Towards Goal-Driven Integration of Explanation and Action

David B. Leake

Computer Science Department, Indiana University
Lindley Hall 215, Bloomington, IN 47405

leake@cs.indiana.edu

812–855–9756

May 17, 1995

To appear in Ram, A. and Leake, D., Eds., *Goal-Driven Learning.* MIT Press/Bradford Books. In press.

1 Introduction

In order to learn effectively in complex situations, learners must be able to determine the features of those situations that are relevant to their goals and the reasons that they are relevant. Explanations can play a key role in making that determination. For example, in explanation-based generalization, explanations are used to identify operational features of a training example that are relevant to membership in a target concept (Mitchell et al., 1986); in explanation-based story understanding, explanations determine the features of a situation that are important for establishing coherence with prior knowledge and generating accurate predictions to facilitate future understanding (DeJong, 1986; Leake, 1992; Mooney, 1990; Ram, 1993; Schank, 1986; Wilensky, 1983); in planning and problem-solving, explanations are used to identify the sources of problems and successes, to repair current problems, exploit successful solutions, and avoid problems in the future (Birnbaum et al., 1990; Hammond, 1989; Leake, 1992; Minton, 1988); in multistrategy learning, explanations are used to guide selection of learning methods (Hunter, 1990b; Hunter, 1990a; Ram and Cox, 1994). In addition, explanations are also often used by learning systems to guide selection of appropriate indices for storing what is learned (Barletta and Mark, 1988; Hammond, 1989; Leake and Owens, 1986; Ram, 1993; Schank, 1982). Thus explanations can guide a learner's decisions of what to learn, which learning methods to apply, and how the results should be stored. These tasks have different information requirements, which in turn place different requirements on the information that explanations must provide in order to be useful. Consequently, the process that generates explanations must be sensitive to the goals that explanation serves. This chapter examines how the problem of generating useful explanations can be addressed by goal-driven learning systems in real-world domains.

Generating useful explanations is a difficult problem for several reasons. First, it may not be clear which targets to choose as the focus of explanation (e.g., (Ajjanagadde, 1991; Leake, 1992; Schank, 1986)). Even if an appropriate choice is made, generating a valid explanation for the chosen target may be difficult because of missing knowledge or incomplete or intractable domain theories (e.g., (Chien, 1989; Leake, 1992; Mitchell et al., 1986; Rajamoney, 1993)). In addition, there is no guarantee that the resulting explanation will actually be useful to the performance system (e.g., (Keller, 1988; Leake, 1991a; Leake, 1992; Minton, 1988)). Consequently, a model of real-world explanation must address the questions of *when to explain*, *what to explain*, and *how to focus explanation generation* (Leake, 1992). This chapter addresses those questions within the goal-driven learning framework described in Chapter 1 of this volume. In so doing, it provides one set of answers to the goal-driven learning questions of *when learning is needed, what should be learned,* and *how to guide the learning process.*

The chapter focuses on the problem of abductive explanation (e.g., (Pierce, 1948; Josephson and Josephson, 1994)), and in particular on the role of explanation in guiding and facilitating goal-driven learning. It advances that this role requires the explanation process to be fully integrated with other processing and allowed to interact with other tasks, goals, and the external environment. First, situating the learning process in the context of the larger environment and the reasoner's capabilities for action makes explanation more effective, by allowing multistrategy explanation generation. Second, the interaction between explanation and other processes makes it possible to dynamically refocus reasoning, action, and even the goals that drive explanation as they interact during the explanation process.

The model builds on several strands of research on aspects of explanation and diagnosis, placing them in a framework integrating the performance and explanation system as interacting components and replacing the traditional purely top-down explanation process with one in which the explanation process itself dynamically refocuses learner goals. This chapter discusses the tenets of this model and the processing capabilities and knowledge required to support this real-world explanation process.

1.1 Isolated vs. Interactive Explanation

In learning systems that use explanations, the explanation generation process is generally treated as an isolated subroutine. That subroutine may be provided with descriptions of what should be explained, such as target concepts (Mitchell et al., 1986) or particular anomalous states (Hammond, 1989); it may be provided with additional criteria for what constitutes a useful explanation, such as operationality criteria (Mostow, 1983; Keller, 1988); and it may also be able to access system background knowledge to verify particular assumptions. However, given this information, the actual explanation generation process is done without incremental monitoring of the course of explanation—of how the partial hypotheses that have been generated affect current system knowledge goals, of the cost incurred in constructing the explanation so far and the utility of continuing, and of the aspects of the situation that become more important based on the current partial explanation. Nor can the explanation system call upon the performance system to take action in the surrounding environment in support of explanation, or suspend explanation to allow interleaving of explanation with action and processing for other tasks. Instead, the explanation process simply proceeds in isolation, based on the goals initially provided to it, until it either terminates successfully or reports failure. In the remainder of this chapter that model will be referred to as the "isolated explanation" model.

In contrast to the isolated explanation model, this chapter argues for a model in which the reasoner goals that guide explanation, as well as the explanation generation process itself, are shaped dynamically by ongoing interactions with other tasks, goals and the changing external environment that form the surrounding circumstances in which explanation is done. Psychological data show that an important component of human model-building is the joint application of active strategies and prior knowledge that mutually constrain the model-building process (Schauble et al., 1992); the proposed model takes a related view of explanation generation as involving mutually contstraining knowledge and action. The model is called *goal-driven interactive explanation* (GDIE), to emphasize that its ongoing processing is not only driven by existing goals, but interacts fully with current goals, other goal-driven internal processing, and the external world.

Goal-driven interactive explanation has two main tenets. The first is that the role of goals in real-world explanation is not adequately modeled by purely top-down influences fixed at the start of the explanation process. Instead, effective real-world explanation depends on being able to change the criteria used to judge good explanations, and even the overarching goals determining those criteria, in response to incremental results of the on-going explanation process, in turn causing the explanation process to be refocused. As a result, the goal-driven interactive explanation process is more "situated" (e.g., Suchman, 1987) in its environment than conventional explanation models. However, despite the importance of environmental factors, the model is committed to strategic and deliberative reasoning rather than purely reactive activity. The second tenet is that explanation generation should be fully integrated with other processing, making explanation generation a multistrategy process that can draw upon all the capabilities for reasoning and action of the system building the explanation. This places the pursuit of knowledge goals and task goals on an equal footing, with each to be guided by the same process. The two principles of goal-driven interactive explanation allow not only more effective explanation generation, but also flexible revision of decisions of when and what to explain, in light of all available information.

