Scientific Discovery in the Layperson

Michael J. Pazzani
Department of Information and Computer Science
University of California
Irvine, CA 92717
pazzani@ics.uci.edu

Margot Flowers
Department of Computer Science
Los Angeles, CA 90024-1560
flowers@cs.ucla.edu

Abstract

In this chapter, we argue that the process of scientific discovery is in many ways similar to the process that the layperson uses to learn and evaluate predictive relationships, with the point that that these techniques should be incorporated into models of scientific discovery. We review some findings that illustrate constraints on human learning and show how these findings manifest themselves in scientific discoveries. We concentrate on two aspects of the discovery process: evaluation of evidence, and scientific rhetorics, illustrating these aspects by two programs that perform common-sense reasoning tasks. OCCAM uses a theory of causality to constrain the learning of causal relationships. We show that this system can use the same mechanism to learn about the process of catalysis. ABDUL/ILANA models the process of argumentation. We show how the system can represent competing scientific hypothesis and use argumentation strategies to direct the reasoning process. We conclude with some observations about the consequences of this view of discovery.

1.0 Introduction: Everyday reasoning and scientific reasoning

The process of scientific discovery can be broken down into two interacting processes:

- *Hypothesis generation*, which we view as a bottom-up, data-driven process in which one propose an explanation for some unexplained, unusual or unexpected occurrence in term of some causal mechanism.
- *Hypothesis evaluation*, which we view as a top-down, theory-driven analysis in which one evaluates a hypothesis against further data or against alternative hypotheses.

However, these component processes are not unique to science. In the course of learning to master the environment, the layperson also performs these same activities. Thus, we claim that progress on understanding the process of scientific discovery can be aided by understanding discovery in the layperson.

In our view, the primary difference between discovery in the layperson and the scientist is one of context: a scientist's discovery must be novel to both the scientist and society, whereas a discovery in the layperson need be novel only to the layperson. Our claim, simply stated, is that there is minimal difference between the processes used to invent the wheel and the processes used to reinvent the wheel (when one is previously unaware of wheels). The process of scientific discovery involves analysis of data, formation of new hypotheses, testing of hypotheses by experimentation, the assertion and retraction of hypotheses as they are confirmed or contradicted, and revision of incorrect hypotheses. These are precisely the same kinds of activities used in everyday situations by non-scientists. Similarly, both scientists and non-scientists are subject to similar sources of error: incorrect analysis of data, rejection of evidence, influence of self-interest, and influence of prior expectations. The processes used when a scientist interprets evidence and forms hypotheses differ only in degree and emphasis, and not kind, from those used by people in everyday life (Kuhn, Amsel & O'Loughlin, 1988).

Our belief is that a model of the discovery processes used by the layperson will require only minor modification to model the scientific discovery process. [1] We will present examples of everyday discoveries by the layperson, compare these activities to those of scientists and present examples of computational systems that model the reasoning processes of the layperson. Finally, we will discuss how these programs may be used to model the scientific discovery process.

In most instances, it is difficult to construct an accurate model of the cognitive processes of a human subject in a controlled experimental situation which occurs over a small period of time. Although we are encouraged that some are studying the reasoning processes of famous scientists over long periods of time (e.g., Tweeney, 1989; Hadamard, 1954; Hovey, 1962; Jenkings, 1983; Jenkins & Jeffrey, 1984; Osowski, 1986; Wertheimer 1959), it is an exceedingly difficult and time consuming task. There are many advantages to studying discovery in the layperson:

- There are more laypeople than scientists, providing more examples to study. For example, Thagard (1989) presents the seven examples of revolutionary changes in the history of science. Yet each year, thousands of children learn there is no Santa Claus.
- Negative as well as positive examples can be studied. The mistakes that people make provide useful information about constraints on cognitive processes (e.g., Klahr & Dunbar, 1988). Scientists who fail to make discoveries are quickly lost to obscurity and thus examples of their failures are hard to capture for study.
- Repeatable experiments are possible. Subjects can be observed making discoveries under different conditions. Such experiments can reveal the utility of a piece of information in a particular discovery. Protocols of subjects making discoveries can be recorded and analyzed (e.g., Klahr & Dunbar, 1988).

In this chapter, we identify a number of relevant cognitive traits and biases used by the layperson. We support the claim that scientific reasoning can be viewed as a special case of

1. One obvious place for refinement is the amount of effort a scientist spends in confirming hypotheses.

everyday reasoning techniques, giving examples from the domain of science. We illustrate how these techniques can be incorporated into models of scientific reasoning and how these models better capture the processes involved in scientific discovery.

We first discuss the notion of the intuitive scientist, how laypersons can be viewed as scientists, and ways in which the reasoning of scientists reflects everyday reasoning strategies. Next, we review the computational models of learning and argumentation used by OCCAM and ABDUL/ILANA, two programs that model common-sense reasoning tasks. Then, we examine a number of case studies of biases and argumentation in laypersons and scientists to show how the knowledge-intensive approach to learning can be applied to scientific discovery and how the evaluation of scientific hypotheses can be approached with techniques for understanding arguments.

1.1 Background: layperson as intuitive scientist

Our claim bears some resemblance to the notion of the "intuitive scientist" (Kelley, 1967; Kelley, 1973; Nisbett & Ross, 1980). Kelley advocated the idea that the layperson makes causal inferences in a manner similar to that of the formally trained scientist, by detecting covariation between effects and potential causes. However, this work does not address the issue of how accurately people perceive covariation between events and prior expectations have been shown to influence the perception of covariation (Chapman & Chapman, 1969; Jennings, Amabile & Ross, 1980; Kuhn, Amsel & O'Loughlin, 1988).

One recent experiment demonstrates the dominant role of prior knowledge on the detection of a predictive relationship (Pazzani & Schulenburg, 1989). In this experiment there were two different learning tasks that used the same stimuli. Subjects were shown photographs of a person with a balloon. The photographs varied in four ways:

- Age: Either an adult or a small child was performing an action.
- Action: The person either dipped the balloon in a bowl of water or stretched the balloon.
- Color: Either a yellow or purple balloon was shown.
- Size: The balloon was either very small or very large.

