Adaptive Similarity Assessment for Case-Based Explanation.*

David B. Leake
Computer Science Department
Indiana University
Bloomington, IN 47405
812–855–9756
leake@cs.indiana.edu

The International Journal of Expert Systems Research and Applications. 8(2):165–194,1995.

Key words: Case-based reasoning, similarity assessment, abductive explanation

*This research was supported in part by the National Science Foundation under Grant No. IRI-9409348. I would like to thank the anonymous reviewers for *International Journal of Expert Systems Research and Applications* for their helpful comments.

Abstract

Guiding the generation of abductive explanations is a difficult problem. Applying case-based reasoning to abductive explanation generation—generating new explanations by retrieving and adapting explanations for prior episodes—offers the benefit of re-using successful explanatory reasoning but raises new issues concerning how to perform *similarity assessment* to judge the relevance of prior explanations to new situations. Similarity assessment affects two points in the case-based explanation process: deciding which explanations to retrieve and evaluating the retrieved candidates. We address the problem of identifying similar explanations to retrieve by basing that similarity assessment on a categorization of anomaly types. We show that the problem of evaluating retrieved candidate explanations is often impeded by incomplete information about the situation to be explained, and address that problem with a novel similarity assessment method which we call *constructive similarity assessment.* Constructive similarity assessment contrasts with traditional "feature-mapping" similarity assessment methods by using the contents of memory to hypothesize important features in the new situation, and in using a pragmatic criterion—the system's ability to adapt features of the old case into features that apply in the new circumstances—as the basis for comparing features. Thus constructive similarity assessment does not merely compare new cases to old; instead, based on adaptation of prior cases in memory, it addresses the problem of incomplete input cases by building up and reasoning about augmented descriptions of those cases.

Contents

1 Introduction

Abductive explanations play a crucial role in knowledge-based systems to support tasks such as understanding (e.g., Charniak, 1988; Leake, 1992; Ng & Mooney, 1990; Pazzani, 1990; Ram, 1993; Schank, 1986, Wilensky, 1983) and diagnosis (e.g., Console & Torasso, 1990; de Kleer & Williams, 1989; Poole, 1989; Pople, 1973; Stern & Luger, 1992; Weintraub, 1991). In these tasks, a difficult question is how to generate explanations effectively. Abductive explanation systems often generate their explanations by reasoning from scratch, using backwards chaining methods (Hobbs, Stickel, Appelt, & Martin, 1993; Kautz & Allen, 1986; Konolige, 1990; Levesque, 1989; McDermott, 1974; O'Rorke, 1994; Poole, 1989; Wilensky, 1983). However, the computational cost entailed by these methods is a significant concern, as has been established both by theoretical analysis (Bylander, Allemang, Tanner, & Josephson, 1991; Selman & Levesque, 1990) and experiments applying diagnosis systems to real-world abductive tasks (Tuhrim, Reggia, & Goodall, 1991).

One approach to controling the cost of explanation generation is to use prior knowledge to guide the explanation process. Such knowledge may be provided in pre-specified schemas, as in schema-based diagnosis (Stern & Luger, 1992; Turner, 1994), or generated by reasoning from scratch and generalizing the results to store as new rules, as in explanation-based schema acquisition (DeJong & Mooney, 1986; Mooney, 1990; Pazzani, 1990). Another alternative is to store ungeneralized traces of previous explanations and adapt them as needed to generate new explanations, as in case-based reasoning (CBR) approaches to explanation (Kass, 1992; Leake, 1992b; Schank, 1986; Schank & Leake, 1989; Schank, Riesbeck, & Kass, 1994). Case-based reasoning solves new problems by a combination of retrieval of relevant stored "cases" reflecting prior problem-solving episodes, analogical mapping from those cases to the new situation, and adaptation of prior solutions to fit new circumstances (Kolodner, 1993; Riesbeck & Schank, 1989).

Successful case-based reasoning depends on being able to determine whether stored cases are relevant to current problems. This is done by performing *similarity assessment* to compare old and new situations. This article examines similarity assessment methods that we have developed for case-based explanation of anomalous events noticed during understanding, discussing the problems that they address and the ramifications of our solutions. In so doing, it also illustrates some important commitments of the case-based approach: the use of analysis of the *content* of explanations to organize memory and focus hypothesis generation, and the use of stored cases reflecting prior explanatory episodes to augment incomplete domain information. These principles apply to many case-based reasoning tasks, but applying a case-based approach to explanation raises a new problem as well: how to assess similarity between stored cases and incompletely-described input situations.

Models of CBR often judge similarity by comparing a fixed problem description, provided as input and assumed to include all important features, to prior cases. For case-based reasoning tasks such as case-based planning, this assumption is reasonable. However, when performing diagnosis or explaining anomalies during story understanding, only incomplete information about the new situation will be available—explanation is being done precisely *because* the current situation is incompletely understood. Thus for case-based explanation, the selection of the previous case to apply must be based on incomplete information. Consequently, for this task, simply comparing the features explicitly provided in an input case with those in stored cases is generally not sufficient to determine whether the stored cases apply.

This article describes a two-step process for selecting potentially relevant cases and deciding which ones apply, despite incomplete input information. The first step determines the similarity of the current anomaly to explain and the anomalies explained by prior cases stored in memory. In that step, the anomaly to explain is described by an *anomaly characterization* reflecting an *a priori* judgement of features likely to be important, and the anomaly characterization is compared with anomaly characterizations for previously-explained episodes to determine which stored explanations are relevant. Because the anomaly characterizations define a way to map between features of the old and new situations, the approach also provides the added benefit of addressing the classic *analogical mapping* problem of forming correspondences between analogous features of old and new situations.

After a stored explanation for a similar anomaly has been retrieved, the second step examines the similarity of the current situation to the previously-explained situation. The new situation and retrieved case are compared by a novel approach that we call *constructive similarity assessment,* which is the primary focus of this article. This approach differs in two ways from traditional CBR similarity assessment methods. First, constructive similarity assessment uses prior cases to decide which types of features should be investigated in the new situation, augmenting the input case description with hypothesized features. Second, its criterion for similarity is not the match between features of the situations per se, but instead a much more flexible criterion: whether the previous explanation can be adapted to a new explanation fitting the new circumstances. Thus constructive similarity assessment does not merely compare new cases to old: using prior cases as its guide, it dynamically carves augmented descriptions of new cases out of memory. Its judgment of "similarity" is pragmatic: whether new and old cases are considered similar depends directly on whether the old case helps to generate a (possibly new) explanation of the current situation.

As background for our discussion, we begin with a brief sketch of the case-based explanation process and its relationship to other approaches for generating abductive explanations. We then examine the role of anomalies in motivating and focusing explanation. We next address the two similarity assessment problems for case-based explanation of anomalous events: determining similarity of anomalies and determining similarity of explanations. We close with additional perspective on case-based explanation and on the contributions of constructive similarity assessment, both in the context of other similarity assessment methods and in terms of the commonsense reasoning issue of determining relevant features in complex situations that are only partially specified.

2 A case-based approach to explanation generation

The problem of abductive explanation is to form plausible hypotheses about the factors that account for an observed state or event. Because many competing alternatives may be generated, choosing between possible candidates is a central issue. As a result, the abductive explanation process is often characterized as "inference to the best explanation" (Harman, 1965).

In explanation systems, abductive explanations are generally modeled as deductive proofs whose premises include "abductive assumptions." For example, an explanation of a device malfunction would be a proof showing that a set of assumptions about particular malfunctioning components entailed the observed device behavior. In abductive explanation systems, explanations are usually generated by a theorem proving process that chains backwards from the event to be explained. The computational expense of this approach is a well known problem. Many responses have been proposed to help control that cost within the chaining framework, such as limiting the number of chaining steps allowed before abandoning an attempt to explain, using probability estimates to focus attention, or using a simplified domain theory (Bhatnagar & Mostow, 1990; Charniak, 1986; Chien, 1989; de Kleer & Williams, 1989; Hobbs et al., 1993; Mooney, 1990; Ng & Mooney, 1990; Norvig, 1989; Tadepalli, 1989; Wilensky, 1983). Nevertheless, in complex domains the cost of explanation generation remains a crucial concern.

An alternative approach is to use knowledge about the explanations that tend to apply to a particular domain to guide the search for new candidate explanations. Stern and Luger (1992), for example, advocate a template-based explanation method in which templates at different levels of generality guide the formation of diagnoses. Explanation-based learning (EBL) augments explanatory knowledge by generalizing explanations and storing the generalizations as new rules for future use, allowing them to be re-used without having to be re-derived (Mitchell, Keller, & Kedar-Cabelli, 1986; DeJong & Mooney, 1986).

The case-based approach also learns and re-uses information from explanation episodes. However, rather than re-using generalizations it re-uses specific prior explanations: It uses *remindings* of prior episodes as the starting point for new explanations (Schank, 1982, 1986). Remindings give very specific guidance towards candidate explanations, as illustrated by the following episode, which will be used as a running example throughout this article. In early 1994, figure skater Nancy Kerrigan was attacked after completing a figure skating practice session. Given only the limited information in initial news reports, countless possible explanations could be hypothesized for the attack. However, remindings of similar prior events suggested candidate explanations from similar contexts. For example, the *Washington Post* observed:

> This is the second time in little more than eight months that a high-profile female athlete has been attacked. On April 30, tennis star Monica Seles was stabbed during a changeover on the court at a tournament in Hamburg. Her assailant, Guenter Parche, ... stabbed Seles because he is a fan of her rival, Steffi Graf ...

This reminding focused attention on a small subset of the many possible explanations: attacks aimed at aiding one of the victim's rivals. Psychological studies support that human explainers tend to favor explanations based on remindings (Read & Cesa, 1991). The case-based approach to explanation is inspired by this human use of specific prior episodes to guide the explanation process.