1.2 Chapter Overview

The chapter begins with a discussion of the reasons that goal-driven interactive explanation is needed for real world explanation. It next considers the issues that must be addressed to model that process. It presents a brief case study of a computer model addressing one part of the process, the problem of generating knowledge goals to drive explanation, and directions in which that model is being extended to develop a goal-driven interactive explanation system. The chapter closes with a discussion of the relationship of this process to broader issues in goal-driven learning, extending the discussion of GDIE to propose that goal-driven learning in general requires that the learning process be situated in the context of other tasks and goals in the environment. Situating the learning process in the context of the larger environment and the reasoner's capabilities for action makes learning more effective, by making it possible for the learner to apply both reasoning and action to support the learning process, and by making it possible to dynamically re-focus learning in response to incremental changes in goals and circumstances.

2 Real-World Explanation: An Everyday Example

The following example illustrates the need for GDIE during everyday explanation. When a person attempts to explain an automobile break-down, the explanation process is likely to involve strategic use of multiple explanation strategies and to be strongly influenced by changing factors in the world as the effort to explain proceeds. This process is both a goal-driven learning process in which new information is gathered and existing knowledge is transformed to satisfy knowledge goals, and a process in which the explainer's goals are themselves transformed as the explanation effort proceeds, refocusing processing.

Initially, when beginning to explain a car problem, a driver is likely to generate new knowledge goals as part of the explanation process (e.g., the goal to find out whether there is a wiring problem), and to try to satisfy them by chosing among many possible strategies for satisfying them, either by internal processing (e.g., by retrieving relevant prior cases), or by taking action to gather more information (e.g., by looking under the hood to note anomalies such as charred parts, or to verify the condition of electrical connections). The choice of strategies will be affected both by the explainer's internal state and by the external circumstances in which explanation is done. For example, internal factors may include the expected usefulness of satisfying the knowledge goal, given current task goals and the state of current hypotheses; environmental factors may include the availability of assistance, the costs of satisfying the knowledge goal (e.g., whether rain means that getting out of the car to look under the hood will result in getting wet), and so forth. Thus existing plans, goals, knowledge, and external circumstances change the course of explanation generation.

However, the flexible pursuit of explanatory goals is only half of the everyday explanation process: Partial results of the explanation process may themselves change current plans, goals, and circumstances, with repercussions outside of the current explanation problem. For example, attempting to generate an explanation by looking under the hood for problems requires getting out of the car and opening the hood, which may provide an opportunity to repair a pre-existing problem. Carrying out that repair may prompt the generation of new knowledge goals and new plans to satisfy them. Likewise, if someone rushing to the airport begins to suspect, based on a partial explanation, that the car problem is serious (even if it has not been completely diagnosed), continuing the attempt to explain is pointless, and the attempt to explain may be abandoned in favor of taking a taxi instead. This change gives rise to a new goal—getting a taxi—that makes explanation of the break-down unimportant. The new goals in turn generate new goals for explanation. For example, the goal to get a taxi may prompt a goal to explain why no taxis are driving by, which may also require actions in the world (e.g., going to a policeman and asking for information).

This example illustrates that the everyday explanation process both influences and is influenced by current circumstances and goals. The key issue for GDIE is to determine the particular criteria that are important in a given situation, given current knowledge goals and circumstances, and to be able to change those goals and adjust the explanation processing accordingly as explanation proceeds. This explanation process both serves *and determines* higher-level goals in a goal-driven learning system.

3 Explanations and Explanation Generation

Artificial intelligence systems often model explanations as dependency chains that derive a conclusion to be explained from either known or assumed information. For example, if a mechanic finds a defective engine component when trying to diagnose bad acceleration, an explanation of the acceleration problem might trace how the observed acceleration problem was caused by the defects in the component. The type of derivation may be interpreted as deductive proof (e.g., Mitchell et al., 1986), or as plausible reasoning (e.g., Pearl, 1988; Schank, 1986), but the basic structure remains the same. Thus a fundamental question for such models of explanation is how to effectively generate explanatory chains. Explanatory chains are often generated by a single strategy—backwards chaining from the conclusion to be explained (e.g., Kautz and Allen, 1986; Mitchell et al., 1986)—though case-based approaches have also been proposed (see Schank et al. (1994) for an overview), as has the combination of these strategies (Hammond and Seifert, 1991).

In both chaining and case-based approaches, explanation generation has been treated as an isolated subroutine to be called by the main processing system. The environment and the goals of the explainer are assumed to be fixed at the start of the explanation generation process, so that the task of explanation generation is to use prior knowledge to build the needed explanation according to fixed goals. This reflects an idealized conception of explanation as an act of pure reasoning, following the spirit of Sherlock Holmes' directive that "the art of the reasoner should be used rather for the sifting of details than for the acquiring of fresh evidence" (Doyle, 1976, p. 1). Accordingly, the explanation generation component cannot call upon other parts of the reasoner to aid in the explanation effort; as a result, it has no means to obtain any information beyond the information provided to it at the start of the explanation process. In this way, the explanation generation component is isolated from its surrounding world.

Likewise, although isolated models of explanation may reflect the goals of the explanation system (e.g., in requiring that explanations explain membership in a given target concept and that the result be operational (Keller, 1988; Mitchell et al., 1986)), their goal-based constraints are fixed at the beginning of the explanation attempt. The ramifications of partial results from the explanation process on goals *beyond* the explanation process are not monitored during explanation, and consequently cannot influence either task goals or the knowledge goals driving explanation until either a complete explanation satisfying the original goals has been generated or the explanation effort has failed. Isolation from the environment and from the full range of system goals can be serious impediments to real-world explanation.

The need for interaction with the environment to support explanation

When real-world explainers start explaining, they often lack the information that they need to complete the explanation. In such situations, explaining by reasoning alone is impossible. Consequently, a robust explanation mechanism must be able to take action in the external world. In fact, even the earlier quote from Sherlock Holmes in favor of pure reasoning occurs in a context that emphasizes the need for action and belies Holmes' edict: the statement is made as Sherlock Holmes is traveling to the scene of the crime to investigate for himself.

In addition, even if an explainer *could* eventually reach the needed conclusion by reasoning in isolation, explaining in isolation might still not be worthwhile. The most effective means for drawing a conclusion may be to verify it in the world, rather than to infer it from what is already known. For example, an engineer designing an airplane wing may have all the information needed in principle to calculate how it will perform, but a wind tunnel test may still be the most practical means for determining its characteristics—or for identifying surprising aspects of its behavior on which to focus further reasoning.

The need for incremental interaction between the explanatory process and explainer goals

In traditional explanation systems, the explanation component generates an explanation in isolation until it returns the final result of its processing to the system that requested the explanation. However, explanation can be directed more effectively if the reasoner monitors the incremental results of explanation generation in order to revise its goals accordingly and, in turn, to redirect the explanation process as needed.

