plausibility might actually be undesirable in a humorous explanation.) Given a task, ACCEPTER examines the causes outlined in a candidate explanation, to determine whether those causes are sufficient to accomplish the task. For example, suppose the explainer's task is to repair a malfunctioning device. An explanation must account for the symptoms of the malfunction, and show how those symptoms result from abnormalities in the device. For example, if a television set is smoking, a good explanation for repair must say which parts are causing the smoke, rather than simply saying the smoke is the result of combustion within the device.

ACCEPTER evaluates explanations for four purposes: predicting an event, controlling its future occurrence, repairing an undesirable state, and assigning responsibility. Depending on which purpose is in effect, the system requires different information to be included in an explanation. For example, one episode for which the system evaluates explanations involves a car recall. It is offered two possible explanations for the recall: that there was bad quality control when the car was manufactured, and that the car has a defective transmission. Both explanations are plausible when taken out of context, but when ACCEPTER evaluates them *for the repair task*, it rejects the first explanation and accepts the second, since only the second provides the information needed for the repair.

ACCEPTER's evaluation of usefulness is dynamic, depending on the current task. It may require that an explanation include causes that are observable, predictive, and distinctive (for the task of predicting and avoiding future problems), repairable (for repairing a bad state), controllable (for preventing or causing an outcome), or desirable/undesirable (for assigning praise or blame). All these criteria reflect the content of the explanation, rather than structure alone. A more detailed description can be found in Leake (1989b; 1990).

Conclusion

Abduction, or inference to the best explanation, is a central component of the reasoning process. The "best" explanation is not one that is the most "correct," if correctness is even measurable in the domain of interest, but one that is most useful to the process that is seeking the explanation. Consequently, criteria for evaluating the goodness of explanations must depend on a theory of the types of uses to which explanation may be applied. We have argued that two main classes of purposes must be considered: those based on knowledge goals reflecting internal desires for information, and those based on goals to accomplish tasks in the external world.

The role of knowledge goals in explanation evaluation has been explored in the AQUA and ACCEPTOR programs. AQUA is a computer model of the theory of question-driven understanding. AQUA builds explanations in order to find answers to questions raised by gaps in its domain knowledge, and learns by incrementally improving its understanding of the domain. The relationship between tasks and requirements for explanations has been investigated in ACCEPTER, a program to evaluate the usefulness of explanations of anomalies for major classes of overarching goals. We are currently investigating the relationships between tasks and knowledge goals, and exploring how a system might generate different knowledge goals depending on its current tasks. Both AQUA and ACCEPTER take a strongly context-dependent view of evaluation of hypotheses: the final determinant of an explanation's goodness is whether it provides the desired information.

References

J.G. Carbonell. 1979. *Subjective Understanding: Computer Models of Belief Systems*. PhD thesis, Yale University, New Haven, CT.

E. Charniak. 1986. A Neat Theory of Marker Passing. In *Proc. of the Natl. Conf. on Artificial Intelligence*, Philadelphia, PA.

F. Hayes-Roth & V. Lesser. 1976. Focus of attention in a distributed logic speech understanding system. In *Proc. of the IEEE Intl. Conf. on ASSP*, Philadephia, PA.

A. Kass, D. Leake & C. Owens. 1986. *SWALE: A Program That Explains.* In Schank (1986), 232–254.

H.A. Kautz & J.F. Allen. 1986. Generalized plan recognition. In *Proc. of the Natl. Conf. on Artificial Intelligence*, Philadelphia, PA.

R. Keller. 1988. Defining operationality for explanation-based learning. *Artificial Intelligence*, 35.

K. Konolige. 1990. A General Theory of Abduction. In *Proc. of the AAAI Spring Symp. on Automated Abduction*, Stanford, CA.

J.R. Hobbs, M. Stickel, D. Appelt & P. Martin. 1990. *Interpretation as Abduction*. Tech. Note 499, SRI International.

D. Leake. 1989. Anomaly detection strategies for schema-based story understanding. In *Proc. of the 11th Annual Conf. of the Cognitive Science Soc.*, 490–497, Ann Arbor, MI.

D. Leake. 1989. *Evaluating Explanations*. PhD thesis, Yale University, New Haven, CT.

D. Leake. 1990. Task-based criteria for judging explanations. In *Proc. of the 12th Annual Conf. of the Cognitive Science Soc.*, 325–332, Cambridge, MA.