To study how prior knowledge can facilitate the prediction task, one group of subjects was instructed to predict whether or not the balloon would be inflated, whereas another group of subjects had to perform a classic concept acquisition task (e.g., (Bruner, Goodnow & Austin 1956)) and was instructed to learn a way to identify which balloons belonged to an arbitrary category called *Alpha*. These two groups were further subdivided into one that had to learn a conjunctive relationship (color = yellow and size = large) and one that had to learn a disjunctive relationship (age = adult or action = stretched). The latter was designed to be consistent with a subject's prior knowledge about the ease of inflating balloons[2], whereas the conjunctive relationship was not.

The subjects in this experiment were 88 undergraduates who received extra credit in an introductory psychology course for their participation. Each subject was shown a photograph, asked to make a prediction (or classification) and then shown the correct answer. This process continued until the subject made 6 correct predictions in a row. For each subject, the number of the trial on which the last error error was made was recorded.

If subjects relied only on the detection of covariation, one would expect that result would be the same for the prediction and classification task. Instead, Figure 1 shows that the task of learning a predictive relationship is influenced by prior knowledge. This effect is so strong that it dominates the well-known finding that conjunctive concepts are easier to learn than disjunctive concepts (Bruner, Goodnow & Austin, 1956).

2. Note that "consistency" does not imply "deducible from". In particular, when there are multiple interacting factors, the influence of each factor may not be obvious.

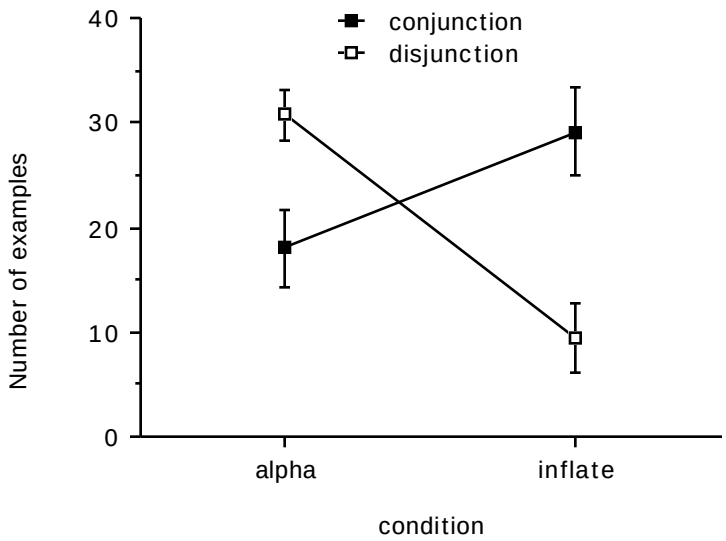

Figure 1. The ease of acquiring predictive (inflate) and descriptive (alpha) concepts. The disjunctive relationship is consistent with prior knowledge on the ease of inflating balloons, whereas the conjunctive relationship violates these beliefs.

1.2 Scientist as layperson

Kelley has revised the notion of the intuitive scientist to deal with two related findings:

- The role of prior knowledge in the perception of covariation as described in the previous section.
- The perseverance of beliefs in the face on new evidence. Once a belief becomes accepted, new evidence is not treated in the same manner as prior evidence (Gorman, 1986; Klahr & Dunbar, 1988; Lord, Ross & Lepper, 1979; Nisbett & Ross 1980; O'Brien 1986).

Early scientific discovery programs (e.g. BACON (Langley , Bradshaw & Simon 1983)), which concentrated on empirical techniques for making scientific discoveries, were not intended to account for the biases that people have in learning and evaluating causal relationships. However, scientists exhibit similar biases (Chapman & Chapman, 1969; Eddy, 1982; Elstein, Shulman & Sprafka 1978; Faust, 1984; Greenwald et al., 1986; Mahoney & Kimper, 1976; Schwartz &

Griffinn 1986). Rather than viewing these biases as non-normative shortcomings (Cohen, 1981; Nisbett & Ross, 1980; Stitch & Nisbett, 1980), we view them as essential parts of the scientific discovery process. In particular, prior knowledge provides an important constraint on the interpretation of the results of an experiment.

It is not an accident that chemists make chemical discoveries. It is not simply because chemists have beakers, perform chemistry experiments, and analyze the results of chemical experiments. Rather it is because they have prior knowledge of chemistry and the types of factors that are likely to influence chemical processes. This knowledge enables a chemist to focus on a small number of potentially relevant features. Hypotheses that are consistent with the data, but that are inconsistent with knowledge of chemistry, are not actively pursued. This focusing effect is especially important in scientific discovery, in which novelty to society is stressed. Reliance on prior knowledge serves as an effective means of controlling distributed problem solving techniques. Different background knowledge of different scientists causes each scientist to focus on a different hypothesis.

Similarly, it is important that scientists be more skeptical of data that disagrees with their hypotheses and prior knowledge. Unexpected occurrences need explanations. When data are consistent with a hypothesis, the hypothesis serves as an explanation. However, when data are not consistent with a hypothesis, there are a number of possible explanations: The hypothesis may be incorrect, there may be measurement error, equipment error, self-deception, or fraud. Until other alternatives are sufficiently substantiated, a hypothesis is not usually abandoned. This bias causes energies to be focused on discoveries within an existing framework, rather than continuously switching frameworks or expending considerable effort triple-checking equipment that produced the expected result.

2.0 Using causal knowledge and argument structure

In order to model the effects of bias and subjective evaluation of evidence in scientific discovery, we propose the use of causal knowledge and argument structure. As background, in this section we discuss two process models that address employ these kinds of knowledge.

2.1 The OCCAM system

OCCAM (Pazzani 1990; Pazzani, Dyer & Flowers 1986) can be viewed as a model of the intuitive scientist, that incorporates a learning technique called *theory-driven learning* (TDL). The system addresses the problem of learning causal knowledge by observing examples of events and their consequences. The task domain involves acquiring simple rules of causation, such as those that describe the outcome of common events in the life of a small child (e.g., when a cup made of glass is dropped, it usually breaks).

A distinction can be made between a theory of *causality* -- general principles that lead one to believe that a particular class of actions has a necessary consequence, and a theory of *causation*[3], indicating specific inference rules that indicate the effects of a particular class of actions. *OCCAM's objective is to construct a theory of causation, given a theory of causality and a number of observations.* The model of theory-driven learning was influenced by a number of experimental studies in cognitive, social, and developmental psychology that have explicated a number of principles which people use when learning causal relationships (Bullock, Gelman & Baillargeon, 1982; Shultz & Kestenbaum, 1985).