2.1 Steps of the case-based explanation process

Starting from an input anomaly,[1] the case-based reasoning process for generating explanations involves the following steps:

1. **Anomaly characterization:** Categorize the anomaly to explain using a pre-defined vocabulary of anomaly classes. Based on the anomaly class, generate a knowledge structure to describe the anomaly and the surrounding context. This knowledge structure is called an *anomaly characterization*.

   For example, in the Kerrigan example, the anomaly characterization might represent that the attack was anomalous because there was no apparent motivation for it.

2. **Explanation retrieval:** Retrieve the stored explanation whose anomaly characterization is most similar to the characterization for the current anomaly, from a library of prior explanations stored in memory represented as *explanation patterns* (XPs) (Schank, 1986).[2]

   For example, in the Kerrigan episode, the attack on Monica Seles is a retrieval candidate because it shares the anomaly of being an apparently unmotivated attack, and because

[1]Existing case-based explanation systems attempt to explain each anomaly as soon as it is detected, so that only one anomaly is under consideration at any given time. As has been pointed out by Schank (1986), the conjunction of a combination of anomalies can provide useful information for focusing the generation of explanations. In the case-based framework, explanations could be indexed by combinations of anomalies with preference given to the explanation that simultaneously accounts for the largest set of anomalies, but this has not been implemented.

[2]This process presupposes a library of prior explanations to use as its starting point. In general, that library could be provided by external sources (e.g., by being told explanations or reading about explained episodes) or initially built up by backwards chaining methods that generate explanations starting from scratch (Koton, 1988). After that initial information is provided, the library is built up by storing the new explanations that the case-based explanation process generates.

it also involved a famous female athlete as its victim.

3. **Explanation evaluation:** Evaluate the similarity of the factors described in the retrieved explanation and in the current situation. Generate new anomaly characterizations for any anomalies or believability problems that arise from the new explanation.[3]

   The explanation for the attack on Seles is not completely applicable to the Kerrigan episode, because Kerrigan's attacker was not a fan. Consequently, adaptation is needed.

4. **Explanation adaptation:** If problems were found when attempting to apply the retrieved explanation to the new situation, use their anomaly characterizations to select and apply adaptation strategies for repairing the problems, and return to the explanation evaluation phase to evaluate the adapted explanation. If no applicable adaptation strategies are found, return to the retrieval phase and attempt to retrieve another candidate explanation.

   In the Kerrigan example, the explanation can be adapted by hypothesizing that instead of being a fan, the attacker was an *agent* for a fan. The details of the adaptation process and issues it involves are beyond the scope of this article; see Kass (1990, 1992, 1994).

We note that explanation retrieval is based on similarities between anomalies to explain and explanation evaluation is based on similarities between explanations. We also note that the decision of whether a case is appropriate depends not only on the initial retrieval but also on the success of explanation adaptation; if adaptation fails, alternative retrievals will be tried until one of the retrieved explanations can be adapted successfully.

This case-based approach to explanation was originally developed in the SWALE system (Kass, 1986; Leake & Owens, 1986; Schank, 1986; Schank & Leake, 1989). The methods developed in that work were further refined in the system's descendents ABE (Kass, 1990, 1994) and ACCEPTER (Leake, 1992b, 1994a). The current version of ACCEPTER is a story understanding program explaining anomalies in the domain of deaths, defects, and destruction. ACCEPTER detects anomalies in the stories it processes, characterizes the

[3]The evaluation phase also checks whether the explanation provides the information needed for tasks beyond understanding, but that evaluation is beyond the scope of this article. See Leake (1991b, 1992b, 1994a).

anomalies to facilitate explanation retrieval and adaptation, retrieves candidate explanations, and evaluates candidate explanations and candidate adaptations of those explanations. The examples it processes include the death of basketball star Len Bias (who died the day after being first choice in the basketball draft), the explosion of the space shuttle Challenger, the episode in which the American warship Vincennes shot down a civilian airliner, and the recall of a prestigious car. All these examples require explaining in a complex domain for which both input information and domain rules are likely to be incomplete or imperfect, suggesting recourse to prior experience to suggest plausible explanations.

2.2 The role of similarity assessment

The similarity assessment steps involved in case-based explanation are illustrated diagrammatically in figure 1. Two types of similarity assessment are involved: an initial similarity assessment that compares the current anomaly to be explained with previously-explained anomalies, to guide retrieval of explanations for similar anomalies; and a later similarity assessment step that compares the new situation to the situation explained by the retrieved explanation. Thus the two types of similarity assessment are done in two different spaces, the space of anomalies and the space of explanations.

As summarized previously, when ACCEPTER detects an anomaly during story understanding, the system generates a characterization of that anomaly and uses it to determine which of the anomalies explained by stored explanations is most similar to the current anomaly. Once the most similar previously-explained anomaly has been found, the explanation for that anomaly is retrieved, and attention shifts to the space of explanations. Similarity assessment within the space of explanations considers not only how well the factors mentioned in the retrieved explanation match the current circumstances, but the more flexible criterion of whether the old explanation can be adapted to fit the new circumstances. If so, the result is a new version of the retrieved explanation, adapted to fit the new situation. The following sections discuss our similarity criteria and methods for judging similarity of anomalies and similarity of explanations.

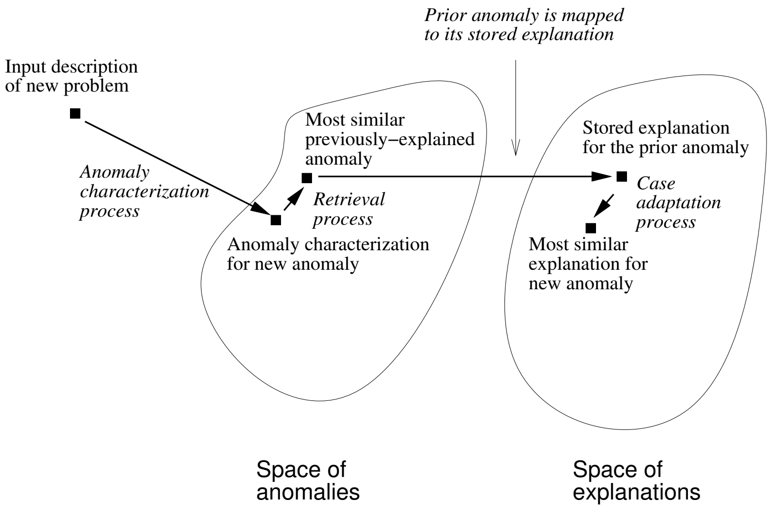

Figure 1: Using similarity of anomalies and similarity of explanations in the case-based explanation process. Comparisons between anomalies determine candidate explanations. Those explanations are then evaluated for relevant similarity to the new situation.

3 Similarity of anomalies

3.1 Why focus on anomalies?

In most AI models of abductive explanation, accounting for the state or event being explained is the sole focus of explanation (Charniak, 1986; Charniak & Goldman, 1991; de Kleer & Williams, 1989; Kautz & Allen, 1986; Ng & Mooney, 1990; Pearl, 1988; Thagard, 1989; Wilensky, 1983). In such models, two explanation problems are similar if they involve similar states or events. However, it is becoming recognized that any state or event can be explained from many different perspectives, each of which provides different information. As a result, it is important to generate explanations whose focuses are most appropriate for their intended uses (Chandrasekaran, 1994; Krulwich, Birnbaum, & Collins, 1990; Leake, 1991b, 1992b; Ram & Leake, 1991). Consequently, similarity criteria must reflect those intended uses.

The use for explanations addressed by this article is revising flawed beliefs to improve the accuracy of the understander's beliefs and expectations. In the case-based explanation systems we describe, belief revision is triggered by detection of *anomalies*—conflicts between new information and prior beliefs or expectations. This approach is consistent with a large body of AI research that uses *expectation failures* (Schank, 1982) as a trigger for learning (Collins & Birnbaum, 1988; Collins, Birnbaum, Krulwich, & Freed, 1993; Hammond, 1989; Redmond, 1992; Ram, 1993; Ram & Cox, 1994; Riesbeck, 1981; Schank, 1982, 1986). Learning in response to expectation failures can be supported on functional grounds: when expectations or beliefs turn out to be wrong, their failure reveals a defect in previous knowledge, showing the need for learning in order to have an accurate world model. For example, if an understander already knows the motivation for the attack on Kerrigan but is surprised by the choice of weapon used, what must be explained is the anomalous weapon. In this context, having the attacker's motivation re-explained, as would be done by explanation-based schema acquisition systems that always explain motivations for the event being explained, provides no useful information.

Revising beliefs in response to anomalies also parallels a number of models of human learning. For example, using the accuracy of expectations as the criterion for the adequacy of understanding is quite similar to the perspective of Kelly's personal construct theory, which treats the correct anticipation of events as a fundamental human goal and as its criterion for determining whether representations are adequate (Adams-Webber, 1989; Kelly, 1970). Likewise, research from educational psychology corroborates that the ability to detect anomalies plays an important role in triggering students' learning (Otero & Campanario, 1990).

The value of anomalies in focusing human belief revision suggests that the ability to generate explanations with the right focus for resolving anomalies will have applications to expert system explanation as well. Constructivist perspectives on expert system explanation, for example, stress the need for explanation to support the user's attempt to build up a predictive model (Ford, Canas, & Adams-Webber, 1991). Likewise, experiments by Roth, Bennet, and Woods (1987) show that the effectiveness of the combination of an expert system and its human user can be significantly improved when the human user monitors the performance of the system and has an accurate model of the reasons why the system acts as it does. Otherwise, performance may be degraded by problems such as the user misunderstanding the system's questions or misjudging the extent of its competence. Expert system explanations aimed at resolving anomalies noticed by a user could enable the user to refine predictions about the expert system's behavior, contributing to the user making more effective use of that system in the future.