R. Mehlman & C. Snyder. 1983. Excuse theory: A test of the self-protective role of attributions. *Journal of Personality and Social Psychology*, 49(4):994–1001.

S. Morris & P. O'Rorke. 1990. An Approach to Theory Revision using Abduction. In *Proc. of the AAAI Spring Symp. on Automated Abduction*, Stanford, CA.

H. Ng & R. Mooney. 1990. On the role of coherence in abductive explanation. In *Proc. of the 8th Natl. Conf. on Artificial Intelligence*, 337–342, Boston, MA.

A. Ram. 1987. AQUA: Asking Questions and Understanding Answers. In *Proc. of the 6th Natl. Conf. on Artificial Intelligence*, 312–316, Seattle, WA.

A. Ram. 1989. *Question-driven understanding: An integrated theory of story understanding, memory and learning*. PhD thesis, Yale University, New Haven, CT. Research Report #710.

A. Ram. 1990. Incremental Learning of Explanation Patterns and their Indices. In *Proc. of the 7th Intl. Conf. on Machine Learning*, Austin, TX.

A. Ram. 1990. Explanation Patterns: A Theory of Volitional Explanation. In *Proc. of the 12th Annual Conf. of the Cognitive Science Soc.*, Cambridge, MA.

A. Ram. 1990. Knowledge Goals: A Theory of Interestingness. In *Proc. of the 12th Annual Conf. of the Cognitive Science Soc.*, Cambridge, MA.

C. Rieger. 1975. Conceptual Memory and Inference. In R.C. Schank (ed.), *Conceptual Information Processing*, North-Holland.

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

D. Sperber & D. Wilson. 1986. *Relevance: Communication and Cognition*. Language and Thought Series. Harvard University Press.

M.E. Stickel. 1990. A Method for Abductive Reasoning in Natural-Language Interpretation. In *Proc. of the AAAI Spring Symp. on Automated Abduction*, Stanford, CA.

P. Thagard. 1989. Explanatory coherence. *Behavioral and Brain Sciences*, 12(3):435–502.

H. Zukier. 1986. The Paradigmatic and Narrative Modes in Goal-Guided Inference. In R. Sorrentino & E. Higgins (ed.), *Handbook of Motivation and Cognition: Foundations of Social Behavior*, 465–502, Guilford Press.

- **Choosing a response to an unexpected event** – learn causes that allow discrimination between possible plans, by predicting events or identifying current circumstances
- **Repairing an undesirable state** – learn repairable causes of that state
- **Causing recurrence** – learn achievable causes
- **Preventing recurrence** – learn blockable causes
- **Assigning credit or blame** – learn particular actors' influence on an outcome
- **Replicating another actor's success** – learn motivations of the observed actor's unusual planning decisions

Since each of these tasks requires different information, explanations to serve them must be evaluated in the context of those tasks. Evaluation criteria based on knowledge organization goals, knowledge acquisition goals, and tasks are collectively called *utility-based criteria* in this paper.

Two case studies

We have developed two computer programs to test our theory. Although space limitations preclude detailed descriptions of these programs, we describe them briefly to illustrate the kind of abductive reasoning that our theory of explanation evaluation supports. It should be noted that both programs deal with real world examples in domains in which provably correct domain theories are unavailable. Both programs construct explanations in service of other tasks (e.g., learning, natural language understanding), and hence must evaluate explanations in the context of these tasks.

The AQUA program

Our first example program is AQUA, a story understanding program that learns from what it reads (Ram, 1987; Ram, 1989). AQUA reads newspaper stories about terrorism, such as the "blackmailed into suicide bombing" story mentioned above (New York Times, Nov 27, 1985). In order to understand text, AQUA must integrate the text, which is often ambiguous, elliptic and vague, with its world knowledge, which is often incomplete and possibly incorrect. In order to learn from what it reads, it must detect perceived anomalies in the text which may identify flaws or gaps in its model of the domain, formulate explanations to resolve those anomalies, confirm or refute potential explanations, and possibly learn new explanations or modify incorrect ones.

The process of natural language understanding generates reasoning goals or *questions*, representing what the understander needs to know in order to perform an understanding task, be it explanation, learning, or some other cognitive task. These questions constitute the specific knowledge goals of the understander generated during a parsing experience, and are used to focus the reasoning processes on aspects of the input that are actually relevant. These goals are also used to focus the learning process so that the understander learns what it needs to know in order to better carry out its tasks.