Actions during explanation may result in opportunities that are best addressed before completing the initial explanation. For example, if a car owner disassembles the car's engine in order to explain a strange noise, the engine disassembly process may provide the opportunity to explain other problems at low cost (e.g., why the car is burning oil). Likewise, if a car owner is explaining why his car burns oil in order to repair it, and partial explanation reveals other severe problems before the oil problem is diagnosed, the owner may abandon the repair goal and generate a new goal—selling the car—before having explained why the car is burning oil. Given the change in goals, it may become necessary to explain other problems instead (e.g., any problems that did not bother the original owner but that would dissuade potential buyers). Ongoing reasoning about multiple goals, current information, and current information needs is crucial to guiding this explanation process, and an explanation system that completed each explanation before considering its ramifications would waste much effort.

4 Issues in Building a Model of Goal-Driven Interactive Explanation

A model of goal-driven interactive explanation must be able to reflect the effects of learning goals on explanation generation, and to reflect the effects of explanation generation on learning goals. In particular, an architecture for a GDIE system must enable the following behaviors:

1. **Dynamic refocusing of explanatory effort:** The cost and value of pursuing a particular explanation may change markedly as a result of changes in the explainer's knowledge, capabilities and environment. Consequently, the explanation generation process must be able to switch its attention between candidate explanations as circumstances change.

   Maintaining a set of alternative explanations and reasoning about the relative utility of pursuing the competing alternatives changes the explanation process from focusing on single explanations to focusing on a group of competing explanations, in a process more like differential diagnosis in medicine (e.g., (Harvey and Bordley, 1970)), in which competing hypotheses are simultaneously considered and weighed against one another. Although differential diagnosis is a common approach in medical expert systems (e.g., (Clancey and Shortliffe, 1984)), it has received very little attention in other applications of abductive explanation such as abductive understanding. Some research has examined the change of focus during explanation based on incremental plausibility estimates for the tasks of diagnosis (e.g., (de Kleer and Williams, 1989)) and understanding (e.g., (Ng and Mooney, 1990)); plausibility considerations must be augmented with utility criteria reflecting the information contained in a partial explanation, current knowledge goals, their relative priorities, and factors in the external world affecting the difficulty of pursuing the candidate explanation by alternative means. A difficult open problem is how to estimate the utility of particular actions in complex and imperfectly understood situations. Likewise, the process must take into account how new information changes the set of explainer goals (Turner, 1994).

As circumstances change, the best means for pursuing a particular explanation may change. For example, the arrival of a tow truck may make the driver's efforts to reconstruct the circumstances of a similar previous problem superfluous, because the tow truck driver is likely to have more experience organized more usefully, and simply needs to be asked for a diagnosis. Refocusing explanatory methods in a multistrategy system requires the ability to generate, suspend, continue, and cancel explanatory tasks based on goal-based considerations. Decisions on how to refocus effort must reflect not only local concerns but how those decisions affect the goal-driven reasoner's global performance.

2. **Interweaving explanation generation with other tasks outside explanation:** Rather than waiting for generation of a complete explanation, as in the isolated explanation model, GDIE interweaves the explanation effort with other tasks. For example, when responding to car problems it is often appropriate to act based on a tentative hypothesis that can later be revised. If a car fails to start on a cold day, it is reasonable to attribute the problem to the cold as an initial hypothesis and to respond to it by keeping the car in the garage on cold nights, but to store other alternative hypotheses (e.g., a weak battery) to be considered if information relevant to the choice between them becomes available opportunistically (Ram, 1993). In GDIE the explanation task itself is pursued not as the highest-level reasoner goal, to be satisfied for its own sake, as is generally the case in diagnosis systems, but explicitly in service of other goals. As a result, the explanation process itself is in competition with other tasks serving those goals.

3. **Recognizing useful partial results and applying them as soon as they are available:** In order to effectively interweave explanation generation with other tasks, a GDIE system must monitor the current results of explanation to recognize and use the information from partial explanations and tentative hypotheses as soon as that information is useful.

4. **Revising system task and knowledge goals based on incremental information:** New information obtained during explanation—either from the explanation effort itself, or from other sources—should be able to change system goals. For example, a driver whose car breaks down while the driver was rushing to the airport may initially seek an explanation that will enable him or her to repair the problem. If the explainer determines that the problem involves the power train, that explanation is insufficient for the original purpose—it is too vague to direct a repair—but the information it provides is nevertheless sufficient to show that the original goal needs to be changed. Power train repairs are time-consuming, so that executing a transmission repair would be incompatible with the higher-level goal of reaching the airport in time. Consequently, the repair goal (and explanation towards that goal) should simply be abandoned.

4.1 Content Theories for GDIE

Understanding the dynamic focusing of explanation effort, the interweaving of explanation with action, and the revision of system goals based on incremental information, requires developing theories of the *content* of the knowledge required to guide these behaviors. In particular, it is necessary to account for the following four types of knowledge:

1. **Types of goal-based needs for information:** In many diagnosis and explanation systems, the condition to be explained is provided as an external input to the system (e.g., (Clancey and Shortliffe, 1984; de Kleer and Williams, 1989; Ng and Mooney, 1990)). Other systems generate their own explanation tasks in response to processing failures, expectation failures, or performance failures (e.g., (Birnbaum et al., 1990; Hammond, 1989; Leake, 1992; Ram, 1993; Ram and Cox, 1994; Riesbeck, 1981; Schank, 1986)). However, information needs may arise even in the absence of failures *per se*. For example, the policy to attempt to improve performance at a task may motivate learning even if current performance matches expectations and is successful, simply because there may be a way to achieve even better performance. Thus a model of goal-driven explanation must be based on a theory of the tasks of a reasoner and the classes of information needs that are generated by particular tasks. (This has been done for a number of tasks in (Leake, 1991a), Chapter 9). It also requires developing mechanisms for identifying each type of need when it arises, either in response to explicit failures or due to other causes.

2. **Strategies for determining how far to elaborate explanations:** Events can be explained at many levels of granularity, and explanatory chains can often be traced arbitrarily far back from the conclusion being explained. For example, a car breakdown could be explained at different levels of granularity by "a mechanical failure", "a transmission failure", "a failure of part X within the transmission," etc.; the failure of particular part X might be traced to a stress failure, or farther back to the soft metal from which the part was made, or still further to the transmission designers' decision to economize in order to keep the cost of the transmission below that of a competitor's transmission, which might in turn be explained by management decisions about competition strategies, and so on. Consequently, principles must be formulated to decide how far to trace an explanation, and the level of detail to use in the explanation, given current knowledge goals and other circumstances. For example, a reasoner might have rules to guide failure driven-explanation based on other current tasks, such as *when time is limited, explain only as far back as necessary to resume execution of the plan whose failure triggered the explanation effort.* Other rules might give guidance for elaborating partial explanations, as in *if a candidate explanation involves a device failure, and the device is used in a current plan or pending plan, explain until the failure is confirmed or ruled out.*

3. **Classes of methods for pursuing explanations and criteria for choosing which to apply:** Once a reasoner has decided what information to search for, it must decide how to guide its explanatory process. Because a GDIE system can generate goals to be satisfied by any of the components of the performance and learning system, it can

either explain by internal reasoning, as done in the traditional isolated explanation model, or prompt action in the world to determine that information. Choosing which strategies to apply requires reasoning about the types of knowledge goals driving the system and the functionality of different types of strategies for performing the needed learning (Ram and Cox, 1994); it also requires reasoning about the resources required to perform particular information-seeking actions, the circumstances under which they apply, and their tradeoffs (Hunter, 1990b).