3.2 Criteria for deciding the similarity of anomalies

For a case-based explanation system, two anomalies are usefully similar if the explanation for one is helpful when explaining the other. Consequently, similarity criteria for anomalies need to reflect what is necessary to explain those anomalies: two anomalies should be judged similar if they have similar explanations.

In order for an explanation to resolve an anomaly, it must address not only the state or event being explained but also what makes the state or event surprising, which changes depending on the context in which the anomaly occurs. For example, to someone who had not heard about the attack on Kerrigan, the condition of her knee in later practices would be surprising because of its impaired function, requiring an explanation of why that function was impaired. To someone who had heard about the attack, however, the condition of her knee might be surprising because it had healed faster than expected. Explaining the state of her knee in the latter context requires focusing not just on which muscles were damaged, but on the factors that affected the healing process (e.g., that because of her training, she was in exceptionally good physical condition).

In light of this example we can identify two questions that an explanation of an anomaly must answer (Leake, 1992b, 1994a). The first is why the surprising event (or surprising features of the event) occurred. That question is the one answered by the explanations generated by traditional abductive explanation systems (Charniak, 1986; Charniak & Goldman, 1991; Charniak & Shomony, 1994; Hobbs et al., 1993; Josephson & Josephson, 1994; Kautz & Allen, 1986; Konolige, 1990; Levesque, 1989; McDermott, 1974; O'Rorke, 1994; Mooney, 1990; Poole, 1989; Wilensky, 1983).

The second question is why there was the observed difference between actual events and expectations. Answering this second question requires providing an account of the flaw in reasoning that led to the reasoner's erroneous belief or expectation (e.g., that standard expectations for the rate of healing do not apply to athletes in peak condition). Both parts of the explanation are needed in order to update the understander's world model and improve its ability to avoid erroneous predictions in the future. Answering the first question identifies important features that were overlooked, updating current understanding; answering the second helps point out how to avoid forming similar faulty predictions in the future.

The need for an explanation to simultaneously account for the event and for a flaw in reasoning suggests that explanations of anomalies have three components. The first is a derivation of how the surprising features of the new situation follow from background knowledge and abductive assumptions. The second component is a description of the reasoning that led to the erroneous previous belief or expectation, that is, of the reasoning chain on which the erroneous prior belief or expectation was based. The third component relates these two, explaining why the prior reasoning does not apply and the new reasoning does: Why the new conclusion should hold and the previous one should not.

For example, suppose that the anomaly to be addressed is that an athlete was attacked after practice, despite expectations for total safety because of tight security. In that case, the first component would be a reasoning chain accounting for why the attack could have been expected to happen (e.g., including information about a motive and about the enabling condition of an unlocked door). The second component would be an account of why the attack was considered impossible (e.g., because all entrances were guarded and the guards were expected to bar anyone without a special pass, to prevent unauthorized access). The third is the reason that the first reasoning chain (which showed factors supporting the occurrence of the attack) takes precedence over the reasoning chain in the second component (which led to the belief that the attack was impossible). For example, the third component might point to the fact that the first chain contained the assumption that all entrances were guarded, but the second reasoning chain included the new information that one entrance was actually unattended. Our anomaly characterization scheme provides the information needed to retrieve explanations of similar anomalies: it reflects what actually happened (which needs to be accounted for), and what was expected (whose absence must be accounted for).

3.3 Characterizing anomalies

Our anomaly characterization scheme represents anomalous situations using a fixed set of features for each *anomaly category* in a vocabulary of anomaly types that we have defined. The same characterization scheme is used to represent *what needs to be explained* in a new anomalous situation and *what is explained* by explanations already stored in memory, allowing it to be used to guide retrieval of stored explanations by comparing anomaly characterizations.

The aim of the characterization scheme is to assure that any two anomalies that are characterized in similar ways are likely to have similar explanations as well—i.e., that the explanation for the previous anomaly can easily be adapted into an explanation for the anomaly at hand. When that property holds, only the anomaly characterizations need to be considered, rather than the explanations themselves, when retrieving candidate explanations for a new situation.

The anomaly characterization scheme used in ACCEPTER represents two aspects of each anomaly. The first is the *event* that was anomalous. The aim of including the event in the anomaly characterization is to make sure that, for example, an anomaly involving an attack will be considered similar to other anomalies involving attacks. The second aspect of the anomaly that is represented is the *conflict* between the event and the prior reasoning that makes the event anomalous.

ACCEPTER's anomaly characterizations are represented using an anomaly category and a frame-like knowledge structure filled in with specific information about the situation. For each anomaly category, ACCEPTER's associated knowledge structure includes a pre-defined set of roles that reflect an *a priori* estimate of which features of the situation are likely to be relevant to explaining the anomaly. We discuss those knowledge structures in the following section and then sketch the anomaly categories.

3.3.1 The structure of anomaly characterizations

Associated with each anomaly category is a knowledge structure for representing specific information relevant to anomalies of that type. The anomaly categories and knowledge structures have been developed based on examination of approximately 200 anomalies informally collected from computer science graduate students. Thus the categories reflect the content of a set of commonly-reported types of everyday anomalies and explanations.

As an example, one of the anomaly categories, SURPRISING-PLAN-CHOICE, groups anomalies that arise when an actor selects a plan that conflicts with the understander's expectations.

Role	**Filler**
Goal	Proximity to Kerrigan
Actor	Kerrigan's attacker
Plan	Enter athletes' area
Blocked step	Show pass to guard
Impediment	UNAVAILABLE-ROLE-FILLER: Pass

Table 1: The characterization structure for a BLOCKED-PLAN anomaly.

Associated with the SURPRISING-PLAN-CHOICE anomaly category is a knowledge structure with roles for information that is often relevant to describing anomalies involving surprising choices of plans: the *actor* who is doing the planning and the *goal* for which the actor is planning, the surprising *plan* chosen, and the *conflict* between the surprising plan and expectations. The reason this information is included is that when explaining a surprising plan choice it is often useful to consider other choices of plans that were surprising in similar ways and under similar circumstances: when explaining the apparently unmotivated attack on Nancy Kerrigan, it is useful to consider the explanation for the attack on Monica Seles.

A more specific sub-category of the SURPRISING-PLAN-CHOICE anomaly category is BLOCKED-PLAN, which describes instances of SURPRISING-PLAN-CHOICE in which the selection of the plan was surprising because the plan appeared to be blocked. The knowledge structure for describing BLOCKED-PLAN anomalies has roles to be filled by the actor, by the goal of the chosen plan, by the chosen plan itself, by the step that is blocked, and with a description of what made the plan appear to be blocked: the condition that was expected to prove to be an impediment to executing the plan. The combination of the anomaly category and the filled-in knowledge structure describing the circumstances is called an *anomaly characterization*. Table 1 provides an example of the characterization of a particular BLOCKED-PLAN anomaly involving Kerrigan's attacker entering the athlete's area of the arena where she was practicing, even though that should have been prevented by the attacker lacking the proper pass.

3.4 A vocabulary of anomaly categories

ACCEPTER's vocabulary of anomaly categories is organized hierarchically. At the top level, the categories reflect only the domain of knowledge involved in the reasoning that led to the faulty expectation (e.g., knowledge about planning, as in SURPRISING-PLAN-CHOICE anomalies, or knowledge about biological processes such as the healing process, as in PROCESS-EXECUTION-FAILURE). Reflecting the domain of knowledge involved in the flawed reasoning is important for guiding explanation, because different explanations are needed to account for failures of expectations based on reasoning involving different domains of knowledge. For example, the single attack on Kerrigan might be anomalous because it conflicts with knowledge and expectations about normal planning (e.g., if it seems to be an unmotivated act or one unlikely to succeed), requiring an explanation focusing on the attacker's motivations. Alternatively, it might be anomalous because it conflicts with knowledge and expectations about physical injuries (e.g., if the blows caused less damage than expected), requiring an explanation addressing the physical effects of the blows. Table 2 enumerates the domains of knowledge reflected by ACCEPTER's anomaly categories, shows an example for each one, and lists the item in ACCEPTER's anomaly vocabulary for that anomaly category.

At the next level of the hierarchy, the categories reflect both the domain of knowledge involved and how the input situation differed from expectations. For example, categories such as SURPRISING-PLAN-CHOICE are sub-divided according to why the choice of a plan is surprising, with different subcategories for choices that are surprising because the chosen plan is expected to fail (BLOCKED-PLAN), because there is no need to carry it out because the goal will be accomplished anyway (REDUNDANT-PLAN), or because the plan, though otherwise reasonable, simply conflicts with the actor's normal plan selection preferences, as in the case of the choice of a laborious plan by someone known to be impatient, when faster plans are available (PREFERENCE-FAILURE). Figure 2 illustrates some anomaly categories and their specifications.

ACCEPTER's anomaly categories were developed to describe the types of anomalies arising in stories of everyday events. If explanation were being done in a specialized task domain, additional subcategories could be generated as appropriate for the task, resulting in

Domain of knowledge underlying expectation	**Related anomalies and vocabulary of anomaly categories**
Plan choice	From knowledge of the modus operandi of Kerrigan's attacker, another criminal might have expected the attack to be done in a more secluded place. Conflicts with expectations about plan selection are instances of the anomaly category SURPRISING-PLAN-CHOICE.
Plan instantiation	The criminal might also have expected the attacker to use a knife instead of a club because knives are more effective for disabling a victim. The category for anomalies involving use of unusual tools or other objects is SURPRISING-PROP-CHOICE.
Plan execution	A confederate might have known of the plan in advance and known that the attack was scheduled to be carried out earlier. The category for explanations of deviations in a plan's execution details, once it has begun, is PLAN-EXECUTION-FAILURE. If the confederate previously expected the plan to be blocked (*e.g.*, by security at the rink), and it succeeded anyway, the explanation would also have to account for the BLOCKAGE-VIOLATION.
Progressive processes	A fan would not have expected the knee to recover from the attack, resulting in a later anomaly when the knee did. The category for deviations from models of physical, biological or chemical processes, such as the healing process, is PROCESS-EXECUTION-FAILURE.
Device models	If metal detectors were in use at the rink, they might have prompted the expectation that no weapons could be brought into the area. In that case, their failure to function as expected would have been a DEVICE-FAILURE anomaly.
Inductive generalizations	If the skates worn by a skater were unusual compared to others encountered by the understander, their deviations would prompt UNUSUAL-OBJECT-FEATURE anomalies. When such generalizations concern the persistence of features (McDermott, 1982), that anomaly would be an instance of a STRANGE-FEATURE-CHANGE anomaly.
Information transmission	After the attack, a fan might have heard a rumor that the attack broke Kerrigan's leg and then heard a correct report involving less damage. In that case, the anomaly would be that the original source was unreliable, causing a BAD-INFORMATION anomaly.