AQUA uses the following criteria to evaluate hypotheses. Although AQUA uses a case-based approach using explanation patterns to construct explanations (Ram, 1990b), the criteria listed here are applicable to other kinds of explanation construction methods which rely on domain knowledge in the form of inference rules, cases, schemas, or other types of knowledge structures.

1. **Believability:** Do I believe the domain knowledge from which the hypothesis was derived? This is an issue for any learning program in a realistic domain for which a correct domain theory is not yet known.
2. **Applicability:** How well does the domain knowledge (the particular rules, cases or schemas) apply to this situation? Did it fit the situation without any modifications?
3. **Relevance:** Does the hypothesis address the underlying anomaly? Does it address the knowledge goals of the reasoner? The hypothesis is evaluated in the context of both knowledge acquisition and organization goals.
4. **Verification:** How definitely was the hypothesis confirmed or refuted in the current situation? Does the hypothesis spawn new knowledge goals (requiring further information to help verify the hypothesis)?
5. **Specificity:** Is the hypothesis abstract and very general, or is it detailed and specific? This is a structural criterion in the sense that it is based on the structure, and not the content, of the hypothesis. However, the structure of the hypothesis is evaluated in the context of the organization of causal memory.

Intuitively, a "good" explanation is not necessarily one that can be proven to be "true" (criterion 4), but also one that seems plausible (1 and 2), fits the situation well (2 and 5), and is relevant to the goals of the reasoner (criterion 3).

AQUA is a dynamic story understanding program that is driven by its questions or goals to acquire knowledge. Rather than being "canned," the program is always changing as its questions change; it reads similar stories differently and forms different interpretations as its questions and interests evolve. AQUA judges the interestingness of the input with respect to its knowledge goals (Ram, 1990c), and learns about the domain by answering its questions (Ram, 1990a). Both these processes are goal-based. Here, we are proposing that the evaluation of explanations be goal-based as well. It is important for the evaluation criteria to be sensitive to the current goals of the reasoner.

The ACCEPTER program

Our second example is ACCEPTER (Leake, 1989b), which was designed as the central understanding component for a case-based explanation system (Kass *et al.*, 1986) to both detect problems in explanation and guide adaptation to resolve them. ACCEPTER processes stories of episodes of death (such as the death of the star racehorse Swale), damage (such as car problems), and destruction (such as the explosion of the Space Shuttle Challenger).

Like AQUA, ACCEPTER is a story understanding program that detects anomalous events in the stories it processes, and evaluates the goodness of candidate explanations for those anomalies. However, ACCEPTER's evaluation of explanations includes task-based criteria, according to user-selected explainer goals. The system's evaluation criteria involve criteria to determine:

1. **Relevance** of an explanation to an anomaly (i.e., whether the explanation accounts for why reasoning failed)
2. **Plausibility** of an explanation
3. **Usefulness** of an explanation for the current task

All three criteria must often be satisfied for an explanation to be satisfactory, but in some cases not all are important (e.g.,

useful if it allows the reasoner to learn, or to accomplish current tasks. The claim here is that *an explanation must be both causal and relevant in order to be useful*.

An explanation of an anomaly, therefore, must answer two types of questions:

1. Why did things occur as they did in the world? This question focuses on understanding, and learning about, the causal structure of the domain.
2. Why did I fail to predict this correctly? This question focuses on understanding, and improving, the organization of the reasoner's own model of the domain.

The answer to the first question is called a *domain explanation* since it is a statement about the causality of the domain. The answer to the second question is called an *introspective* or *meta-explanation* since it is a statement about the reasoning processes of the system.

Each of the above questions relates to a need to collect or organize the missing information that caused the anomaly, and that utility-based evaluation criteria must address. The first question gives rise to *knowledge acquisition goals*, which are goals to collect information or knowledge about the domain that the anomaly has signaled as being missing. The second question gives rise to *knowledge organization goals*, which are goals to improve the organization of knowledge in memory. Let us consider the second question first.

Introspective explanations: Addressing knowledge organization goals. One of the questions an explanation must address is why the reasoner failed to make the correct prediction in a particular situation. This could happen in three ways:

1. **Novel situation:** The reasoner did not have the knowledge structures to deal with the situation.
2. **Incorrect world model:** The knowledge structures that the reasoner applied to the situation were incomplete or incorrect.
3. **Mis-indexed domain knowledge:** The reasoner did have the knowledge structures to deal with the situation, but it was unable to retrieve them since they were not indexed under the cues that the situation provided.