OCCAM represents a theory of causality as a set of causal patterns. A total of approximately 25 causal patterns have been implemented. The causal patterns make use of temporal and spatial constraints to suggest causal relationships. One such causal pattern is given below:

3. A theory of causation is also termed a domain theory (Mitchell, Kedar-Cabelli & Keller 1986).

```
When similar actions on an object are followed by a state change,
and a feature of the object correlates with the state change,
then that feature is need for the action to result in the state change.
```

In TDL, a causal pattern that matches a training example proposes a hypothesis which is then tested tested against new data and either accepted or rejected depending upon the accuracy of the hypothesis. Theory-driven learning can be viewed as a form of explanation-based learning (DeJong & Mooney, 1986) in which the domain knowledge (i.e., the set of causal patterns) is known to be overly general. Since it is possible for the theory of causality to propose hypotheses that are not true, these hypotheses are evaluated against further examples.

To gain an understanding of the role and scope of TDL, it is necessary to describe the overall learning architecture of which the process is a component. One important task for OCCAM is to determine what learning strategy is appropriate for each new example. The learning strategies are ordered in OCCAM by the amount of knowledge they require. Thus, OCCAM prefers explanation-based learning (EBL) if it can produce an explanation. If an explanation cannot be produced and the event fits a known causal pattern, then OCCAM uses TDL. As a last resort, OCCAM attempts similarity-based learning (SBL). Thus, TDL's role is to restricted to those observations that cannot be explained by the current knowledge of causation (otherwise, EBL would be used) and that meet the constraints of a potential causal relationships (i.e., match a causal pattern). Regularities between observations which cannot be explained, and which do not match a causal pattern (e.g., the opening of a garage door by pressing the button on a remote control) can be detected and generalized by SBL in OCCAM.

In OCCAM, the different learning strategies operate over the same memory. The memory that stores the results of SBL and TDL is the same memory that is used to create explanations for EBL. This lets the system use the results of data-intensive empirical strategies to enable later knowledge-intensive learning. OCCAM's memory is organized as a hierarchy of explanatory

schemata similar to Schank's MOPs (Schank 1982). The hierarchy of schemata serves as a discrimination net for making predictions and finding explanations. There are several advantages to this architecture for integrating empirical and explanation-based learning in OCCAM. First, the empirical learning component can acquire the background knowledge required by the explanation-based component. Second, when there is sufficient background knowledge, only a small number of examples are required to create a set of schemata that make accurate predictions.

In spite of its utility, explanation-based learning alone cannot adequately explain many scientific discoveries. In discovery tasks, the goal is to produce a complete and correct background theory, thus such a theory does not exist ahead of time to be utilized as required by EBL methods. Nonetheless, as demonstrated by the experiment in Section 1.1, prior background knowledge can focus the hypothesis generation process.

There is a major difference between a schema that encodes knowledge whose only support is a number of examples, and a schema whose support also includes an underlying theory. The differences between relevance and simple correlation affect how OCCAM handles exceptions to schemata, consistent with OCCAM's preference for knowledge-intensive strategies over data-intensive strategies. Exceptions to schemata are treated differently if they have proved relevant, that is, if it has been have previously explained why a certain regularity holds, than if they are merely correlational. When new data doesn't agree with a correlational schema, the schema itself is questioned. In contrast, when new data contradicts a schema supported by an underlying theory, the data itself is questioned. Only when the data withstands such questioning, or when the amount of data is overwhelming, is a schema supported by an underlying theory abandoned. This is an indication that the underlying theory is incorrect and needs revision to account for new empirical findings.

2.2 The ABDUL/ILANA System

In science, the questioning and confirmation of data and theories can be viewed as a kind of argument. Thus, aspects of scientific creativity can be analyzed and modeled by incorporating techniques and heuristics of argument analysis and response formation. In this section, we illustrate our basic approach to arguments with an example drawn from the domain of political argumentation and present an overview of ABDUL/ILANA (Flowers, McGuire & Birnbaum, 1982; Birnbaum, Flowers & McGuire, 1981; McGuire, Birnbaum & Flowers, 1981). Later we will show how the process of creating, evaluating, rejecting, and updating scientific hypotheses can be viewed as a similar kind of argument, and thus how argument structure analysis can be useful in analyzing the relationships between theory and data, and in determining what research steps are sensible responses.

ABDUL/ILANA views an argument as a dynamic process in which a network of beliefs, linked by support and attack dependencies (de Kleer et al., 1979) is incrementally elaborated during the course of an argument. This *argument graph* is an episodic memory (Tulving 1972) of the argument. Here, we briefly examine ABDUL/ILANA's use of such graphs for argument analysis. Consider this argument about responsibility for conflict in the Middle East:

```
A:  Who started the 67 war?
I:  The Arabs did, by the blockade.
A:  But Israel fired first.
I:  According to international law, blockades are acts of war.
A:  Were we supposed to let Israel import arms through the Straits
of Tiran?
I:  Israel was not importing arms through the straits...
```

ABDUL/ILANA models the process of participating in this argument, including natural language analysis, argument analysis and strategy choice, reasoning, and language generation. In different modes it can model either the Arab or the Israeli point of view. The argument's strategic content, which can be expressed graphically (see Figure 2), consists of various propositions or beliefs, justified or attacked by other propositions.