As a concrete example of the type of reasoning that must be performed, an automobile mechanic must be able to reason about the tradeoffs between different means of determining whether a particular part is defective, such as attempting to remember and compare the symptoms of previous failures, disassembling the part, swapping the suspect part with a replacement and observing the effects of the swap, etc.; this explanation process can be modeled as a planning process (Turner, 1994). In general, which operations are appropriate will depend strongly on both the current external environment and system capabilities and resources (e.g., the time and tools available to the mechanic and the mechanic's competing obligations). Thus in combination with knowledge of the functionality of different types of learning strategies, utility-based considerations should play a role in deciding what should be learned and which learning methods to apply (desJardins, 1992; Gratch and DeJong, 1993; Pryor and Collins, 1992).

4. **Ways new information can impinge on current plans and goals:** In order for a GDIE system to monitor new information and re-adjust its goals accordingly, its processing must be based on a theory of the ways in which new information can relate to prior plans and goals and of how new information is analyzed to reveal threats and opportunities and prompt the generation of new goals.

5 A Simple Illustration

To illustrate the process, consider how GDIE changes the explanation process of an explanation system that starts its attempt to explain a car problem by using case-based reasoning to retrieve and adapt prior explanations stored in memory (e.g., (Leake, 1992; Schank, 1986; Schank and Leake, 1989; Schank et al., 1994)). Suppose that the case-based explainer retrieves three candidate explanations: the battery is dead, a wire to the starter is disconnected, or the starter itself is defective. In a case-based explanation system that explains by pure reasoning, in the isolated explanation model, it is reasonable to pursue the explanation for the most similar prior problem as the one to pursue first, on the assumption that the most similar problem is the one most likely to apply. For example, if the situation significantly resembles one with starter problems, the system would first attempt to apply that explanation, only abandoning that explanation if it was impossible to apply.

The previous strategy is appropriate for the isolated explanation model, in which the explainer cannot act to verify its choice: given the regularities of the world, the explanation for the most similar case probably *is* the one most likely to apply. However, in this instance, it is inefficient for an explainer that *can* take action. Much effort might have to be expended to confirm or disconfirm the bad starter (at the minimum exiting the car and examining the starter, the connecting wires, etc.), while another alternative can be checked at almost no cost: it is possible to test whether the battery is dead by turning on the radio. Consequently, even if the dead battery is less likely, it may be worthwhile to check it first, if there is any chance that it might be the problem, to try to repair it before doing additional verification. Note that knowing which checks *are* simpler depends on knowing the circumstances in which they will be performed; the situated nature of the GDIE process makes it possible to dynamically base decisions on reasoning about current circumstances and system capabilities.

The explanation process for the starting problem might prompt other learning goals as well, such as the goal to learn generalizations about how similar episodes can be prevented or predicted; it might also change the goals that prompted the effort to explain. Which goals are pursued must depend on other system goals, their priorities, and the availability of resources for conducting the learning. For example, if the car that failed to start was a taxi and the potential explainer is a passenger, trying to help explain the starting problem may be worthwhile if the taxi is in an isolated place and the driver appears to be an incompetent mechanic, but not if another taxi is nearby—unless the passenger recently bought a car for the first time and is trying to learn about common car problems. New learning goals that are generated must in turn be planned for, both by internal, inferential actions, such as attempting to form new generalizations about which brands of cars tend to have starting problems, and by actions in the external world; that planning process must be done in light of other goals and tasks competing for the explainer's processing resources.

6 A Broader Perspective

The GDIE model described in the previous sections is both an instance of a goal-driven learning model for a particular task—explanation—and a general framework for goal-driven learning processes.

First, GDIE is a goal-driven learning process that applies at least four types of learning strategies, according to reasoning about goal-based needs: Learning by gathering information, learning by transforming existing knowledge (e.g., during the inference process used to elaborate explanations), learning by forming new connections between beliefs, and learning by storing and re-indexing the explanations and information-gathering plans that it builds. In addition, the GDIE process should be integrated with other goal-driven learning processes to provide it with access to the learning strategies of the overarching system. For example, inductive generalization might be used to hypothesize new causal rules to be used by the explanation component. Conversely, because explanations are required as the starting point for other learning methods such as explanation-based generalization (Mitchell et al., 1986; DeJong and Mooney, 1986) and explanation-based indexing (Barletta and Mark, 1988; Leake and Owens, 1986), GDIE should also be viewed as a new tool to be added to goal-driven learners' repertoires of methods for forming the explanations required to support desired learning.

In the GDIE model, the choice between types of learning is dictated by mediating between the factors that shape any goal-driven learning process: task goals, knowledge goals, and the environment in which processing occurs. Consequently, the GDIE framework has ramifications for any goal-driven learning process: the theory of how to perform this mediation, as well as the mechanisms needed to enable this task, apply to any goal-driven learning process.

7 Progress Towards the Model

Our previous research addresses one set of issues for developing the GDIE Model: How a reasoner with certain tasks in the world can guide explanation generation according to current needs. In particular, it addresses (a) how a reasoner can determine that explanation is needed in support of the understanding task, and (b) how, if explanation is needed, the system can generate knowledge goals characterizing the information needed to achieve adequate understanding and to satisfy knowledge goals arising from other tasks. The following sections sketch how these questions are addressed in the story understanding system ACCEPTER (Leake, 1988; Leake, 1991b; Leake, 1992; Leake, 1994) and discuss research on applying the model in a more general framework.

7.1 ACCEPTER

ACCEPTER is a story understanding program. It takes as input simple (1-4 line) news stories, detects anomalies in those stories, characterizes the anomalies and retrieves stored explanations for the anomalies based on their characterizations, and then evaluates candidate explanations' plausibility and usefulness for tasks selected by a human user. The episodes it processes include the death of the star racehorse, Swale, after winning a major race, the death of basketball star Len Bias, who died the day after being selected as first choice in the basketball draft, the explosion of the Space Shuttle *Challenger*, two versions of the story of the accidental shootdown of an Iranian airliner by the American warship *Vincennes* (one story reflecting initial news reports that the ship had shot down an attacking military plane, and the other reflecting the story after all facts were known), and the recall of Audi 5000 cars for transmission problems.