When an explanation is built, the reasoner needs to be able to identify the kind of processing error that occurred and invoke the appropriate learning strategy to prevent recurrence of the error. For example, if an incomplete knowledge structure is applied to a situation, the knowledge activated by the resulting processing error must represent both the knowledge that is missing, and the fact that this piece of knowledge, when it comes in, should be used to fill in the gap in the original knowledge structure. Similarly, if an error arose due to a mis-indexed knowledge structure, the explanation, when available, should be used to re-index the knowledge structure appropriately.

Knowledge organization goals can be categorized by the type of gap that gave rise to them, or by the type of learning that results from their satisfaction:

- **Missing knowledge** – learn new knowledge to fill gap in domain model
- **Unconnected knowledge** – learn new connection or new index
- **Implicit assumption** – learn heuristics for when to check assumption explicitly
- **Calculated simplification** – learn heuristics for when to check assumption in detail
- **Explicit assumption** – learn new knowledge to correct the assumption
- **Conjunctive assumptions** – learn new interactions

A hypothesis is evaluated from the point of view of knowledge organization goals by checking to see if it provides the information necessary for the type of learning that the reasoner is trying to perform. For example, suppose the reasoner reads a newspaper story about a Lebanese teenager who, it turns out, is blackmailed into going on a suicide bombing mission. Even if the reasoner already knows about terrorism, religious fanatics and blackmail, the story may nevertheless be anomalous if the reasoner has never seen this particular scenario before. The difficulty arises from the fact that blackmail is not ordinarily something that comes to mind when one reads about suicide bombing. Here, the reasoner can learn a new connection between the knowledge structures describing suicide bombing and blackmail, respectively. In order to do this, the explanation must provide the information required to identify the conditions under which a suicide bombing is likely to be caused through blackmail.

This type of analysis is essential in determining whether an explanation is sufficient for the purposes of the reasoning task at hand. In this example, the reasoning task is to satisfy a knowledge organization goal, which is a goal to learn by reorganizing existing knowledge in memory.

Domain explanations: Addressing knowledge acquisition goals. Knowledge acquisition goals seek new causal knowledge about the domain. A domain explanation is a causal chain that demonstrates why the anomalous proposition might hold by introducing a set of premises that causally lead up to that proposition. If the reasoner believes or can verify the premises of an explanation, the conclusion is said to be explained. Explanations are often verbalized using their premises or abductive assumptions. However, the real explanation includes the premises, the causal chain, and any intermediate assertions that are part of the causal chain.

In order to be useful, a hypothesis must provide the information that is being sought by the knowledge acquisition goals of the reasoner. For example, if the reasoner has a goal to acquire knowledge about the biochemical properties of a particular virus, a description of a sick patient must provide the biochemical information in order to qualify as an explanation from the point of view of that goal. An alternative hypothesis that provides causal information suggesting how some drug might destroy the virus, while useful from the point of view of curing the patient, may not provide the required information.

Task-triggered knowledge acquisition goals. Knowledge acquisition goals often arise from an explainer's tasks. A doctor wishing to cure a patient will seek an explanation of the patient's symptoms that suggests the plan of action for a cure; an epidemiologist may seek an explanation in terms of environmental factors, to make similar disease outbreaks less likely. In order to reflect such factors in evaluation, we must identify the basic tasks that can be prompted by an anomaly, in turn prompting explanations that focus on particular aspects of a situation. For example, each of the tasks below is associated with particular knowledge acquisition goals:

generating suitable explanations for system use, although the range of uses considered has been limited (see Keller (1988) for a discussion of operationality considerations).

From the functional point of view, these ideas are related to the "goal satisfaction principle" of Hayes-Roth and Lesser (1976), which states that more processing should be given to knowledge sources whose responses are most likely to satisfy processing goals, and to the "relevance principle" of Sperber and Wilson (1986), which states that humans pay attention only to information that seems relevant to them. These principles make sense because cognitive processes are geared to achieving a large cognitive effect for a small effort.

Utility-based evaluation of explanatory hypotheses attempts to maximize the utility of explanations by explicitly considering the needs of the reasoner in forming the explanation. Further examples of the task-sensitive nature of human explanation can be found in psychological research on excuse theory, which demonstrates that people manipulate explanations to displace blame for poor performance (e.g., Mehlman & Snyder (1983)).