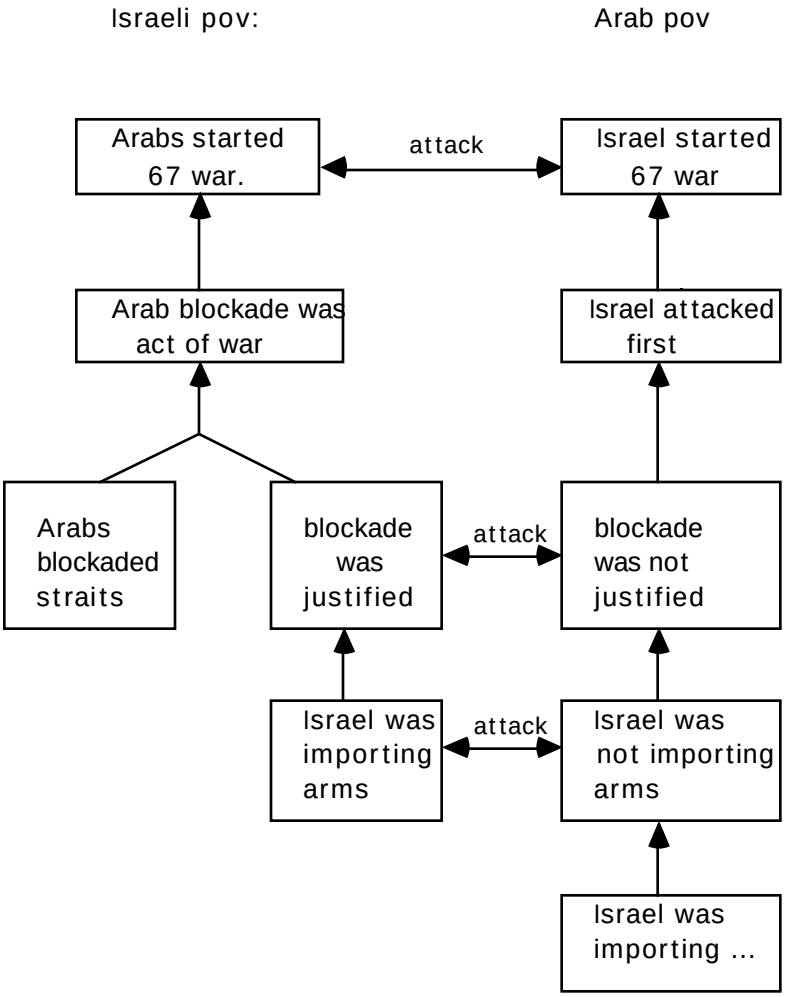

Figure 2. Argument graph illustrating support and attack relationships in ABDUL/ILANA dialog.

Our notion of "belief" is a psychological, pragmatic one; distinct from the formal propositional sense(Abelson 1979). Beliefs consist of a premise, or belief content, an attribution of who believes it, and various kinds of supporting or attacking information. Both support structures and attack structures are themselves built out of beliefs. Support structures consist of a belief, an inference rule (left implicit in Figure 2, also known as "warrants" (Toulmin, 1958; Toulmin, Reike & Janik, 1972)) whose result is the assertion of the belief, and another belief that is the basis of the inference rule, and serves to support the original belief. For example, in Figure 2, the belief that "Israel started the 67 war" is supported by the belief that "Israel attacked first".

Because we are taking a psychological rather than formal approach to modeling belief, a particular belief may have many kinds of support. Consequently, retraction of one kind of support does not automatically mean that belief is retracted. In a political domain, positions are often determined on emotional grounds. So, for example, from the Israeli point of view, retracting the existing support for the assertion that the Arab blockade caused the war does not end up causing the Israeli to retract the belief that the Arabs caused the war.

Attack structures also are relationships between two beliefs, except that they contradict one another according to some inference rule. Attack and support may exist simultaneously for a belief. For example, the Israeli point of view, although believing that the Arabs are responsible for the war, includes a representation of the Arab belief that attacks this notion. Thus, whether or not a belief is believed overall depends upon inferences made over the kinds of attacks and supports. This approach to belief representation is intended to facilitate modeling the kinds of phenomena observed in people, such as selective attention to evidence, and pursuit of hypotheses that are consistent with currently-held theories. Several subsequent computer models have been built which utilize this system of representing beliefs: OPED in the domain of editorial comprehension (Alvarado, Dyer & Flowers, 1985; Alvarado, Dyer, & Flowers ,1986), JULIA in the domain of editorial analogies (August & Dyer, 1985), SHERLOCK for teaching students inference strategies (Feifer, Dyer, & Flowers, 1986; Feifer 1989), and AQUA for giving advice on UNIX commands (Quilici, Dyer ,& Flowers, 1986; Quilici, Dyer & Flowers, in press) .

The approach taken by these computer models can be useful in modeling scientific reasoning. Hypotheses in science also can be viewed as beliefs, because their validity depends upon the balance and relationships of confirming and attacking evidence (Thagard, this volume). For example, the phlogiston theory at one time functioned as a theory to explain observed phenomena, until new experimental evidence led to its retraction. Thus, the processes of science itself can be viewed as arguments in that they involve competing analyses of data, and the

assertion, contradiction, retraction, along with the revision of hypotheses.

3.0 Examples of bias in learning and evaluating evidence and theories

In this section, we provide several examples of biases in both the layperson and the scientist. The purpose of these examples is to illustrate how these biases can be utilized in models of hypothesis generation and confirmation.

3.1 Rickettsialpox

In 1946, a new disease (today known as Rickettsialpox) was first encountered in New York City (Rouche 1984). The first known case was observed by Benjamin Shankman who encountered an unusual patient with symptoms that included fever, a small lesion, rash, and swollen glands. The initial diagnosis for this case was chicken pox. The doctor initially dismissed the fact that this patient had already had chicken pox, assuming that perhaps the first case of chicken pox was misdiagnosed. Later examinations revealed that the lesions differed from those of chicken pox and the symptoms were more severe. The patient was hospitalized and tests ruled out chicken pox, typhus, smallpox, and Rocky Mountain spotted fever. The patient was given penicillin and recovered in a few days. Shortly thereafter, Dr. Shankman encountered another case, which he treated with sulfonamides, and the patient recovered in a few days. Since the sulfonamide treatment appeared to be no less or more effective than penicillin, Dr. Shankman speculated that neither drug affected the disease. This hypothesis was strengthened, when a third patient, treated with aspirin, recovered in about the same length of time.

Soon more cases were encountered, and Dr. Shankman and other physicians noticed that all the patients lived in the same apartment complex. The United States Public Health Service investigated and eventually discovered that a member of the genus *Rickettsia* was responsible for the infections. However, originally the method of transmission remained unknown.

Meanwhile, the media picked up the story and reported the similarity of the disease to Rocky Mountain spotted fever. Charles Pomerantz, an exterminator, read reports and decided to try to help. He canvassed the neighborhood looking for ticks which are known to carry Rocky Mountain spotted fever. When he found no ticks outside, and the local kennel reported that there was no tick problem in the neighborhood, he began to consider other possible mechanisms that might carry the disease. He knew that the apartment complex was infested with mice and rats, but they were not known to transmit rash-producing diseases. He considered the possibility that mice might be the host of mites. He asked for permission to inspect the building and found some mites that were subsequently found to be carrying Rickettsia.