Table 2: Domains of knowledge on which expectations are based and the categories of anomalies that result when the expectations fail.

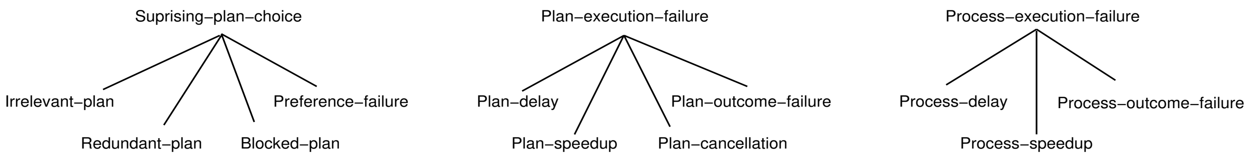

Figure 2: Subcategories for three anomaly categories.

additions to its tree of anomaly categories. For example, for diagnosing failures of electronic devices, the general DEVICE-FAILURE anomaly could be subdivided according to classes of devices that tend to have similar failures, and according to the symptoms of different failure modes.

Note that it is possible for more than one anomaly to arise when understanding a single situation: For example, the choice of a plan might be anomalous both because it is a surprisingly expensive plan for a stingy actor to have chosen (causing a PREFERENCE-FAILURE anomaly) and because it is expected to fail (causing a BLOCKED-PLAN anomaly). In such cases, an explanation or combination of explanations resolving all the anomalies is needed. For a discussion of how ACCEPTER determines whether an explanation resolves a particular anomaly, see Leake (1992b).

3.5 Using the categories to organize explanation memory

In ACCEPTER's memory, stored explanations are organized hierarchically by the categories of anomalies that they explain. When an explanation must be retrieved for a new anomaly, ACCEPTER first identifies the anomaly category involved (by checking the domain of knowledge leading to the expectation that failed). It then generates an instance of the knowledge structure associated with the anomaly category and fills in the roles of that knowledge structure. For example, if the anomaly to be explained were Kerrigan's attacker entering the restricted area of the arena without a pass, because it was not believed that he would be allowed in by the guard, the anomaly would be described as a BLOCKED-PLAN anomaly and the roles illustrated in table 1 would be filled to generate a description of the anomaly. That description would then be matched with other instances of characterizations for BLOCKED-PLAN anomalies to select the one whose role-fillers matched most closely, defining a similarity relationship between anomalies. For the example, this might enable retrieval of the information that a friend of the attacker who was helping plan the attack had previously been able to enter because of a forged pass. Similarity of role-fillers is computed according to a simple distance measure based on the distance between them in a semantic network.

4 Using anomaly characterizations to determine analogical mapping

Retrieval of a stored explanation provides a potential analogue for the anomalous situation to be explained. That analogue can be used to suggest new explanations by analogy. However, doing so requires addressing a central issue in analogical reasoning: how to perform the mapping between two situations. A number of mechanisms have been developed for analogical mapping (Falkenhainer, Forbus, & Gentner, 1989; Holyoak & Thagard, 1989; Hofstadter & Mitchell, 1993; Keane, Ledgeway, & Duff, 1994; Mitchell, 1993); one has even been applied to the task of generating explanations by analogy (Falkenhainer, 1990). In general, determining the appropriate mapping is a difficult problem, given the many correspondances that might apply. Our case-based explanation method obviates the problem by not doing complicated reasoning to determine mappings. Instead, it uses the anomaly characterization structures to make a simple mapping process practical.

As described in the previous sections, each explanation stored in ACCEPTER's memory is associated with an anomaly characterization structure. At the time that the explanation is stored in memory, the slots of its anomaly characterization structure are filled with the corresponding objects in the stored explanation. When a new anomaly is being explained and the explanation retrieval process determines that a stored explanation has a similar anomaly characterization, the matching roles in the old and new anomaly characterizations provide a correspondence between the old and new situations.

For example, if the explanation for the attack on Monica Seles were retrieved as relevant to the Kerrigan attack because both were characterized as involving apparently senseless attacks on athletes, each would be associated with anomaly characterization structures involving the same list of slots. The correspondence between the lists of slots in the anomaly

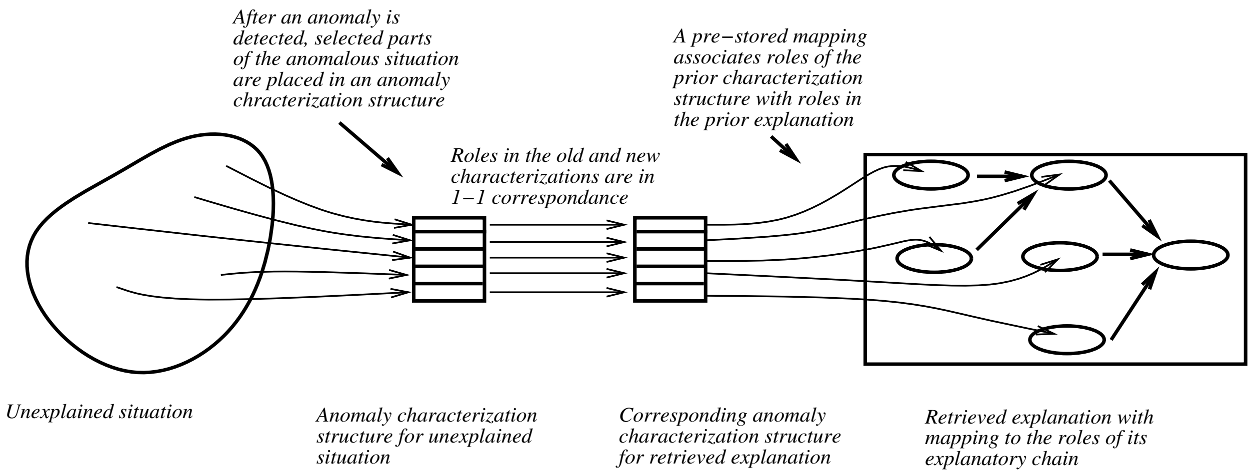

Figure 3: Case-based explanation generation addresses the analogical mapping problem by mapping corresponding role-fillers in anomaly characterization knowledge structures.

characterizations directly provides a mapping between the two events without further processing, by mapping the corresponding slot-fillers. This mapping process is illustrated in figure 3. Thus the anomaly characterization structures provide a way to efficiently determine the correspondences between a retrieved explanation and a hypothesized explanation for a new situation.

The previous sections show that comparing anomaly characterizations is a useful method for assessing the similarity of anomalies and for facilitating mapping between old and new situations. However, no method relying solely on comparing features selected *a priori* (as in the anomaly characterizations) can be guaranteed to capture all the important features for explaining a situation. Consequently, after retrieving a candidate explanation based on similarity of anomalies, a second form of similarity assessment is done in the space of explanations, to judge the similarity of the features of the new situation to those identified as important by the retrieved explanation itself. In the following section we discuss the problems of *a priori* feature selection and traditional feature-matching similarity assessment methods for the explanation task. We then discuss our solution: a constructive similarity assessment method that dynamically determines feature relevance at the same time that it adapts features of the retrieved case into hypothesized features of the new situation.

5 Constructive similarity assessment: Judging similarity of retrieved cases

5.1 Assessing similarity despite partial information

For many applications of case-based reasoning, determining similarity between old and new cases can be done using only a straightforward matching of features in the old and new cases. For example, when case-based reasoning is applied to planning (to generate new plans by retrieving and adapting previous plans), the input to the planner normally describes all relevant goals and constraints. Consequently, determining the usefulness of an old plan can be done by comparing the given goals and constraints to the goals achieved by the retrieved plan and the constraints that the retrieved plan satisfies (Alterman, 1986; Hammond, 1989; Hinrichs, 1992; Kolodner, 1987). Many other case-based reasoning systems determine similarity by a simple feature matching method, in which the value of each feature in the representation of an input case is compared to the value of each corresponding feature in a retrieved case, and an aggregate match score is computed by applying some fixed weighting to the degree of match found for each feature, as discussed by Kolodner (1993, pp. 354-358).

A case-based reasoning system with the task of explaining anomalies, however, cannot rely on such a method because its input case descriptions may be incomplete. In fact, it is intrinsic to the abductive explanation task that instead of an abductive explainer having complete information about the situation it explains, the opposite is usually true: explanation is needed precisely *because* information about the current situation is incomplete and inadequately understood. As a result, if similarity assessment is done by matching known features of the new situation to the features of retrieved cases, it is likely to fail to give good results because of gaps in the current case's description. For example, in the context of diagnosis, there is no guarantee that the features relevant to diagnosing a device failure will be included in the information provided to the explainer—it is impossible to be certain about which features are relevant to the explanation until the explanation has been generated.

It might appear that the problem of incomplete input case descriptions to a CBR system could be solved by preprocessing the new case. If the input case can be elaborated to select important features and fill in missing information, the case-based reasoning process can apply traditional similarity assessment procedures to the resulting case description. However, a circularity vitiates this method: the preprocessing phase of identifying the needed features would need to analyze the current situation in order to select the features to fill in, and that analysis of the current situation involves the entire explanation problem that the system originally needed to solve.