The explanation cycle

The process model for the task of explanation consists of the following steps:

1. **Anomaly detection:** Identification of an unusual fact that needs explanation. (See Leake (1989a; 1989b) and Ram (1989) for our approaches to anomaly detection.)
2. **Explanatory hypothesis construction:** Construction of one or more explanatory hypotheses that would resolve the anomaly and explain the situation. This is typically done by chaining together causal inference rules through a search process (e.g., Rieger (1975), Morris & O'Rorke (1990)), through a weighted or cost-based search (e.g., Hobbs *et al.* (1990), Stickel (1990)), or through a case-based reasoning process in which previous explanations for similar situations are retrieved and adapted for the current situation (e.g., Schank (1986), Kass *et al.* (1986), Leake (1989b), Ram (1990b)).
3. **Hypothesis verification:** Confirmation or refutation of possible explanations or, if there is more than one hypothesis, discrimination between the alternatives. A hypothesis is a causal graph that connects the premises of the explanation to the conclusions via a set of intermediate assertions. At the end of this step, the reasoner is left with one or more alternative hypotheses. Partially confirmed hypotheses are typically maintained in a data dependency network (e.g., AQUA's hypothesis tree (Ram, 1989)).

Evaluation criteria

Regardless of how explanatory hypotheses are constructed, the evaluation of these hypotheses (step 3 in the explanation cycle above) is a central and difficult problem. We categorize evaluation criteria into structural (or syntax-based) and utility-based (or goal-based) criteria.

Structural criteria

Structural criteria use the structural or syntactic properties of the causal chain to evaluate hypotheses. A goodness measure for each hypothesis is computed based on the length of the causal chain, the number of abductive assumptions, or other such structural properties.

Most structural criteria appeal to Occam's razor by requiring *minimality* of hypotheses. Simply stated, a hypothesis that is "minimal" with respect to some criterion is preferred over one that is not (e.g., Charniak (1986), Kautz & Allen (1986)). For example, Konolige (1990) argues that "closure + minimization implies abduction." To take another example, the TACITUS system for natural language interpretation merges redundancies as a way of getting a minimal interpretation, which is assumed to be a best interpretation (Hobbs *et al.*, 1990). Minimality criteria include:

- **Length:** Causal chains with the shortest overall length are preferred.
- **Abductive assumptions:** Explanations requiring the fewest abductive assumptions are preferred.
- **Subsumption:** If two candidate hypotheses are found and one subsumes the other, the more general hypothesis is preferred.

Another approach focuses on the structural relationship of propositions in an explanation rather than minimality:

- **Explanatory coherence:** The cohesion of an explanation is measured, based on the form of connections between an explanation's propositions, and the "best connected" explanation is favored (e.g., Thagard (1989), Ng & Mooney (1990).)

While structural criteria provide an easy way to evaluate the goodness of a hypothesis, they are not very useful in real situations. Explanations are not constructed in a vacuum. Typically, there is a real world task that the reasoner is performing that requires the reasoner to seek an explanation. The reasoner may also need an explanation to help it with a piece of reasoning that it is trying of perform. Both these types of motivations for explanation influence evaluation criteria.

Utility-based criteria

An explainer's motivation for explaining will often place additional requirements on candidate explanations, beyond their form. For example, explanations prompted by anomalies must provide particular information, in order to resolve the anomaly. For example, suppose that we expected team X to win over team Y because of the talent of X's star player, but we are told that team X actually lost. If someone explained the loss by 'Y scored more points than X," the explanation would be inadequate. Although it is a correct explanation, it gives no information about *why* our expectation went wrong. The explanation "X's star was injured and couldn't play" does account for what was neglected in prior reasoning, and consequently is a better explanation. However, this explanation would not be preferred on structural grounds alone. The causal chain underlying that explanation is more complex, so it would not be favored by minimality criteria. Likewise, the explainer of the game has access to only one observation, the fact that team X lost, so coherence metrics that measure how an explanation relates *pairs* of observations, such as those described by Ng and Mooney (1990), give no grounds for preferring the second explanation.

To state our relevance criterion another way, *an explanation must address the failure of the reasoner to model the situation correctly*. In addition to resolving the incorrect predictions, it must also point to the erroneous aspect of the chain of reasoning that led to those predictions. An explanation is

Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, Chicago, IL, August 1991.