This case study illustrates several points about the process of discovery. First, after the doctor came up with an initial diagnosis, he dismissed evidence that did not agree with his hypothesis. Only when the evidence against the initial hypothesis was overwhelming did he consider alternative hypotheses. Second, the exterminator's background knowledge focused the generation of hypotheses to a small number that could potentially explain the concentration of the disease in a certain location (e.g., the water supply of apartment was contaminated). The exterminator's background knowledge put him in a position to advocate the hypothesis that turned out to be correct.

3.2 Piltdown man

In 1912, a fossil that became known as the Piltdown man was discovered in Sussex. The fossil appeared to be the "missing link," since it had an ape-like jaw but human molars. Later, additional fossils were found in Africa and China that differed substantially from Piltdown man. If both these later findings and Piltdown man were accepted, it would imply that there were two distinct evolutionary paths to modern man. Newly developed fluorine tests on the Piltdown man

revealed that the fossil was less than 500 years old. Further examination revealed that a molar had been filed and painted brown, and the skull had been stained with iron. Although the dating technique that revealed the hoax was not available at the time Piltdown man was first discovered, it was possible other signs of fraud could have been detected at the time of discovery.

Although in this case the scientists involved may have been more lax than usual in their standards of scrutinizing evidence, the phenomenon is a common one (Lord, Ross & Lepper 1979). At the time it was discovered, the fossil of the Piltdown man appeared to satisfy a prediction of evolutionary theory. The evidence was more closely scrutinized only when it presented complications for that theory.

We claim that that the general principle of scrutinizing evidence extremely carefully only when it contradicts an expectation is an important part of the discovery process. When an experiment yields results that are expected, the experimental apparatus is often not examined as carefully as when the experiment yields an unexpected result. This closer scrutiny is needed to eliminate error (or fraud) as a possible explanation for the unobserved results. Similarly, when a computational model produces an expected outcome, the average researcher does not (and should not) verify that the compiler converted his program to machine code correctly. Only when a program produces unexpected results is it necessary to take such extreme measures. The benefit of this strategy is that research has more time to spend on new discoveries and additional implications of the theory. Testing these implications is an alternative, and more fruitful, means of validating the theory. One possible difference between the layperson and the scientist is the amount of care that a scientist takes to control an experiment.

3.3 Meteorites

The previous section provided an example of too easily accepting false evidence, but in some cases, scientists have too easily dismissed evidence against a current theory. A classical example

of this is the failure of the scientific community to accept evidence that meteorites had fallen from the sky. Even when a few scholars finally accepted the evidence on the origins of meteorites, the belief was not universally held.

> *I could more easily believe that two Yankee professors would lie than stones fall from heaven. Thomas Jefferson, 1807 (Cerf & Navasky, 1984.)*

Of course, evidence contrary to existing theories can eventually become overwhelming and beliefs do eventually change. However, extraordinary claims require extraordinary amounts of evidence to support them.

4.0 The role of knowledge in discovery

The process of discovery can utilize both theory-driven learning and explanation-based learning. In this section we provide examples that show how OCCAM integrates these two processes, and we discuss some limitations of the approaches.

4.1 TDL in OCCAM: balloons rising

OCCAM learns from a number of examples of children playing with balloons. One example is from a colleague who was babysitting her three-year old niece. She decided to entertain the child by inflating some balloons. The child demanded that she tie a string to the balloon. When she complied, the child asked "How come it doesn't go up?". OCCAM's theory-driven component can acquire the hypothesis that balloons with strings on them rise, by observing a number of examples in which balloons with strings on them rise while those without strings do not. These examples are consistent with children's experiences since string are usually tied to helium balloons. Theory-driven learning in OCCAM can suggest the strings as a causal candidate because it covaries with the rising and it is in spatial proximity to the balloon. OCCAM contains a causal pattern to deal with this sort of example, which can be paraphrased:

```
if an initial action on an object always occurs before a second action that
    precedes a state change,
then the initial action results in an intermediate state that enables the
    second action to result in the  state change
```

In this example, the initial action involves tying a string to a balloon and the second action is letting go of the balloon which precedes the rising of the balloon. The causal pattern hypothesizes that tying the string on the balloon results in some state that enables the balloon to rise when released.

4.2 OCCAM on catalysis: EBL and TDL

A catalyst is a substance that modifies the rate of a chemical reaction without being consumed in the process. In some cases, the catalyst changes the rate of a reaction to such an extent that the reaction occurs only in the presence of the catalyst. Inorganic reactions involving catalysis and organic reactions involving enzymes became the subject of study in the early nineteenth century. The process of catalysis was identified by Jons Jacob Berzelius in 1835:

> *Platinum, when heated to a certain temperature, had the property of supporting the combustion of alcohol vapors when mixed with atmospheric air and gold and silver lack this property...*
>
> *Thus it is certain that substances, both simple and compound, in solid form as well as in solution, have the property of exerting an effect on compound bodies which is quite different from ordinary affinity, in that they promote the conversion of component parts of the body they influence into other states without necessarily participating in the process with their own component parts....*
>
> *I shall ... call it the catalytic power of the substances, and decomposition by means of this power catalysis (Jorpes, 1966).*

Summerlin (1985) describes an example of catalysis with common household items. Nothing happens when a sugar cube is touched by a flame. However, if ash from a cigarette is rubbed on the sugar cube first, the sugar cube bursts into flame. The carbon in the ash serves as a catalyst.

OCCAM has been applied to the problem of catalysis, and can identify the rubbing of ash on a sugar cube as an action that enables the sugar cube to burst into flame. The input to OCCAM is a

conceptual dependency representation (Schank 1973) of a series of actions and state changes. The system tries to account for a state change as the effect of a known causal relationship. OCCAM 's hierarchy of schemata serves as a discrimination net that allows known causal relationships to be recognized. If there is no schema that predicts a state change, the system tries to chain together the effects of different schemata to explain a state change. If an explanation can be found by chaining, a new schema is created by explanation-based learning that allows the state change to be explained in the future via recognition. If an explanation cannot be created with the existing set of schemata, the system tries to create a new schema with theory-driven learning. If all else fails, OCCAM attempts an empirical method derived from UNIMEM (Lebowitz, 1986).