The previous section on anomaly characterization describes one approach to this problem: to simply pre-select features that are often relevant to explaining a given type of anomaly and compare those features to the features in a retrieved explanation when deciding its relevance. Other domain-specific knowledge can also be used to decide features to consider. For example, it seems reasonable to have the input case description for an engine problem include features known to commonly be relevant to explaining engine problems, such as engine noises, and to have the case description omit miscellaneous circumstances (e.g., the particular errands that the driver may have been doing before the problem occured, etc.). Knowledge of the features that are likely to be relevant can also be used prior to retrieval to guide collection of important features to represent in a case description; this process is called *situation assessment* (Kolodner, 1993).

Unfortunately, however, unexpected aspects of cases may also be relevant to generating an explanation, as shown by the following anecdote from a book on automobile repair:

> An elderly customer told a mechanic that her car's starting depended on the type of ice cream she bought: whenever she bought peppermint the car refused to start, but whenever she bought vanilla it started perfectly. The mechanic knew that peppermint could not affect the car's starting, but humored her with a test drive to buy ice cream. She parked at the store, went inside and bought a pint of vanilla; the car started perfectly. She drove around the block a few times, parked and bought peppermint. When she came out, the engine would not start.
>
> (Porter, 1973, pp. 253-254)

In this example, contrary to all *a priori* expectations for which features should be relevant, the errand being done *was* important.

In this example, the explanation for the bizarre correlation is as follows. At the ice cream store in question, popular flavors such as vanilla were prepackaged, but each order of peppermint was packed to order by hand. Consequently, when a customer ordered vanilla it could be delivered immediately, but when a customer ordered peppermint, there was a delay while the order was packed. The time required for hand packing allowed fuel from the carburetor to trickle into the engine, resulting in a flooded engine.

This example illustrates the difficulty of deciding which features of a situation may be relevant to explaining it. Descriptions of a situation that are provided to an explanation system may omit important aspects of the situation: In this example, the correlation between the flavor of ice cream and the problem might easily have been overlooked. Even when that correlation is mentioned, extensive inferencing is needed to determine why it might in fact be relevant. The difficulty of determining relevant features in advance seriously complicates the problem of deciding the similarity of an input case and a previously-explained situation.

We address this problem with a novel similarity assessment method that we call *constructive similarity assessment* (Leake, 1992a). Rather than simply matching known features of old and new cases, constructive similarity assessment uses prior cases to determine which features of an input situation to examine. Using prior cases to provide guidance, it hypothesizes possible features of incomplete case descriptions, filling in missing details. In constructive similarity assessment, the absence of a feature in an input case description is not taken as evidence against the feature being present, but rather as indicative of the need to do more reasoning to determine whether that feature is a plausible one to hypothesize or whether that feature can be adapted into one that *is* plausible. Thus the process treats an old case as similar to a new one provided that either the new situation can be elaborated to match the old explanation or that the old case can be adapted to fit the new situation. We describe this process in the following sections.

5.2 Using memory to build up cases during similarity assessment

Constructive similarity assessment addresses the problem of missing information in new case descriptions by using the retrieved case, which already had its relevant features identified by its explanation, to suggest features to look for or infer about the new situation. When a retrieved explanation identifies some features of a situation as potentially important, and those features are not already included in the description of the new situation, ACCEPTER uses prior knowledge in an attempt to infer the values of those features in the current situation.

For example, consider again attempting to explain the attack on Kerrigan. Once the explanation for the attack on Monica Seles has been retrieved as a candidate explanation, the Seles explanation can guide what to look for in the Kerrigan episode. One feature involved in the explanation for Seles, but left unspecified in the Kerrigan story, is whether the victim of the attack had an important upcoming competition. The fact that Kerrigan has an upcoming competition may be inferred, however, from prior knowledge (e.g., that Kerrigan is an Olympic skater and that 1994 is an Olympic year). Once that information has been inferred, it augments the original input description of the Kerrigan episode and shows another commonality with the situation of the Seles attack. This suggests further aspects of the attack on Kerrigan to consider to determine whether the Seles situation is similar: whether the attack might have been done by a fan to aid one of Kerrigan's competitors in the winter Olympics. In this case the old case is not quite applicable: Kerrigan was not attacked by a fan but by a hired assailant. Once that conflict is known, additional adaptation is needed (e.g., by suggesting that the attacker was a hired assailant). The result of this process is not just an evaluation of the similarity between the input case description and a case description in memory, but an *elaborated picture* of the input case, as diagrammed in figure 4.

In general, a constructive similarity assessment process could elaborate an input case description in either of two ways: by inferring additional features of the input situation, as in the Kerrigan example, or by taking action to seek information externally (e.g., by generating questions in the Kerrigan example, or by reading gauges or performing experiments during device diagnosis). ACCEPTER relies solely on inference to augment its cases, but mechanisms for extending the model to combine inference with action in the world are currently being developed (Leake, 1994b, 1995; Sooriamurthi & Leake, 1994).

5.3 Adaptability as a similarity criterion

Even after a retrieved case has suggested a set of features to consider, constructive similarity assessment does not reduce to simple feature matching between the retrieved case and the

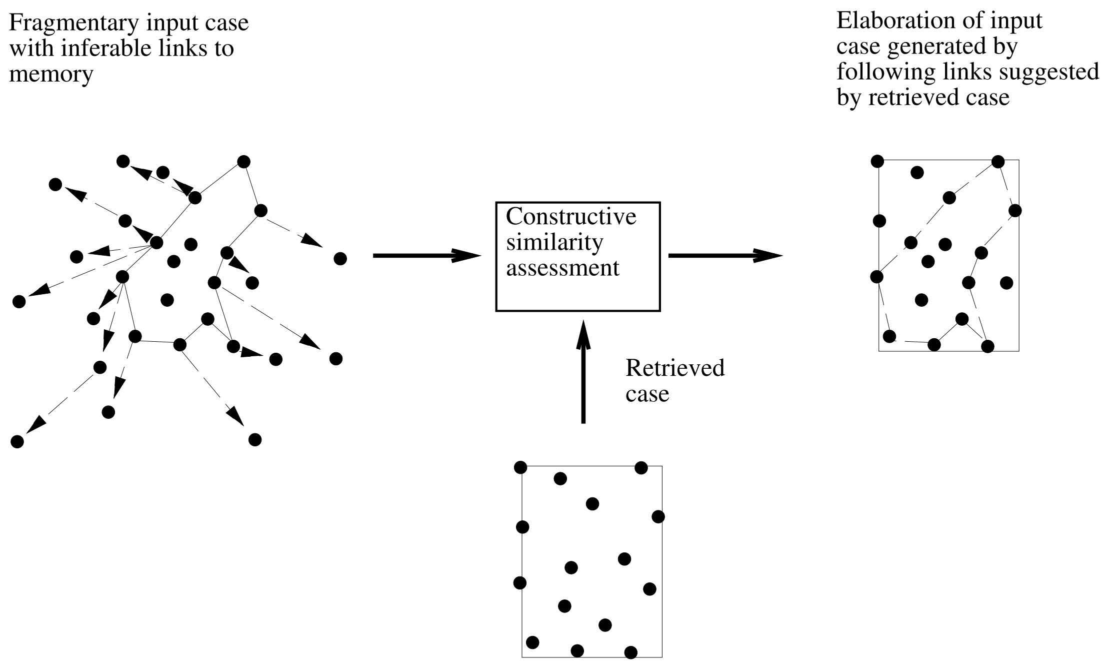

Figure 4: How constructive similarity assessment elaborates new cases based on experience.

new situation. Instead, it uses more flexible criteria: The appropriateness of the retrieved case is determined entirely by its ability to suggest a believable explanation for the new situation. Consequently, two situations may be considered similar despite conflicts between their concrete features, provided that more abstract similarity relationships apply.

For example, suppose that the anecdote connecting peppermint ice cream to starting problems were already in memory and was retrieved as potentially relevant to a new engine problem. With constructive similarity assessment, the judgment of the ice cream story's relevance to the new situation would not depend on whether the new situation involved similar concrete features (e.g., shopping for ice cream), but instead whether it involved features that were *causally equivalent* to the ice cream purchases with respect to the engine problem: events causing delays that could be correlated with the new engine problem.

Thus unlike traditional similarity assessment methods, constructive similarity assessment does not require that the same predicates apply in order to consider two situations similar; they must only match sufficiently to be able to adapt the retrieved explanation to the new situation. Accordingly, similarity assessment need not actually evaluate the match between individual features of the old or new cases at all. What is important is whether, with guidance from the retrieved case, the features of the new case can be adapted into a coherent picture of the current situation. Constructive similarity assessment treats a retrieved case as usefully similar to the new situation, regardless of how dissimilar its individual features might be, if an adaptation of the retrieved case provides a suitable elaboration of the new situation.

Because this method makes similarity assessment depend strongly on the ability to hypothesize elaborations of new cases and to judge the reasonableness of possible elaborations given current knowledge, the judgments made by constructive similarity assessment depend on the contents of memory and the explainer's background domain knowledge. In addition, the results depend on which adaptation strategies the CBR system has available to use when applying a retrieved explanation to a new situation.

5.4 An example from SWALE

Constructive similarity assessment is useful whenever input cases are incompletely described and adaptation is needed to apply old cases to new situations. However, the most dramatic examples of its benefit come in the context of creative explanation, a task for which significant adaptation of existing explanations is necessary. To illustrate the process, we sketch one of the first applications of the method to creative explanation, the namesake example for the SWALE system. That example concerns the story of a racehorse named Swale, a star 3-year-old racehorse who collapsed and died without warning a few days after winning the Belmont stakes.