Evaluation of explanatory hypotheses[1]

Ashwin Ram
College of Computing
Georgia Institute of Technology
Atlanta, Georgia 30332-0280
E-mail: ashwin@cc.gatech.edu

David Leake
Computer Science Department
Indiana University
Bloomington, Indiana 47405-4101
E-mail: leake@cs.indiana.edu

Abstract

Abduction is often viewed as inference to the "best" explanation. However, the evaluation of the goodness of candidate hypotheses remains an open problem. Most artificial intelligence research addressing this problem has concentrated on syntactic criteria, applied uniformly regardless of the explainer's intended use for the explanation. We demonstrate that syntactic approaches are insufficient to capture important differences in explanations, and propose instead that choice of the "best" explanation should be based on explanations' utility for the explainer's purpose. We describe two classes of goals motivating explanation: knowledge goals reflecting internal desires for information, and goals to accomplish tasks in the external world. We describe how these goals impose requirements on explanations, and discuss how we apply those requirements to evaluate hypotheses in two computer story understanding systems.

In order to learn from experience, a reasoner must be able to *explain* what it does not understand. When a novel or poorly understood situation is processed, it must be interpreted in terms of knowledge structures already in memory. As long as these structures provide expectations that allow the reasoner to function effectively in the new situation, there is no reason to revise them. However, the expectations may fail to apply. In that case, the reasoner is faced with an *anomaly* — a conflict between expectations and new information — and learning is needed to prevent future failures. In order to revise its knowledge, the reasoner needs to know *why* it made the mistaken predictions, and to explain *why* the failure occurred. In other words, it must identify the knowledge structures that gave rise to the faulty expectations, and understand why its domain model was violated in this situation. Once revised, the knowledge can then be stored in memory for future use. *Abduction*, the construction of explanations, is a central component of this learning process.

Abduction is often viewed as inference to the "best" explanation. However, there is often no one "right" explanation; the definition of "best" is dependent on the *goals* of the reasoner in forming the explanation. In these situations, the "best" explanation must be more than a causal chain that correctly describes the domain; it must also address the reason that an explanation was required in the first place. The extent to which it does so determines how effectively the reasoner can learn from the explanation.

[1]Ashwin Ram's research was supported in part by the National Science Foundation under contract IRI-9009710. Both authors' research was also supported in part by the Defense Advanced Research Projects Agency and the Office of Naval Research under contract N00014-85-K-0108, and by the Air Force Office of Scientific Research under contracts F49620-88-C-0058 and AFOSR-85-0343.

This paper addresses the problem of evaluating the "goodness" of hypotheses in the context of a reasoning task. We focus on *explanatory hypotheses*, which are causal chains that attempt to justify a given anomalous fact in terms of causal relationships with reasoner's prior beliefs. We begin with discussing the nature of explanatory hypotheses. We then discuss two classes of evaluation criteria for such hypotheses. *Structural criteria* rely on structural or syntactic properties of the causal chain. For example, a reasoner might always choose the shortest causal chain as the best explanation. *Utility-based criteria* select hypotheses according to requirements arising from the system's intended use for an explanation, such as forming predictions about future events.

Although structural criteria have received the most attention in artificial intelligence programs, we will argue in favor of utility-based criteria on functional grounds. Since these criteria judge candidate hypotheses with respect to the particular task and reasoning needs that the reasoner is currently faced with, they are more likely to enable the reasoner to pick explanations that are "right" for the occasion.

We divide utility-based evaluation criteria into two subclasses: those based on *knowledge goals*, and those based on *tasks*. Criteria based on knowledge goals attempt to evaluate a hypothesis based on the reasoner's internal needs for knowledge. For example, if a system is trying to form a distinction between two categories in memory, an explanation that provides the information required make such a distinction is more useful even though it may not be the shortest explanation. Task-based criteria evaluate hypotheses from the point of view of the real world tasks that the reasoner is trying to perform. A detective, for example, might need to build a very different explanation for a stain of blood on a carpet than a cleaner trying to figure out how to remove that stain.

Our theory is based on a functional analysis of the purposes for which explanations will be used, and motivated by empirical psychological evidence. People quite clearly have what psychologists often call "goal orientations," which have a significant effect on the inferences that people draw from their experiences. There is a large body of psychological research on goal direction in focus of attention, particularly from social psychology. Zukier's (1986) review concludes: "Experimental studies have clearly demonstrated that a person will structure and process information quite differently, depending on the future use he or she intends to make of it" (p. 495). AI researchers have also proposed theories of "subjective interpretation" (e.g., Carbonell (1979), Ram (1989)) in which a system's prior beliefs and goals influence the interpretations drawn in a given situation. In addition, AI research on operationality of explanations has considered the question of