To understand the role of explanation-based learning in OCCAM, consider how explanation-based learning applies to the following example from Summerlin (1985). Hydrogen peroxide, iodine, and liquid detergent are added to water producing a large quantity of soap bubbles. Here, we assume there is a background theory that predicts (1) that in the presence of iodine, hydrogen peroxide will produce water and oxygen and (2) that a gas escaping from a solution of liquid soap will create soap bubbles. These two rules can be chained together to explain how the soap bubbles were produced. The result of this chaining can be generalized and saved so that in the future, the effect of this reaction can be found by recognition rather than chaining.

Theory-driven learning differs from explanation-based learning in that the former involves learning at the knowledge-level (Dietterich, 1986). New causal mechanisms (rather than combinations of existing causal mechanisms) are created by this learning process. Consider the example of the sugar cube. There are two training examples used by OCCAM. The first example describes the situation without the use of the catalyst (see Table 1).

TABLE 1. CD representation of touching a cube with a flame

```
(ACT TYPE (PTRANS)
     ACTOR (HUMAN NAME (LYNN)
                  HAIR (BLOND)
                  EYES (BLUE))
     OBJECT (FLAME)
     TO (P-OBJ TYPE (SUGAR-CUBE)
               COLOR (WHITE)))
```

This describes a blond child touching a flame to the sugar cube. There is no state change after this example, so this is nothing for OCCAM to explain. The example is simply indexed in memory under the **`propel`** schema. Next, the example with the catalyst is presented to OCCAM (see Table 2).

When this example is added to memory, OCCAM attempts to explain how the cube burst into flame. In this case, there are no schemata in memory that recognize this situation, nor are there any that can be chained together to produce an explanation. In order to attempts to postulate a mechanism, similar events are retrieved from memory. In this example, the only similar event that is found is the previous example. Next, the causal patterns are checked to see if any could explain the different outcome of the two similar events. In this case, the only causal pattern that

TABLE 2. CD representation of burning a sugar cube

```
(ACT TYPE (PTRANS)
     ACTOR (HUMAN NAME (LYNN)
                  HAIR (BLOND)
                  EYES (BLUE))
     OBJECT (FLAME)
     TO (P-OBJ TYPE (SUGAR-CUBE)
                       COLOR (WHITE))
     BEFORE (ACT TYPE (PTRANS)
                 ACTOR (HUMAN NAME (LYNN)
                              HAIR (BLOND)
                              EYES (BLUE))
                 OBJECT  (P-OBJ TYPE (ASH)
                                COLOR (BLACK))
                 TO (P-OBJ TYPE (SUGAR-CUBE)
                           COLOR (WHITE)))
     AFTER (STATE OBJECT (P-OBJ TYPE (SUGAR-CUBE)
                                COLOR (WHITE))
                  TYPE (BURNING)
                  VALUE (YES)))
```

applies is the one illustrated in Section 3.4 to deal with balloons rising. This causal pattern blames a difference in outcome on a different prior action: positing that rubbing ash on the sugar cube results in some intermediate state that enables the cube to burst into flame when touched by a match. This schema (illustrated in Figure 3) is indexed in memory under the **propel** schema for future use. The next time an event is encountered that meets the pattern described by this schema, OCCAM will be able to explain the outcome via recognition rather than chaining.

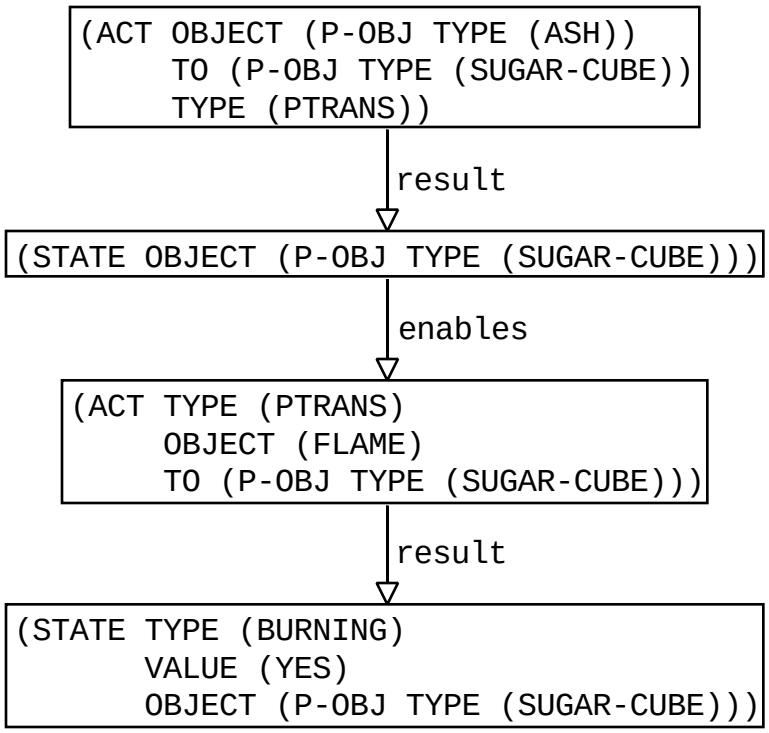

Figure 3. The causal mechanism of a catalytic process.

4.3 Limitations of theory-driven learning

Explanation-based and theory-driven learning are not intended as a complete model of the scientific discovery process. The key idea behind theory-driven learning is that knowledge of configurations that appear to be causal can focus the empirical discovery process. The causal patterns consider only a subset of the attributes that describe a situation as potentially relevant. In this manner, they are an explicit form of bias for predictive relationships. In case this bias rules out the correct hypothesis, OCCAM relies on a purely empirical method that considers all of

its input features to be potentially relevant.

In some cases, the general theory of causality does not provide enough constraints on relationships. For example, in chemical reactions there is an additional constraint that the number of atoms is not changed by the reaction. Thus, if one knows the chemical composition of the substances involved, this serves as an effective constraint on causal relationships (Langley, Zytkow, Simon & Bradshaw, 1986).

Theory-driven learning currently does not make use of quantitative information. As a consequence, it does not learn that catalysts generally increase the rate of a reaction. This was discovered shortly after catalysts were discovered:

> *Platinum sponge is able to ignite a stream of hydrogen as it escapes into the air. This discovery was soon followed by the mutual investigation of Dulong and Thenard that showed several single and compound substances have this property, but to different degrees.... Thus, this power was extended from an exceptional property to a more general one possessed by substances to different degrees.* (Jorpes 1966).