Each candidate explanation can be evaluated according to its usefulness for six types of goals: using the explanation to resolve an anomaly, to determine how to predict similar outcomes in the future, to prevent similar outcomes in the future, to repair the current situation, to highlight a particular actor's responsibility and to assign blame. A complete description of ACCEPTER is available in (Leake, 1992).

7.2 Deciding when to explain

ACCEPTER's basic understanding process is schema-based; it understands routine events in stories by fitting new information into pre-stored schemas in a process along the lines of those discussed by various researchers (Cullingford, 1978; Lebowitz, 1980; Schank, 1982; Schank and Abelson, 1977)). However, it augments that process by using a library of explanations of anomalies to explain novel events when its schema-based understanding is inadequate.[1] As part of the SWALE system, ACCEPTER can also call upon an external module, the TWEAKER (Kass, 1990), to adapt existing explanations to fit new situations when needed; as a stand-alone explanation system, its attempt to explain simply fails if none of its stored explanations applies to the new situation.

One of the questions addressed by ACCEPTER is the question of when to explain. Because the task of a schema-based story understanding process is to account for new information in terms of schema-based expectations, its performance is sufficient as long as the new information is consistent with prior beliefs and expectations; failures occur when expectations or beliefs based on previous understanding prove false. Thus the need to go beyond schema-based understanding is shown by *anomalies*—conflicts with prior beliefs and expectations. Resolving those anomalies requires both (a) understanding the surprising features of the situation, and (b) accounting for what led to erroneous prior beliefs or expectations, in order to repair the flaws in knowledge that led to the erroneous predictions. In ACCEPTER, anomaly detection guides explanation according to the process shown in figure 1. For example, if a reasoner who believes that automatic teller machines are impenetrable receives the input "John broke into an ATM," that input is anomalous with respect to the reasoner's prior belief about ATMs. In order to reconcile the new information with prior beliefs, the break-in must be explained to show why the prior belief was wrong, and to update the reasoner's world model in order to avoid future understanding failures.

Because anomalies are unexpected aspects of a situation, they reveal that the understander's prior world model was flawed. The aspects of the current situation that were not previously known, but that were revealed by the anomalies, may have ramifications beyond simply requiring that the reasoner's knowledge be updated: Updates to prior knowledge may impinge on current goals and plans by revealing threats and opportunities that were not previously apparent. The need to respond to those threats and opportunities may prompt additional information requirements beyond those for merely resolving the anomaly at hand, such as those described in (Leake, 1991a), (Chapter 9 in this volume). In turn, those information requirements may require elaboration of the initial explanation generated for the anomaly, as shown in figure 2. The following sections highlight aspects of the mechanisms for focusing each part of these processes.

7.3 Formulating knowledge goals to guide search for explanations

When explanation generation is prompted by anomalies, the need to resolve those anomalies determines the focus of explanation, determining which of the many valid explanations for

[1] The standard schema-based understanding process can be viewed as "explaining" events by relating them to system schemas, but in the following discussion we reserve the term "explaining" for the process of chaining together inference rules to build chains of rules connecting new information to prior beliefs. In ACCEPTER, these explanations are represented as *explanation patterns* (XPs) (Schank, 1986; Schank and Leake, 1989).

Anomaly
$\Downarrow$ reveals
Flawed understanding
$\Downarrow$ triggers
Goals to make sense of the event and account for belief failure
$\Downarrow$ triggers search for
Explanation providing information to resolve anomaly
$\Downarrow$ enables generation of
Updated world model

Figure 1: How anomalies prompt the tasks to update a world model.

(Updated world model) + (Current goals and plans)
$\Downarrow$ reveals
Threats and opportunities
$\Downarrow$ reveal
Needs for additional information to respond
$\Downarrow$ trigger
Goals to build explanations providing that information
$\Downarrow$ direct
Explanation adaptation

Figure 2: How new information and goals prompt new tasks, requiring that an explainer seek additional information.

a given event are actually relevant. For example, possible explanations for "John broke into an ATM" include:

- John needed money to pay back a loan shark for gambling debts.
- John believes that robberies of ATMs are more likely to succeed than bank robberies.
- The ATM's security camera had been removed for repair.
- John's uncle wouldn't lend him any more money.
- Metal in ATM locks is brittle when frozen, so the locks can be shattered by freezing them and striking them with a mallet.

The previous explanations are only a small subset of those that might be generated, and they (and countless other explanations) might simultaneously be valid. However, whether any one of them is useful varies according to the anomaly that motivated explanation. For example, if the reasoner was surprised that the break-in succeeded despite bank precautions, the explanation that the security camera was missing would be relevant and John's need for money would be irrelevant. If the explanation generation process fails to reflect the reasons for explaining, it is doomed to impartially generate irrelevant candidates as well as relevant ones, exacerbating the problem of explanation generation cost (which may be considerable, especially if action in the world—e.g., tracking down witnesses to the break-in—is needed to form candidate explanations) and resulting in explanations that are irrelevant to the explainer's information needs.

When ACCEPTER detects an anomaly, it formulates criteria for the information that the explanation must provide to resolve the anomaly. Thus unlike explanation-based generalization systems in which the target concept to explain is provided as an input to the explanation system (Mitchell et al., 1986), and unlike other systems in which novel events are explained according to a single fixed perspective (e.g., to find the actor's motivation (DeJong, 1986; Mooney, 1990; Wilensky, 1978)), ACCEPTER dynamically decides which features of the situation must be accounted for, according to its current needs to resolve the anomaly.[2]

ACCEPTER's needs for information to resolve the anomaly are encoded in an *anomaly characterization* that serves as a knowledge goal to guide retrieval of relevant explanations (Leake, 1991b; Leake, 1992). Likewise, when understanding is done in service of additional tasks, ACCEPTER characterizes the information needs for those tasks (see (Leake, 1991a), Chapter 9 of this volume), generating additional knowledge goals and passing them to an adaptation process that selects strategies for adapting the retrieved explanation to satisfy those goals (e.g., by elaborating portions of the explanation that are too vague to provide useful information). However, although the case adaptation process for case-based explanation has been extensively studied for repairing plausibility problems (Kass, 1990), little work has been done on developing adaptation strategies responding to particular types of information needs. Developing effective strategies for other needs is an important open problem.

[2]ACCEPTER explains each anomaly as soon as it is detected; consequently it explains only one anomaly at a time. Another explanation strategy would be to complete examination of an input to find all anomalies before beginning explanation, and to attempt to simultaneously explain those multiple anomalies using strategies such as *coordination of anomalies* (Schank, 1986).