The input story of the death that is provided to SWALE is quite incomplete, providing only the amount of information contained in newspaper headlines: it includes only Swale's age, the fact that Swale won an important victory at Belmont, and the fact that he died a few days later. Consequently, there are many explanations in memory that, from a feature-matching point of view, might match the few supplied features of the input case equally well. To determine the explanation to apply, the system first uses its anomaly characterization process to focus retrieval on similar anomalies (i.e., episodes involving other deaths that were surprising because they were premature). It then uses constructive similarity assessment to hypothesize a more complete description of the new situation and make a more refined similarity judgment between Swale's death and the retrieved cases. In that process it hypothesizes unstated features of Swale's death and then uses knowledge of likely events to evaluate those hypotheses. Unlike methods that treat matches between given features in old and new cases as the key determiners in the relevance of a retrieved case, what is important to this process is simply whether consideration of the old case gives rise to a coherent scenario for the new situation (possibly quite different from the retrieved episode).

For example, one of the explanations that SWALE retrieves for Swale's death is the explanation for the death of the rock star Janis Joplin. Joplin died from an overdose of recreational drugs taken to escape stress, and few features of this explanation match at a literal level. However, through SWALE's adaptation process, the suggestion of a drug overdose from the Joplin explanation leads to a plausible alternative cause of death more directly connected to horses: that Swale might have died from an overdose of performance-enhancing drugs. Another retrieved explanation concerns the death of the runner Jim Fixx, who died because jogging overtaxed a hereditary heart defect. That explanation again fails to match at a literal level (horses are not joggers, and the input case provides no information about a heart defect). Nevertheless, that explanation can also suggest a reasonable hypothesis: that Swale's racing overtaxed a heart defect, leading to heart attack and death. SWALE uses each of these suggestions to guide its search for explanations.

The features implicated in the new explanations that SWALE generates (death from performance-enhancing drugs and death from racing with a heart defect) are not part of the input case presented to SWALE. Consequently, the resultant explanations would be likely to receive low rankings according to traditional similarity assessment criteria. However, constructive similarity assessment determines that the old explanations are in fact relevant: they suggest hypotheses about Swale that can reasonably be added to the input case representation, even though they do not follow from the input information. In that way, the result of similarity assessment is not just a judgment of similarity but a larger picture of the new case that is a plausible elaboration and that shares important features with the retrieved explanations.

For example, when applying the adapted version of the explanation for Joplin's death to Swale, the use of performance-enhancing drugs fits stereotypes about horse racing in the program's background knowledge, supporting that Swale's trainer might have given him a drug overdose; when applying the adapted version of the explanation for Fixx's death, a heart defect could follow from inbreeding, which is associated with purebred animals in the program's background knowledge, also providing support. The two candidate explanations suggest that the hypothesized features of the situation—performance-enhancing drugs in one case, or a heart defect in the other—are links to be pursued to form a more complete picture of the input case that can be either confirmed or refuted. By following such links and considering their ramifications, constructive similarity assessment uses prior cases in memory to incrementally build up a richer picture of a sketchy input case, eventually leading to a plausible candidate explanation.

6 Perspective

6.1 Relationship to other methods for explanation

Approaches to abductive explanation and diagnosis traditionally focus on content-neutral explanation generation methods such as the backwards chaining (e.g., Hobbs et al., 1993; Kautz & Allen, 1986; Konolige, 1990; Levesque, 1989; O'Rorke, 1994; Poole, 1989; Wilensky, 1983) and marker passing (e.g., Charniak, 1986; Norvig, 1989). The case-based approach, which relies on defining a vocabulary of anomaly types and using that vocabulary to guide consideration of candidate explanations, instead uses information reflecting the content of the task domain to organize memory and guide explanation generation. This use of concrete content is consistent with psychological experiments by Blessing and Ross (1994) showing that the concrete content of problem descriptions is influential in guiding human problem categorization for both novice and expert human problem solvers.

In its emphasis on concrete explanations, case-based explanation also contrasts to the other method that has been advanced for re-using explanatory effort, explanation-based schema acquisition (DeJong & Mooney, 1986; Mooney, 1990; Pazzani, 1990; Segre, 1988). The emphasis in case-based explanation on using the most specific available knowledge is motivated by the desire to explain reliably despite an imperfect domain theory: it uses similarity between episodes as a criterion to supplement to the conditions on domain rules when deciding whether an explanation applies, assuming that in a regular world, similar anomalies have similar explanations. In domains that are not precisely codified, such similarities can be important to help distinguish between competing alternatives that are consistent with an incomplete input case description, as they were when attempting to explain the attack on Nancy Kerrigan. Some previous research has addressed the problem of imperfect domain knowledge by refining flawed domain theories (Dietterich & Flann, 1988; Ourston & Mooney, 1994; Pazzani, 1990; Rajamoney, 1993), but case-based explanation addresses an orthogonal issue: how to reason as effectively as possible from a current set of domain rules, given that those rules may be imperfect and input information may be incomplete.

The case-based explanation model also differs in dynamically focusing its effort on resolving a current anomaly, rather than simply accounting for a state or event from a predetermined perspective (in existing systems for explanation-based schema acquisition, that perspective focuses on accounting for the actor's motivations). As discussed previously, explanations from a fixed focus may not satisfy an explainer's needs for information.

In addition, the way that case-based explanation re-uses explanations is more flexible than explanation-based generalization. EBL systems immediately generalize new explanations as far as licensed by the rules used, and the generalization is applied to future episodes that satisfy all the conditions of the generalization. Although case-based explanation adapts explanations only when required to do so to account for new anomalies, it may perform a wider range of modifications than explanation-based generalization alone, including deletion, addition and substitution of selected components of the explanation. This allows re-using relevant sub-parts of existing explanations in a flexible way. For a more complete comparision of case-based explanation to other models and analysis of its benefits, see Leake (in press).

6.2 Relationship to other methods for similarity assessment

Similarity assessment has been the subject of many studies in the context of other case-based reasoning tasks, and the basic principles used here to develop anomaly characterizations have been applied in a number of other contexts as well. (See Bareiss & King, 1989, and Kolodner, 1993 for two surveys of current approaches). Likewise, CBR systems for other tasks have used a two-step retrieval/evaluation process that selects a set of candidate cases and filters the results, as in PARADYME (Kolodner, 1989) and STAIN (Simoudis & Miller, 1991).

However, the constructive similarity assessment process used for ACCEPTER's second phase of similarity assessment differs markedly from most similarity assessment methods.

Most similarity assessment methods share a common property: they compare the features of a retrieved case to the features provided by a pre-determined representation of the input case. That method is reasonable in domains for which input cases are known to provide all the information that needs to be considered during similarity assessment. For example, the standard input to a planner or problem-solver is a set of goals and constraints; descriptions of those goals and constraints provide the essential information that case-based reasoning systems such as CHEF (Hammond, 1989) and JULIA (Kolodner, 1987; Hinrichs, 1992) need in order to judge the similarity of new problems to prior cases stored in their memories. Likewise, in domains such as legal reasoning, input cases are routinely described in legal briefs that include all relevant features of the situation under consideration; those input cases provide all the features that need to be considered during similarity assessment by CBR systems such as HYPO (Ashley, 1990; Ashley & Rissland, 1987) and GREBE (Branting & Porter, 1991). In contexts such as these, for which input cases are guaranteed to include all relevant features, traditional similarity assessment (which simply compares the new case's features with features of a stored case) is appropriate.

However, for task domains in which input information may be incomplete, such as explanation and diagnosis, a different approach to similarity assessment is needed. This is especially true in CBR systems for which the set of cases in the system's case library does not span all alternative possibilities, making adaptation essential to solving domain problems. In such domains, constructive similarity assessment offers important advantages by providing a useful new perspective on four issues: what to include in the representation of an input case, the relationship between feature matching and similarity, the relationship between similarity assessment and case adaptation, and how changing circumstances affect criteria for similarity assessment.

What to include in a case representation: Rather than requiring a new case to be completely specified in advance, constructive similarity assessment uses retrieved cases to suggest new features to hypothesize and add to the input case description. This method can guide internal inference, as in ACCEPTER, or could enable case-based reasoning systems to assist a user in building up a representation of the problem being addressed, by identifying types of features that may be of interest even if their role was not apparent initially.

Using prior cases to guide decisions about how to augment case representations provides significant benefits. If the decision of what to include in the description of a new anomalous situation is made entirely before retrieval, as in most case-based reasoning models for other tasks, there is no way to winnow the features that are important (in light of the reasoner's experience) from the countless aspects of the situation that might *potentially* be relevant. However, when the decision of what to include is made in light of a prior case, a much smaller set of features must be considered (those implicated in the previous explanation) and those features are guaranteed to give useful information: their presence or absence helps confirm or disconfirm the applicability of the prior explanation that the system is attempting to apply.

Some recent case-based reasoning research also uses an examination of the cases in memory to address the question of which features to consider when doing case retrieval (Cunningham, Bonzano, & Smyth, 1995; Owens, 1994). In these approaches, the retrieval process collects all cases matching the given input case and then requests that the user supply information about features that it selects as most useful to discriminate between the retrieved candidates. The spirit of this approach is similar to constructive similarity assessment, but with a key difference: In constructive similarity assessment the final similarity judgement is determined not by how closely the features of the new situation match the features stated in the cases in memory, but by the system's ability to adapt the old features to form a reasonable picture of the new situation.

The relationship between feature matching and similarity: When traditional case-based reasoning systems determine similarity, their judgement is based on the match between values of corresponding features in the old and new cases. Their criterion for matching feature values is that the same predicate for a feature applies to both the old and new situations. (They can accept cases as similar when only some of the features match; explanation-based learning systems are less flexible, re-applying stored explanations only if *all* the predicates match.) Constructive similarity assessment, however, uses more flexible similarity criteria than strict feature matching, allowing old cases to be applied to new situations not only if they match the facts of new situations, but if their features are related on a much more abstract level.