5. Discovery as argumentation

Argument analysis can help model the process of scientific reasoning in two ways. First, it provides a useful framework within which to interpret the data and steps in the process of scientific reasoning. But more importantly, heuristics that operate over argument structures can be applied in the domain of science to help specify productive research and theory formation steps to take next. We illustrate this below with two examples, the ontogeny of white dwarfs,and the chemical structure of insulin.

5.1 White dwarfs

One contemporary case, the origin of hydrogen in white dwarf stars, is an example in which new evidence forces revision of a current theory. There are two types of white dwarfs: DBs, with a

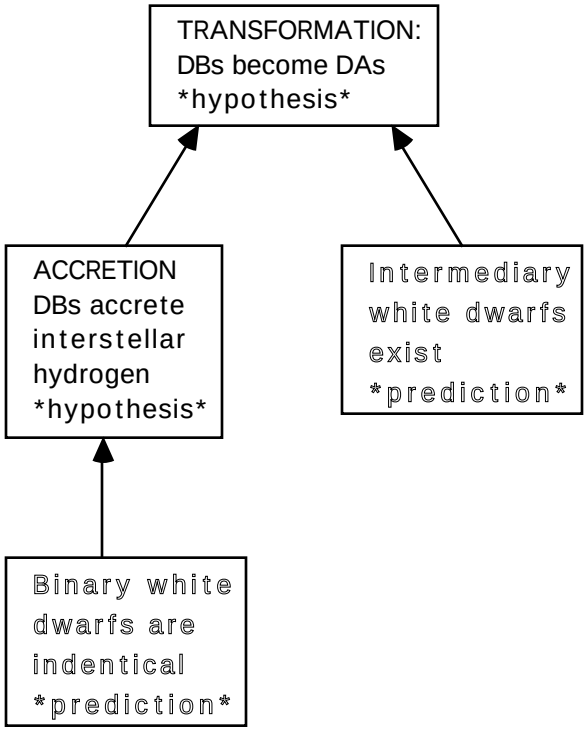

Figure 4. Theories about the origin of white dwarfs and thier predictions.

surface layer composed almost entirely of helium, and DAs, with a surface layer almost entirely of hydrogen. A current hypothesis, which we refer to as the *accretion* theory, is that all white dwarfs are created as DB's which then accrete hydrogen from the interstellar medium. A generalization supported by the accretion theory is a less specific hypothesis that DBs become DAs (by some unspecified method), which we will call the *transformation* theory. This theory predicts that intermediary dwarfs (between DA and DB) should exist. A different prediction follows from the accretion theory, that when binary white dwarfs are found, each of them will be of the same class because if accretion does occur, and both stars are in the same area of interstellar space, then they both should acquire the same type of matter. These predictions, if substantiated, each support in different ways the basic theory that DBs become DAs. Figure 4 depicts a diagram of these skeletal relationships, with the predicted support assertions outlined

and attributions of “hypotheses” or “prediction” marked. One can use these relationship to organize the analysis and incorporation of subsequent data in regards to these theories.

Recently, evidence in favor of the transformation theory has been found. A white dwarf has been discovered (G200-39 in Bootes) (Kenyon, 1988) that has both helium and hydrogen in its surface layer. In addition, the ratio of calcium to hydrogen is approximately the same as that in the interstellar medium. This provides specific evidence to support the prediction of intermediary dwarfs, thus supporting the theory that DBs become DAs. Thus, the prediction of intermediary dwarfs becomes an assertion supported by the specific observational evidence.

Another recent discovery (Oswalt, 1988), that of a binary white dwarf system, calls the accretion theory into question. In this system (L151-81A and L151-81B), one of the stars is of type DB and the other is of type DA. A mechanism that could explain how only one of the binary stars could accrete hydrogen from interstellar space occupied by them both has not been proposed. As a consequence, the theory that white dwarfs accrete hydrogen is attacked by this evidence.

The argument structure resulting from the evidential role of these two observations is depicted in Figure 5 (with the new observational evidence highlighted in bold). The prediction of intermediary dwarfs is confirmed, but finding non-identical binary dwarfs attacks the prediction of identical binary dwarfs. Because this prediction is the only support for the accretion theory, the entire theory is called into question. Because the accretion theory is called into question, it no longer serves as support for the transformation theory, but since alternative support for the transformation theory has been found, that theory is still a viable one. Thus, determining the current validity of any particular theory, is context sensitive, depending on the full set of possible supports and attacks.

