the battery B1 is more than the amount of the resistance dividing the voltage of the battery B1 in the subcircuit containing the light L1 and the battery B1,
- therefore the rate of flow of the electric current through the light L1 is more than the rate of flow of the electric current through the light L2.

Figure 3: Prototype Current Model Text Display

Currently, we use the display of the dialogue history as the medium on which subsequent activities of the explainer and questioner are focused. However, we believe that we could make better use of IKit as a medium of interaction. When two people interact to come to a shared understanding, they often create external representations such as diagrams and text. These representations make models explicit to be commented on and modified, thereby serving to coordinate the activities of the dialogue participants. For this purpose, the dialogue history is inadequate as a shared representation of the conceptual models under construction. It records the sequence of interactions that led to the current models, but does not make these models explicit. If the "live information" interface is to realize its full potential as a medium for coordinating the activities of an automated explainer and a human user, both the machine and the human should be able to make their models explicit. Towards this end, we are experimenting with a "current model" displayer (Figure 3). The model underlying this display is updated by merging the current explanation plan with the previous contents of the current model. Elaboration models replace their simplified counterparts during the merge. Text plan annotations are also merged, and the result is run through the usual text generator to display the machine's version of the current model. For example, Figure 3 shows the textual portion of the current model displayed at the end of the dialogue of Figure 1. The user can select and query this representation in the same manner as the dialogue history. However, we currently lack the crucial ability for the user to edit and add to this model. Once such capabilities are added, the current model display could serve the same function as other external representations used in dialogue, including dialogue between *multiple users* of the system, as well as between users and machine.

Conclusions and Further Research

We have discussed three technological contributions towards supporting interactive model construction dialogues: the separation of various constraints on the content and organization of explanation; sensitivity of operators to the dialogue history in a manner that simulates a human explainer's efforts to support cooperative model construction; and a live information interface that serves as the representational medium through which the activities of the machine and human are coupled. These contributions synergize as follows. An explanation generator based on independently applied constraints rather than schematic plans provides the flexibility necessary for explanations to emerge out of the requirements of the dialogue. The elements of the interface representation of the explanation can be fully annotated with the concepts and intentions behind each element. This annotated display then serves as a medium through which the activities of the machine and human are coordinated. The human participant's subsequent queries and assertions can unambiguously reference active model components through direct manipulation of those components, and the machine's model selection operators elaborate on the same components in subsequent explanations.

The current system is limited in various ways that could be addressed with further research. HEP does not provide for curriculum planning or goal-oriented tutorial behavior over extended interactions. Cawsey [2] has shown how to model teaching exchanges using higher-level top-down refinement rules, an approach that would not be difficult to add to the architecture. However, an exchange-level planner still lacks a sense of curriculum and long-term tutorial strategy. HEP can be embedded in a tutor with such capabilities, for example that of [13]. Modification of the planning operators to be sensitive to the tutor's long-term goals would be necessary. As discussed above, the use of the interface to couple the activities of human and machine could be improved. Also, the natural language parser is too weak to interpret arbitrary user additions to the current model display, and there are information consistency issues to resolve concerning when user editing invalidates the pre-existing underlying model. More fundamentally, a tighter coupling would be obtained if the machine was designed to reason *with* the interface representations, rather than merely reasoning *about* them in separate, internal data structures. Finally, as the technology and implementation become sufficiently robust for applications, we will need to evaluate the effectiveness of explanation operators and interface design in teaching actual students.

Acknowledgments

The authors appreciate comments on prior drafts from Klaus Schultz, David Skalak and the anonymous reviewers. Support was received from the National Science Foundation under grant number MDR8751362 and from External Research, Apple Computer, Inc., Cupertino CA.

References

[1] J. Brown, A. Collins & P. Duguid: Situated Cognition and the Culture of Learning. *Educational Researcher* **18:** 32-42, 1988.

[2] A. Cawsey: Generating interactive explanations. *Proceedings of the Ninth National Conference on Artificial Intelligence*, pp. 86-91. Anaheim, California, 1988.

[3] W. J. Clancey: Situated Cognition: Stepping out of Representational Flatland. *AI Communications - The European Journal on Artificial Intelligence*. 4(2/3): 109-112, 1991.

[4] M. Cornell, B. Woolf & D. Suthers (submitted): Using ``live information" in a multimedia framework. *Intelligent Multimedia Interfaces*, Edited by Mark Maybury.

[5] E. Hovy: Pragmatics and natural language generation. *Artificial Intelligence*. 43(2): 153-197, 1990.

[6] J. Lave: *Cognition in Practice*. Cambridge University Press, Cambridge, 1988.

[7] J. Lester & P. Porter: A revision-based model of instructional multi-paragraph discourse production. *Proc. 13th Cognitive Science*. Chicago, pp. 796-800, 1991.

[8] M. Maybury: *Planning multisentential english text using communicative acts*. Unpublished doctoral dissertation, Cambridge University, 1990.

[9] K. McCoy. Generating context-sensitive responses to object-related misconceptions. *Artificial Intelligence*. 41(2): 157-195, 1989.

[10] K. McKeown, M. Wish & K. Matthews: Tailoring explanations for the user. *Proc. 9th IJCAI*. Los Angeles, California, pp. 794-798, 1985.

[11] J. Moore & W. Swartout: A Reactive Approach to Explanation. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*. Detroit, Michigan, 1989.

[12] J. Moore & W. Swartout: Pointing: A way toward explanation dialogue. Proc. 10th National Conference on Artificial Intelligence, Los Angeles, August 1990.

[13] T. Murray, & B. Woolf: Encoding domain and tutoring knowledge via a tutor construction kit. *Proc. 10th AAAI*, San Jose, CA, 1991.

[14] S. Nirenburg, V. Lesser, & E. Nyberg: Controlling a language generation planner. *Proc. 11th Int. Joint Conf. on Artificial Intelligence*, Detroit, pages 1524-1530, 1989.

[15] C. Paris Combining Discourse Strategies to Generate Descriptions to Users along the Naive/Expert Spectrum. P*roceedings of the 10th International Joint Conference on Artifiical Intelligence*, August, Milan, Italy, pp. 626-632, 1987.

[16] E. Rissland (Michener): Understanding understanding mathematics. *Cognitive Science*. 2(4), 1978.

[17] L. Suchman: *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge Press, Cambridge, 1987.

[18] A. Souther, L. Acker, J. Lester & B. Porter: Using view types to generate explanations in intelligent tutoring systems. *Proc. Cognitive Science Conf.*, Montreal, 1989.

[19] D. Suthers: Providing multiple views of reasoning for explanation. Proc. ITS-88, Montreal, 1988.

[20] D. Suthers: A task-appropriate hybrid architecture for explanation. *Computational Intelligence*. 7(4), 1991.

[21] D. Suthers. *An analysis of explanation and implications for the design of explanation planners*. Ph.D. dissertation, University of Massachusetts, Amherst, 1992.

[22] B. White & J. Frederiksen: Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence*. 42(1), 99-157, 1990.

[23] I. Zukerman: A predictive approach for the generation of rhetorical devices. *Computational Intelligence*. 6:25-40, 1990.