Figures 3 and 4 illustrate this process in context of the previous example of understanding an ATM break-in. As before, we assume that the reasoner had a prior belief that ATMs were impenetrable, so that the anomaly in "John broke into an ATM" is that the break-in was possible. Figure 3 shows the sequence of steps leading to the explanation that ATM metal becomes brittle when frozen, allowing the break-in to succeed. This new information might suggest new opportunities for an explainer who was a criminal: carrying out the ATM break-ins that previously seemed impossible. In response, that criminal would perform further goal-driven explanation to elaborate on how the freezing was carried out, in order to replicate the crime.

ACCEPTER addresses questions of how the basic understanding task triggers explanation, how that and other tasks impose requirements on the information that must be provided by good explanations, and how those requirements guide internal search for explanations through case-based reasoning. However, it does not address questions of how tasks beyond understanding are selected to guide explanation (it simply takes them as inputs from a human user), or of the interplay between ongoing explanation and internal and external tasks. In a goal-driven interactive explanation system, the process described in this section would serve as one of a set of multiple processes for guiding learning. Those processes not only would reflect overarching tasks, but would interact with those tasks and the goals those tasks serve to re-focus explainer attention and system goals.

7.4 Extending the Model

Extending ACCEPTER's model to an active GDIE system requires integrating ACCEPTER's reasoning about when and what to explain into a planning and execution system that can generate explicit knowledge goals, based on task goals, plan for their achievement, and monitor the course of explanation and adjust goals accordingly. Research is underway to perform that integration, integrating explanation with a planner and execution system in a simple simulated world (Sooriamurthi and Leake, 1994). The system explains vehicle malfunctions from the perspective of stranded motorist who is explaining in order to further the goal of reaching a destination under time and resource constraints, in a dynamically changing simulated environment.

In general, many alternative learning methods may apply to generating initial hypotheses, such as asking for possible explanations or looking them up in a manual, using case-based reasoning, or using backwards chaining from the symptoms to be explained; a goal of the new system is to model reasoning about how to mediate between possibilities for reasoning and action to satisfy information needs (e.g., to seek help or to form a more precise characterization of the symptoms to explain). After the driver has formed an initial explanation, additional learning is needed to verify the explanation and to elaborate it until it is sufficiently detailed to suggest a response. Verification and elaboration in turn may require additional reasoning, such as additional chaining or retrieval of cases, or additional information-gathering (e.g., by memory search, knowledge transformation, or by search in the world). During its processing, the system not only will learn about the external world, but also, by storing the traces of its reasoning process (Veloso and Carbonell, 1993; Ram and

Anomaly

John succeeded in breaking into an ATM contrary to expectations for their security

$\Downarrow$ reveals

Flawed understanding

Expectation for secure ATM failed

$\Downarrow$ triggers

Goals to make sense of the event and account for belief failure

Goals to find ATM's weakness and to account for forming the previous false prior belief

$\Downarrow$ triggers search for

Explanation providing information to resolve anomaly

Metal in the ATM lock is brittle when frozen, so it can be shattered by freezing and striking

$\Downarrow$ enables generation of

Updated world model

Model includes *The locks of ATMs can be defeated by freezing*

Figure 3: Explaining to repair the flawed understanding revealed by an anomaly.

(Updated world model) + (Current goals and plans)

(The locks of ATMs can be defeated by freezing) + (Goal to obtain wealth)

$\Downarrow$ reveals

Threats and opportunities

The opportunity for a successful ATM break-in elsewhere

$\Downarrow$ reveal

Needs for additional information to respond

The need to know a plan for freezing the lock

$\Downarrow$ trigger

Goals to build explanations providing that information

The goal to elaborate the explanation of John's technique.

$\Downarrow$ direct

Explanation adaptation

Explanation including details of how John froze the lock

(E.g., that John poured liquid nitrogen into the keyhole.)

Figure 4: How information about ATM vulnerability prompts new goals, giving rise to new information requirements.

Cox, 1994), will improve its reasoning strategies and to transform and reorganize existing knowledge to facilitate its application in the future.

8 Relationship to Other Approaches

The GDIE framework relates to a number of other approaches. The idea that target concepts for explanation-based learning should be based on information requirements from current plans was proposed in (Kedar-Cabelli, 1987). That paper presents a model of explanation-based generalization in which target concepts are derived from explicit annotations in plans about the requirements for artifacts used in those plans. However, GDIE bases its knowledge goals on more general reasoning about the relationships between plans, goals, and information needs. In addition, as discussed in Chapter 1, target concepts serve a specifications of the desired output of a learning strategy rather than the actively-sought knowledge goals that drive GDIE. Rather than being limited to backwards chaining to build explanations, the goal-driven learning model of GDIE allows flexible management of how actions should be integrated with explanatory reasoning.

The need to obtain additional information during explanation has long been recognized in research on diagnostic expert systems (see Clancy and Shortliffe (1984) for a sampling of some of these approaches), but has not been modeled in systems that use explanation for other tasks (e.g., (Hammond, 1989; Kautz and Allen, 1986; Mooney, 1990)). Also, GDIE not only allows the explainer to request information, but models explanation as a goal-driven learning process that strategically uses multiple methods to seek new information and transform prior information. In that way, it increases the range of methods available to the reasoner and takes into account the context outside the diagnostic task; applying them successfully depends on answering new questions about the knowledge needed to guide the explanation process for tasks outside of restricted diagnostic domains.

Although models of abductive explanation generally consider only one explanation at a time, some existing models of explanation refocus attention on different explanations in order to focus on the most likely candidates (de Kleer and Williams, 1989; Ng and Mooney, 1990). GDIE adds the capability to focus attention according to usefulness considerations rather than only plausibility. Krulwich et al. (1990) propose that goal-based considerations should drive hierarchical explanation generation, but their framework is a an isolated explanation process in which explanation is done by reasoning alone and goal-based influences exert only top-down influence. It does not consider the role of changing goals during explanation, the influence of competing tasks and alternative explanations, the use of multiple explanation strategies, or utility-based considerations of how the cost of explanatory actions affect the explanation process.

Goal-based considerations have begun to be examined in research on diagnosis (e.g., (Freitag and Friedrich, 1991; Redmond, 1992; Rymon, 1993; Turner, 1994)). Such models of diagnosis include doing explanation as part of a cycle of diagnosis and repair (Rymon, 1993) and interleaving reasoning with action (Redmond, 1992; Turner, 1994); and making context-sensitive choices of what to diagnose first (Turner, 1994). A goal of GDIE is to integrate all these aspects of diagnosis into a uniform framework, and to apply that framework to guide explanation and action in support of all types of system knowledge goals.