The case-based reasoning systems GREBE (Branting & Porter, 1991) and PROTOS (Bareiss, 1989), also go beyond requiring literal feature matches. Both those systems decide similarity between features in a new case and a retrieved case by generating explanations of why particular features are relevant and seeing if the same explanations apply to both (e.g., PROTOS can accept that the legs of one chair match the pedestal of another, because in both cases their role can be explained as serving the same function of supporting the seat). Likewise, CASEY (Koton, 1988) accepts new features in an input situation if they support the existence of states mentioned in the old case, and accepts the absence of features that were not used in the previous explanation.

Constructive similarity assessment, however, goes beyond even flexible comparison between new features and the features provided in an input description. In fact, retrieved cases may suggest hypothesizing new features that may not be present in either the old case or the input information. For ACCEPTER, old cases are applicable if *adaptations* of their features yield a picture of the new situation that fits other knowledge in memory. For example, the explanation of a recreational drug overdose is considered relevant to the explanation of Swale's death because it can be adapted to suggest a *new* picture of the death that makes sense in light of the system's prior stereotypes for horse racing—that Swale was drugged by his trainer—even though no information about drugging was supplied by the input. Thus the ability to suggest useful features, rather than feature matching itself, forms the basis of constructive similarity assessment.

The relationship between similarity assessment and usefulness: When a CBR system retrieves a stored case, the functional value of that case is its ability to suggest a solution to a new problem. Consequently, from a pragmatic viewpoint, the purpose of doing similarity assessment in CBR is to determine the parts of a retrieved case that have to be adapted to fit a new situation and determining how difficult the retrieved case will be to adapt to fit the new situation.

Despite the pragmatic importance of linking similarity assessment criteria to the capability to adapt new cases, the similarity criteria of CBR systems are generally not directly linked to their adaptation knowledge. The criteria instead simply match the features of pre-determined case descriptions of the new and retrieved cases, weighting particular matches according to a pre-determined estimate of their importance for determining ease of adaptation. In systems such as PARADYME and CHEF, for example, the importance of ease of adaptability is recognized as a central concern, but is judged according to fixed knowledge which features are generally hard to adapt.

The constructive similarity assessment framework described in this article instead directly links similarity judgments to adaptation procedures: two cases are similar precisely to the extent that the solution of one can be adapted to fit the other. This gives more flexibility and a more perspicacious judgement. This flexibility will become especially important in systems that can improve their adaptation capabilities by learning new adaptation rules (Leake, 1993, 1994c): As the ability of these systems to adapt cases improves, constructive similarity assessment will automatically reflect that improvement by treating a wider class of cases as usefully similar.

How changing circumstances affect similarity assessment: Other research has examined how to change similarity criteria according changing system goals for applying retrieved cases; in those models current goals determine which features to consider important in a given situation (e.g., Ashley, 1990; Ashley & Rissland, 1987, Kolodner, 1989; Leake, 1991a, 1992). The constructive similarity assessment process is dynamic in a different way: It elaborates new situations according to the current contents of memory—specific cases, general background knowledge and specific beliefs—and dynamically alters how the reasoner understands and represents a given new case, based on the current contents of memory. In this respect, the flavor of this model is very similar to that of Schank's (1982) dynamic memory theory, in which the process for retrieving information from memory is largely reconstructive.

6.3 Relationship to the qualification problem

Constructive similarity assessment also relates to general issues of commonsense reasoning. It addresses the problem of deciding which potentially-relevant features of an event to consider when attempting to explain an anomalous situation, which can be seen as the inverse of McCarthy's (1980) classic *qualification problem.* The qualification problem is the problem of determining the effects of an action, given that those effects may always be modified by infinitely-many conditions reflecting possible contexts. The problem addressed by constructive similarity assessment is the problem of reconstructing events and context from information about their outcome. Just as a reasoner attempting to determine the outcome of an action in advance cannot examine every possible factor that might affect that outcome, an explainer cannot examine every possible event that might have been important to the observed outcome.

McCarthy proposed circumscription as a general mechanism for addressing the qualification problem, but noted that in any task domain, a reasoner using circumscription will require domain-dependent heuristics to determine the contextual factors to consider. Constructive similarity assessment addresses the problem of obtaining the requisite domain-dependent information by relying on cases rather than heuristics to determine the features to examine: it uses specific prior cases, which encapsulate information about what was relevant in similar prior situations, to hypothesize a context and to determine the contextual factors to account for an anomalous outcome.

Thus constructive similarity assessment differs from previous similarity assessment methods, as well as from prior methods for determining feature relevance, by (1) using the cases in memory to elaborate the representation of a partially-specified new case during the similarity assessment process itself, and (2) using a pragmatic criterion—whether the retrieved case can be adapted into a useful solution—as the basis for its similarity judgments when comparing cases.

7 Conclusions

Abductive explanations can play a crucial role in understanding and learning by knowledge-based systems, but generating the needed explanations is a difficult problem. The case-based approach to explanation generation is designed to control the cost of generating explanations, to increase robustness of explanation generation by exploiting regularities in the domain that may not be explicitly reflected by the explainer's rule base, and to maintain flexibility to explain novel events by allowing prior explanations to be applied in a flexible way when needed.

Case-based explanation depends on assessing similarity between new situations and previous explanations to determine their relevance. We have presented similarity assessment methods for guiding both initial retrieval of candidate explanations and the evaluation of retrieved candidates. Addressing the retrieval problem depends on judging the similarity between the anomaly to explain and the anomalies explained by explanations stored in memory. We have presented a vocabulary of anomalies associated with knowledge structures that are used to characterize new anomalies in a uniform way. Once anomalies have been described by anomaly characterizations, the similarity assessment problem is reduced to the task of comparing anomaly characterizations. This method both facilitates initial similarity assessment and alleviates the classic problem of analogical mapping when applying previous explanations to new situations.

Evaluating the appropriateness of retrieved cases requires determining the relevant similarities between old and new situations. This problem is particularly acute for the case-based explanation task because of the need to decide similarity despite incomplete input case descriptions. We address this problem with a novel approch to similarity assessment, *constructive similarity assessment*. Constructive similarity assessment overcomes the problem of incomplete information by using prior cases and case adaptation rules to hypothesize important new features in the input case—features that may not be present in either the input case description or the retrieved case. The resulting criterion for "similarity" between the new situation and the retrieved case is a purely pragmatic one: whether reasoning starting from the previous case leads to a useful picture of the new situation. Together, the two similarity assessment methods facilitate explanation of the imperfectly-described situations faced by explainers, allowing explainer knowledge to guide decisions in a flexible and pragmatic way.

References

Adams-Webber, J. (1989). Kelly's pragmatic constructivism. *Canadian Psychology*, *30*, 190–193.

Alterman, R. (1986). An adaptive planner. In *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 65–69). Philadelphia, PA.

Ashley, K. (1990). *Modeling legal argument: reasoning with cases and hypotheticals.* Cambridge: MIT Press.

Ashley, K. & Rissland, E. (1987). Compare and contrast, a test of expertise. In *Proceedings of the Sixth Annual National Conference on Artificial Intelligence* (pp. 273–284). Seattle, WA.

Bareiss, R. (1989). *Exemplar-Based Knowledge Acquisition: A Unified Approach to Concept Representation, Classification, and Learning.* San Diego: Academic Press.

Bareiss, R. & King, J. (1989). Similarity assessment in case-based reasoning. In Hammond, K. (Ed.), *Proceedings of the Case-Based Reasoning Workshop* (pp. 67–71). Pensacola Beach, FL.

Bhatnagar, N. & Mostow, J. (1990). Adaptive search by explanation-based learning of heuristic censors. In *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 895–901). Boston, MA.

Blessing, S. & Ross, B. (1994). Problem content affects the categorization and solutions of problems. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 51–55). Atlanta, GA.

Branting, K. & Porter, B. (1991). Rules and precedents as complementary warrants. In *Proceedings of the Ninth National Conference on Artificial Intelligence* (pp. 3–9). Anaheim, CA.

Bylander, T., Allemang, D., Tanner, C., & Josephson, J. (1991). The computational complexity of abduction. *Artificial Intelligence*, *49*, 25–60.

Chandrasekaran, B. (1994). Functional representation and causal processes. In Yovits, M. (Ed.), *Advances in Computers*. New York: Academic Press.

Charniak, E. (1986). A neat theory of marker passing. In *Proceedings of the Fifth National Conference on Artificial Intelligence* (pp. 584–588). Philadelphia, PA.

Charniak, E. (1988). Motivation analysis, abductive unification, and nonmonotonic equality. *Artificial Intelligence*, *34*, 275–295.

Charniak, E. & Goldman, R. (1991). A probabilistic model of plan recognition. In *Proceedings of the Ninth National Conference on Artificial Intelligence* (pp. 160–165). Anaheim, CA.

Charniak, E. & Shomony, S. (1994). Cost-based abduction and MAP explanation. *Artificial Intelligence*, *66*, 345–374.

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* (pp. 590–595). Detroit, MI.

Collins, G. & Birnbaum, L. (1988). An explanation-based approach to the transfer of planning knowledge across domains. In *Proceedings of the 1988 AAAI Spring Symposium on Explanation-based Learning* (pp. 107–111). Stanford, CA.

Collins, G., Birnbaum, L., Krulwich, B., & Freed, M. (1993). The role of self-models in learning to plan. In *Foundations of Knowledge Aquisition: Machine Learning* (pp. 83–116). Boston: Kluwer.

Console, L. & Torasso, P. (1990). Hypothetical reasoning in causal models. *International Journal of Intelligent Systems*, *5*(1), 83–124.

Cunningham, P., Bonzano, A., & Smyth, B. (1995). An incremental case retrieval mechanism for diagnosis. Technical Report TCD-CS-95-01, Trinity College, Dublin.