One example of the use of goal-driven information search in non-diagnostic contexts is provided by the story understander AQUA (Ram, 1991; Ram, 1993), which varies the depth of its reading of input stories in order to actively seek desired information in the stories it reads. Another model in similar spirit is Quilici's (1994) QUACK, which models the goal-driven learning process by which novice UNIX users acquire expertise. In that model both action and reasoning, including explanation, are used to satisfy the learner's knowledge goals.

Existing models sometimes take an active approach to verifying explanations in the external environment, using experimentation to choose between candidate explanations (e.g., Rajamoney, 1993). In GDIE, action in the world is also used to help *generating* explanations, rather than only being used as part of a selection process for choosing between the candidate explanations that have already been built.

As already discussed, the spirit of goal-driven interactive explanation relates to theories of situated activity (e.g., Suchman, 1987) in its emphasis on the role of interactions with the environment in guiding action. In GDIE, explanation is viewed as part of an ongoing process of interaction between explanation, other internal processes, and the external world. However, rather than being behavior-based, the GDIE framework is deliberative, goal-based and strategic.

The role of one type of information-seeking interaction during explanation—interactive dialogue with an expert—has been investigated in depth by Redmond (1992), and research on expert system explanation has also considered the role of dialogue in providing the users of expert systems with useful explanations. For example, Moore and Swartout (1989) argue that expert system explanation should be *reactive,* in the sense of being an on-going dialogue with the user, with previous user questions determining how later questions are interpreted. However, a goal-driven interactive explanation system corresponds both to the explainer and the *user* of a dialogue system in generating requests for explanation, attempting to generate its own answers, and evaluating the suitability of those answers in order to refine its goals and to generate new knowledge goals to guide further explanation search as needed to satisfy those goals.

9 Conclusions

Explanation-based methods can play an important role in guiding useful learning. In order to be successful in that role, such methods must start from explanations that are relevant to overarching goals. Generating useful explanations effectively requires focusing the explanation generation process on satisfying particular knowledge goals. As the explanation process attempts to build explanations satisfying those knowledge goals, it can itself give rise to new knowledge goals, requiring learning through multiple strategies such as internal inference, memory search, and active information search in the world. Thus explanation can both serve existing learning goals and generate new learning goals to be pursued through additional goal-driven learning.

The beginning of this chapter proposes a framework for a goal-driven interactive explanation generation process that can both serve a reasoner's goal-driven learning and motivate new action and learning. This process is shaped by the environment in which explanation is conducted—both by internal goals and external environmental factors. In this process, explanation and action are combined in a unified goal-driven process of information search, and the reasoning process that traditionally constitutes the entire explanation mechanism is only one of a range of information-seeking actions that can be pursued according to reasoning about the knowledge that is needed and how to acquire that knowledge.

Goal-driven interactive explanation facilitates explanation in complex and imperfectly understood domains. First, it increases the range of phenomena that can be explained, by including external information-seeking actions in the explanation process. Second, by making explanation effort sensitive to utility-based considerations, it increases the effectiveness of explanation generation. Third, by reflecting the interplay between new information, environmental factors, and current goals, it increases the effectiveness of explanation effort by dynamically shifting focus in response to changing needs for information.

Although the GDIE model focuses on explanation generation, it has wider ramifications for goal-driven learning as a whole. Fundamental GDIE issues include how to reason about information needs, how best to satisfy those needs, and how to recognize and respond to changes in current circumstances and the learner's knowledge. Consequently, goal-driven interactive explanation is a rich area for not only its ramifications for explanation, but also its ramifications for the general question of how to focus goal-driven learning.

10 Acknowledgements

The author would like to thank Ashwin Ram and anonymous MIT Press reviewers for very helpful comments on a previous draft of this chapter.

References

Ajjanagadde, V. (1991). Incorporating background knowledge and structured explananda in abductive reasoning: A framework. Wilhelm Schickard-Institute für Informatik WSI 91-6, Universität Tübingen.

Barletta, R. and Mark, W. (1988). Explanation-based indexing of cases. In Kolodner, J., editor, *Proceedings of a Workshop on Case-Based Reasoning*, pages 50–60, Palo Alto. DARPA, Morgan Kaufmann, Inc.

Birnbaum, L., Collins, G., Freed, M., and Krulwich, B. (1990). Model-based diagnosis of planning failures. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 318–323, Boston, MA. AAAI.

Chien, S. (1989). Using and refining simplifications: Explanation-based learning of plans in intractable domains. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 590–595, Detroit, MI. IJCAI.

Clancey, W. and Shortliffe, E., editors (1984). *Readings in Medical Artificial Intelligence: The First Decade*. Addison-Wesley, Reading, MA.

Cullingford, R. (1978). *Script Application: Computer Understanding of Newspaper Stories*. PhD thesis, Yale University. Computer Science Department Technical Report 116.

de Kleer, J. and Williams, B. (1989). Diagnosis with behavioral modes. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1324–1330, Detroit, MI. IJCAI.

DeJong, G. (1986). An approach to learning from observation. In Michalski, R., Carbonell, J., and Mitchell, T., editors, *Machine Learning: An Artificial Intelligence Approach*, volume 2, pages 571–590. Morgan Kaufmann Publishers, Cambridge, MA.

DeJong, G. and Mooney, R. (1986). Explanation-based learning: An alternative view. *Machine Learning*, 1(1):145–176.

desJardins, M. (1992). Goal-directed learning: A decision-theoretic model for deciding what to learn next. In *Proceedings of the Machine Discovery Workshop, Ninth International Machine Learning Conference*, San Mateo. Morgan Kaufmann.

Doyle, A. C. (1976). The adventure of Silver Blaze. In *Memoirs of Sherlock Holmes*, chapter 13, pages 1–16. Schocken Books, New York.

Freitag, H. and Friedrich, G. (1991). Goal-driven structural focusing in model-based diagnosis. In *Working Notes of the Second International Workshop on Principles of Diagnosis*, Milano, Italy. CISE Tecnologie Innovative and Dipartimento di Informatica of the Università di Torino.

Gratch, J. and DeJong, G. (1993). Assessing the value of information to guide learning systems. In *Proceedings of the ML-93 Workshop on Knowledge Compilation and Speedup Learning*, pages 65–71, Amherst, MA.

Hammond, K. (1989). *Case-Based Planning: Viewing Planning as a Memory Task*. Academic Press, San Diego.

Hammond, K. and Seifert, C. (1991). Towards a content model of strategic explanation. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, pages 149–154, Chicago, IL. Cognitive Science Society.

Harvey, A. and Bordley, J. (1970). *Differential Diagnosis; The interpretation of clinical evidence.* Saunders, Philadelphis.

Hunter, L. (1990a). Knowledge acquisition planning for inference from large datasets. In Shriver, B., editor, *Proceedings of the Twenty Third Annual Hawaii International Conference on System Sciences*, pages 35–45, Kona, HI.