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

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

Dietterich, T. & Flann, N. (1988). An inductive approach to solving the imperfect theory problem. In *Proceedings of the 1988 AAAI Spring Symposium on Explanation-based Learning* (pp. 42–46). Stanford, CA.

Falkenhainer, B. (1990). Abduction as similarity-driven explanation. In O'Rorke, P. (Ed.), *Working Notes of the 1990 Spring Symposium on Automated Abduction* (pp. 135–139). Technical Report 90-32, Department of Information and Computer Science, University of California, Irvine.

Falkenhainer, B., Forbus, K., & Gentner, D. (1989). The structure-mapping engine: algorithm and examples. *Artificial Intelligence*, *41*, 1–63.

Ford, K., Canas, A., & Adams-Webber, J. (1991). Explanation as a knowledge acquisition issue. In Fishman, M. (Ed.), *Proceedings of the 1991 Florida AI Research Symposium* (pp. 185–191). Cocoa Beach, FL.

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

Harman, G. (1965). The inference to the best explanation. *Philosophical Review*, *74*, 88–95.

Hinrichs, T. (1992). *Problem Solving in Open Worlds: A Case Study in Design*. Hillsdale, NJ: Lawrence Erlbaum.

Hobbs, J., Stickel, M., Appelt, D., & Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, *63*(1-2), 69–142.

Hofstadter, D. & Mitchell, M. (1993). The copycat project: a model of mental fluidity and analogy-making. In Holyoak, K. & Barnden, J. (Eds.), *Advances in Connectionist and Neural Computation Theory*, Vol. 2. Norwood, NJ: Ablex.

Holyoak, K. & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*, 295–355.

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

Kass, A. (1986). Modifying explanations to understand stories. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 691-696). Amherst, MA.

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

Kass, A. (1992). Question asking, artificial intelligence, and human creativity. In Lauer, T., Peacock, E., & Graesser, A. (Eds.), *Questions and Information Processing.* Hillsdale, NJ: Lawrence Erlbaum.

Kass, A. (1994). Tweaker: adapting old explanations to new situations. In Schank, R., Riesbeck, C., & Kass, A. (Eds.), *Inside Case-Based Explanation* (pp. 263–295). Hillsdale, NJ: Lawrence Erlbaum.

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

Keane, M., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping. *Cognitive Science*, *18*, 387–438.

Kelly, G. (1970). A brief introduction to personal construct theory. In Bannister, D. (Ed.), *Perspectives in Personal Construct theory.* New York: Academic Press.

Kolodner, J. (1987). Extending problem solving capabilities through case-based inference. In *Proceedings of the Fourth International Workshop on Machine Learning* (pp. 167–178). Irvine, CA.

Kolodner, J. (1989). Judging which is the “best” case for a case-based reasoner. In Hammond, K. (Ed.), *Proceedings of the Case-Based Reasoning Workshop* (pp. 77–81). Clearwater, FL.

Kolodner, J. (1993). *Case-Based Reasoning.* San Mateo, CA: Morgan Kaufmann.

Konolige, K. (1990). A general theory of abduction. In O'Rorke, P. (Ed.), *Working Notes of the 1990 Spring Symposium on Automated Abduction.* Technical Report 90-32, Department of Information and Computer Science, University of California, Irvine.

Koton, P. (1988). Reasoning about evidence in causal explanations. In *Proceedings of the Seventh National Conference on Artificial Intelligence* (pp. 256–261). Minneapolis, MN.

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

Leake, D. (1991a). ACCEPTER: a program for dynamic similarity assessment in case-based explanation. In Bareiss, R. (Ed.), *Proceedings of the Case-Based Reasoning Workshop* (pp. 51–62). Washington, DC.

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

Leake, D. (1992a). Constructive similarity assessment: using stored cases to define new situations. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 313–318). Bloomington, IN.

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

Leake, D. (1993). Learning adaptation strategies by introspective reasoning about memory search. In *Proceedings of the AAAI-93 Workshop on Case-Based Reasoning* (pp. 57–63). Washington, DC. AAAI Press technical report WS-93-01.

Leake, D. (1994a). ACCEPTER: evaluating explanations. In Schank, R., Riesbeck, C., & Kass, A. (Eds.), *Inside Case-Based Explanation* (pp. 167–206). Hillsdale, NJ: Lawrence Erlbaum.

Leake, D. (1994b). Issues in goal-driven explanation. In Ram, A. & desJardins, M. (Eds.), *Proceedings of the 1994 AAAI Spring Symposium on Goal-Driven Learning* (pp. 72-79). Stanford, CA.

Leake, D. (1994c). Towards a computer model of memory search strategy learning. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 549–554). Atlanta, GA.

Leake, D. (1995). Goal-driven integration of explanation and action. In Ram, A. & Leake, D. (Eds.), *Goal-Driven Learning.* Cambridge, MA: MIT Press/Bradford Books. In press.

Leake, D. (In press). Abduction, experience, and goals: a model of everyday abductive explanation. *The Journal of Experimental and Theoretical Artificial Intelligence.*

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

Levesque, H. (1989). A knowledge-level account of abduction. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1061–1067). Detroit, MI.

McCarthy, J. (1980). Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, *13,* 27–39.

McDermott, D. (1974). Assimilation of new information by a natural language-understanding system. Master's thesis, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Cambridge, MA. Technical Report 291.

McDermott, D. (1982). A temporal logic for reasoning about processes and plans. *Cognitive Science*, *6*, 101–155.

Mitchell, M. (1993). *Analogy-Making as Perception.* Cambridge: MIT Press.

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

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

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

Norvig, P. (1989). Marker passing as a weak method for text inferencing. *Cognitive Science*, *13*, 569–620.

O'Rorke (1994). Abduction and explanation-based learning: case studies in diverse domains. *Computational Intelligence*, *10*, 295–330.

Otero & Campanario (1990). Comprehension evaluation and regulation in learning from science texts. *Journal of Research in Science Teaching*, *27*, 447–460.

Ourston, D. & Mooney, R. J. (1994). Theory refinement combining analytical and empirical methods. *Artificial Intelligence*, *66*, 273–309.

Owens, C. (1994). Retriever and Anon: retrieving structures from memory. In Schank, R., Riesbeck, C., & Kass, A. (Eds.), *Inside Case-Based Explanation* (pp. 89–126). Hillsdale, NJ: Lawrence Erlbaum.

Pazzani, M. (1990). *Creating a Memory of Causal Relationships: An Integration of Empirical and Explanation-Based Methods*. Hillsdale, NJ: Lawrence Erlbaum.

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

Poole, D. (1989). Normality and faults in logic-based diagnosis. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1304–1310). Detroit, MI.

Pople, H. (1973). On the mechanization of abductive logic. In *Proceedings of the Third International Joint Conference on Artificial Intelligence* (pp. 147-152). Stanford, CA.

Porter, J. (Ed.). (1973). *The Family Car*. New York: Time-Life Books.

Rajamoney, S. (1993). Designing experiments to extend the domain theory. In DeJong, G. (Ed.), *Investigating Explanation-Based Learning* (pp. 166–189). Boston: Kluwer.

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

Ram, A. & Cox, M. (1994). Introspective reasoning using meta-explanations for multistrategy learning. In Michalski, R. & Tecuci, G. (Eds.), *Machine Learning: A Multistrategy Approach*. San Mateo, CA: Morgan Kaufmann.

Ram, A. & Leake, D. (1991). Evaluation of explanatory hypotheses. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 867–871). Chicago, IL.

Read, S. & Cesa, I. (1991). This reminds me of the time when ...: expectation failures in reminding and explanation. *Journal of Experimental Social Psychology*, *27*, 1–25.

Redmond, M. (1992). *Learning by Observing and Understanding Expert Problem Solving.* Ph.D. 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* (pp. 115–120). Vancouver, B.C.

Riesbeck, C. & Schank, R. (1989). *Inside Case-Based Reasoning.* Hillsdale, NJ: Lawrence Erlbaum.

Roth, E., Bennett, K., & Woods, D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-Machine Studies*, *27*, 479–525.

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

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

Schank, R. & Leake, D. (1989). Creativity and learning in a case-based explainer. *Artificial Intelligence*, *40,* 353–385.

Schank, R., Riesbeck, C., & Kass, A. (Eds.). (1994). *Inside Case-Based Explanation.* Hillsdale, NJ: Lawrence Erlbaum.

Segre, A. (1988). *Machine learning of robot assembly plans.* Boston: Kluwer.

Selman, B. & Levesque, H. J. (1990). Abductive and default reasoning: a computational core. In *Proceedings of the Eighth National Conference on Artificial Intelligence* (pp. 343–348). Boston, MA.

Simoudis, E. & Miller, J. (1991). The application of CBR to help desk applications. In Bareiss, R. (Ed.), *Proceedings of the Case-Based Reasoning Workshop* (pp. 25–36). Washington, DC.

Sooriamurthi, R. & Leake, D. (1994). Towards situated explanation. In *Proceedings of the Twelfth National Conference on Artificial Intelligence* (p. 1492). Seattle, WA.

Stern, C. & Luger, G. (1992). A model for abductive problem-solving based on explanation templates and lazy evaluation. *International Journal of Expert Systems*, *5*, 249–265.

Tadepalli, P. (1989). Lazy explanation-based learning: a solution to the intractable theory problem. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 694–700). Detroit, MI.

Thagard, P. (1989). Explanatory coherence. *The Behavioral and Brain Sciences*, *12*, 435–502.

Tuhrim, S., Reggia, J., & Goodall, S. (1991). An experimental study of criteria for hypothesis plausibility. *Journal of Experimental and Theoretical Artificial Intelligence*, *3*, 129–144.

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

Weintraub, M. (1991). *An Explanation-Based Approach to Assigning Credit.* Ph.D. thesis, The Ohio State University.

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