Hunter, L. (1990b). Planning to learn. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*, pages 261–268, Cambridge, MA. Cognitive Science Society.

Josephson, J. and Josephson, S. (1994). *Abductive Inference: Computation, Philosophy, Technology.* Cambridge University Press, Cambridge, England.

Kass, A. (1990). *Developing Creative Hypotheses by Adapting Explanations.* PhD thesis, Yale University. Northwestern University Institute for the Learning Sciences, Technical Report 6.

Kautz, H. and Allen, J. (1986). Generalized plan recognition. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 32–37, Philadelphia, PA. AAAI.

Kedar-Cabelli, S. (1987). Formulating concepts according to purpose. In *Proceedings of the Sixth Annual National Conference on Artificial Intelligence*, pages 477–481, Seattle, WA. AAAI.

Keller, R. (1988). Defining operationality for explanation-based learning. *Artificial Intelligence*, 35(2):227–241.

Krulwich, B., Birnbaum, L., and Collins, G. (1990). Goal-directed diagnosis of expectation failures. In O'Rorke, P., editor, *Working Notes of the 1990 Spring Symposium on Automated Abduction*, pages 116–119. AAAI. Technical Report 90-32, Department of Information and Computer Science, University of California, Irvine.

Leake, D. (1988). Evaluating explanations. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 251–255, Minneapolis, MN. AAAI, Morgan Kaufmann Publishers, Inc.

Leake, D. (1991a). Goal-based explanation evaluation. *Cognitive Science*, 15(4):509–545.

Leake, D. (1991b). An indexing vocabulary for case-based explanation. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, pages 10–15, Anaheim, CA. AAAI.

Leake, D. (1992). *Evaluating Explanations: A Content Theory.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Leake, D. (1994). ACCEPTER: Evaluating explanations. In Schank, R., Riesbeck, C., and Kass, A., editors, *Inside Case-Based Explanation.* Lawrence Erlbaum Associates.

Leake, D. and Owens, C. (1986). Organizing memory for explanation. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pages 710–715, Amherst, MA. Cognitive Science Society.

Lebowitz, M. (1980). *Generalization and Memory in an Integrated Understanding System.* PhD thesis, Yale University. Computer Science Department Technical Report 186.

Minton, S. (1988). *Learning Search Control Knowledge: An Explanation-Based Approach.* Kluwer Academic Publishers, Boston.

Mitchell, T., Keller, R., and Kedar-Cabelli, S. (1986). Explanation-based generalization: A unifying view. *Machine Learning*, 1(1):47–80.

Mooney, R. (1990). *A General Explanation-based Learning Mechanism and its Application to Narrative Understanding.* Morgan Kaufmann Publishers, Inc., San Mateo.

Moore, J. and Swartout, W. (1989). A reactive approach to explanation. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1504–1510, Detroit, MI. IJCAI.

Mostow, J. (1983). Machine transformation of advice into a heuristic search procedure. In Michalski, R., Carbonell, J., and Mitchell, T., editors, *Machine Learning: An Artificial Intelligence Approach.* Tioga Publishing Company, Cambridge, MA.

Ng, H. and Mooney, R. (1990). On the role of coherence in abductive explanation. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 337–342, Boston, MA. AAAI.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo.

Pierce, C. (1948). Abduction and induction. In Buchler, J., editor, *The Philosophy of Pierce: Selected Writings*, chapter 11. Harcourt, Brace and Company, New York.

Pryor, L. and Collins, G. (1992). Planning to perceive. In *Proceedings of the AAAI Spring Symposium on Selective Perception*, Stanford, CA. AAAI.

Quilici, A. (1994). Modeling the goal-driven learning of novice unix users. In desJardins, M. and Ram, A., editors, *Proceedings of the 1994 Spring Symposium on Goal-Driven Learning*, Stanford, CA. AAAI.

Rajamoney, S. (1993). Designing experiments to extend the domain theory. In DeJong, G., editor, *Investigating Explanation-Based Learning*, chapter 5, pages 166–189. Kluwer.

Ram, A. (1991). A theory of questions and question asking. *The Journal of the Learning Sciences*, 1(3 & 4):273–318.

Ram, A. (1993). Indexing, elaboration and refinement: Incremental learning of explanatory cases. *Machine Learning*, 10(3):201–248.

Ram, A. and Cox, M. (1994). Introspective reasoning using meta-explanations for multistrategy learning. In Michalski, R. and Tecuci, G., editors, *Machine Learning: A Multistrategy Approach.* Morgan Kaufmann. In Press.

Redmond, M. (1992). *Learning by Observing and Understanding Expert Problem Solving.* PhD thesis, College of Computing, Georgia Institute of Technology. Technical report GIT-CC-92/43.

Riesbeck, C. (1981). Failure-driven reminding for incremental learning. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 115–120, Vancouver, B.C. IJCAI.

Rymon, R. (1993). *Diagnostic Reasoning and Planning in Exploratory-Corrective Domains.* PhD thesis, The University of Pennsylvania.

Schank, R. (1982). *Dynamic Memory: A Theory of Learning in Computers and People.* Cambridge University Press, Cambridge, England.

Schank, R. (1986). *Explanation Patterns: Understanding Mechanically and Creatively.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Schank, R. and Abelson, R. (1977). *Scripts, Plans, Goals and Understanding.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Schank, R. and Leake, D. (1989). Creativity and learning in a case-based explainer. *Artificial Intelligence*, 40(1-3):353–385. Also in Carbonell, J., editor, *Machine Learning: Paradigms and Methods*, MIT Press, Cambridge, MA, 1990.

Schank, R., Riesbeck, C., and Kass, A., editors (1994). *Inside Case-Based Explanation.* Lawrence Erlbaum Associates, Hillsdale New Jersey.

Schauble, L., Glaser, R., Raghavan, K., and Reiner, M. (1992). The integration of knowledge and experimentation strategies in understanding a physical system. *Applied Cognitive Psychology*, 6:321–343.

Sooriamurthi, R. and Leake, D. (1994). Towards situated explanation. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Seattle, WA. AAAI. Research Abstract.

Suchman, L. (1987). *Plans and Situated Actions: the Problem of Human-Machine Communication.* Cambridge University Press, New York.

Turner, R. M. (1994). *Adaptive Reasoning for Real-World Problems: A Schema-Based Approach.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Veloso, M. and Carbonell, J. (1993). Derivational analogy in prodigy: Automating case acquisition, storage, and utilization. *Machine Learning*, 10(3):249–278.

Wilensky, R. (1978). *Understanding Goal-Based Stories.* PhD thesis, Yale University. Computer Science Department Technical Report 140.

Wilensky, R. (1983). *Planning and Understanding.* Addison-Wesley, Reading, MA.