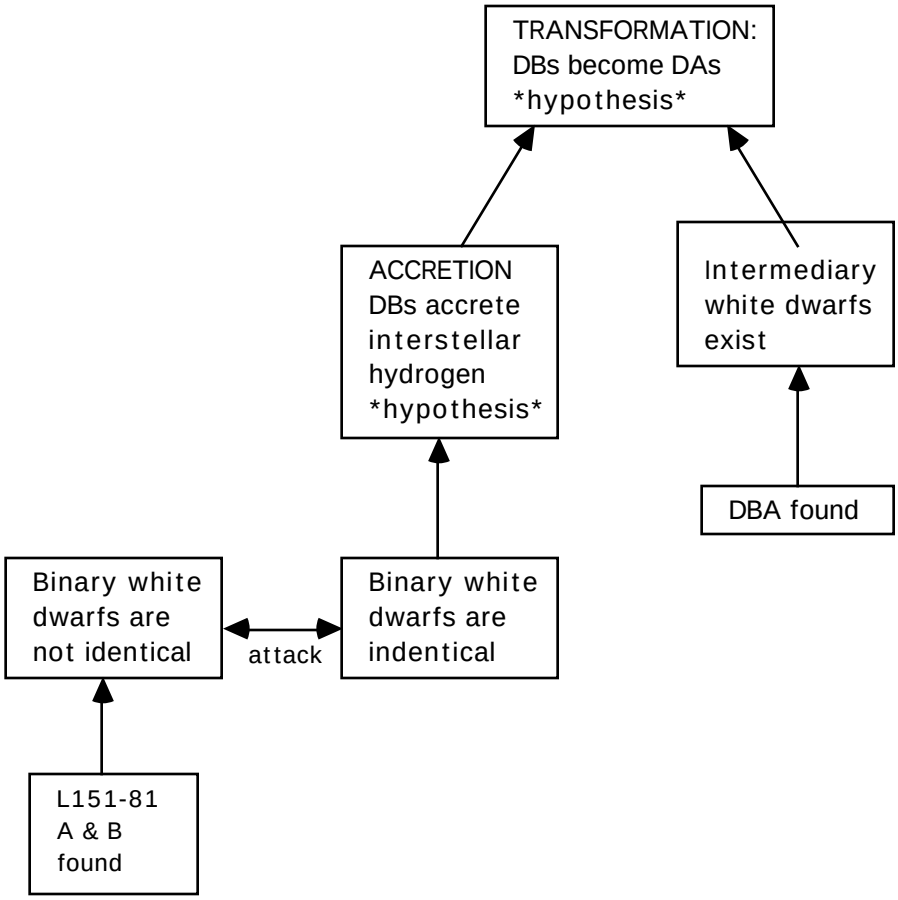

Figure 5: Evidence for White Dwarf Theories

At this point, argument heuristics utilizing this kind of argument analysis can suggest appropriate responses to problematic observations in the scientific domain and focus attention on potentially fruitful research directions. For example, a general argumentation heuristic is that when contradictory evidence is found for a theory, several responses can be made: attack the validity of the evidence itself, find an alternate explanation for the evidence that does not compete with the theory, or accept the contradiction and find an alternate theory. This general argument heuristic provides three reasonable research directions in response to the discovery of L151-81 A and B as a counterexample to the accretion theory. The researcher can look for ways to attack the validity of this finding, such as instrumentation error, or data interpretation error. A second direction is to find an alternative explanation for L151-81 A and B that does not contradict the accretion theory. Finally, the researcher might accept that the accretion theory is incorrect, but

since there is still valid evidence that the transformation theory holds, develop a debugged accretion theory, or an entirely different theory, that is consistent with the transformation theory and is not contradicted by L151-81 A and B (nor by the existence of other extant evidence such as the presence of intermediary dwarfs). A theory-driven learning component, such as that used by OCCAM may be used to patch the theory or find a new theory.

5.2 Insulin

In the 1950s, research into the disposition of insulin in the body led to the discovery that small peptides could stimulate the production of antigens. This research, recounted in Yalow (1981), illustrates the rhetorical interplay between theory formation and the introduction of new data showing how techniques of argumentation participate in the process of scientific discovery.

Yalow's group was investigating Mirsky's hypothesis that diabetes, a disease in which the patient suffers the effects of insufficient insulin, was caused by an abnormally rapid degradation of insulin. The first step of this investigation was to document the specific turnover of insulin in the body. Drawing an analogy between albumin and insulin, they borrowed the technique of labeling from prior studies on the turnover of labeled albumin in the body, and measuring the rate of turnover of labeled insulin. The rhetorical device of analogy -- recognizing similarities between current scientific problems and prior ones, and then adapting known techniques to the new situation -- is fundamental to scientific investigation.

Contrary to expectation, the rate of insulin turnover in the body was found to be slower in diabetics than in normals. This contradicts the initial hypothesis being investigated, with two results. The hypothesis that diabetes is caused by rapid degradation of insulin has no other support, so it was effectively retracted. In addition, the researchers hypothesized a new assertion supported by the observations of the labeled insulin studies: that something about diabetes itself

retards the disappearance of insulin. The argument relationships here and following are depicted in Figure 7.

At this point, selecting what data to focus on was crucial. The retarded insulin hypothesis was also being investigated by another group, who viewed the situation as:

```
If diabetes
   then insulin degradation is slow
   else insulin degradation is normal
```

However, Yalow's group developed a competing hypothesis based on additional observations. They found two populations that did not fit this pattern. *Normal* turnover was found in diabetics who were only recently diagnosed or who were not insulin-dependent, and *slow* turnover was found in non-diabetic schizophrenics. One subject was found who had been recently diagnosed as diabetic: initially he had normal insulin degradation, but a few months later he demonstrated delayed insulin degradation.

Selective focusing on the new data (organized in Figure 6) forced a shift in the analysis of the data. As it turns out, one kind of treatment given to schizophrenics was "insulin shock therapy". Thus, retarded insulin degradation is associated not with whether or not a subject has diabetes, but whether or not the subject has ever been exposed to exogenous insulin. This can be stated as two rules:

```
If prior insulin exposure
   then insulin degradation is slow
   else insulin degradation is normal.

also:

If insulin degradation is slow
   then prior insulin exposure
   else no prior insulin exposure
```

This relationship between insulin exposure and slow insulin degradation forced the assumption of a new hypothesis: that prior insulin exposure led to the development of antibodies which bound to circulating insulin, producing a larger molecule that was thus eliminated from

circulation more slowly.

Insulin Degradation Rate?		
Diabetic?	normal rate	slow rate
no	healthy subject	schizophrenic
yes	non-insulin dependent -or- recently diagnosed	insulin-dependent

Figure 6: Insulin degradation rates from Yalow (1981).

There were two problems with taking this approach. First, insulin therapy had begun in 1921 and the antigenicity of insulin had never been observed. Second, standard immunology said that peptides smaller that 10,000 molecular weight could not be antigenic, yet insulin's molecular weight was about 6000. Each provides support for the assertion that small peptides cannot provoke antigen formation, which contradicts the hypothesis based on the population observations.

At this point, the research focused on two contradictory assertions regarding whether or not antibodies can bind to insulin. Argument heuristics utilizing the supports for the two contradictory assertions specify appropriate subsequent research steps. The objection that "we would have noticed" can be countered only indirectly. However, one can contradict the non-antigenicity of small peptides by demonstrating the binding in insulin (or other small peptides). This in fact is what happened: insulin binding only in those treated with insulin was demonstrated in a number of ways. At this point, one side of the contradiction was grounded in

actual observation, while the other is grounded only in tenet and is thus invalid. Consequently, by inferential analysis, one could assert in general that small peptides can produce antigens.

Assertions that attack the beliefs of others are often rejected, and establishing the validity of novel beliefs that contradict accepted knowledge can be difficult. Yalow's results were rejected by numerous journals, in spite of the numerous studies demonstrating insulin binding, because they violated basic tenants of immunology. It was published only after rewording compromises such as "insulin binding globin" versus "antigen", thus establishing the antigenicity of small peptides.

Much of the argumentation process consists of the interaction of competing hypotheses. In this example there are three major cases: slow versus fast degradation of insulin, diabetic versus

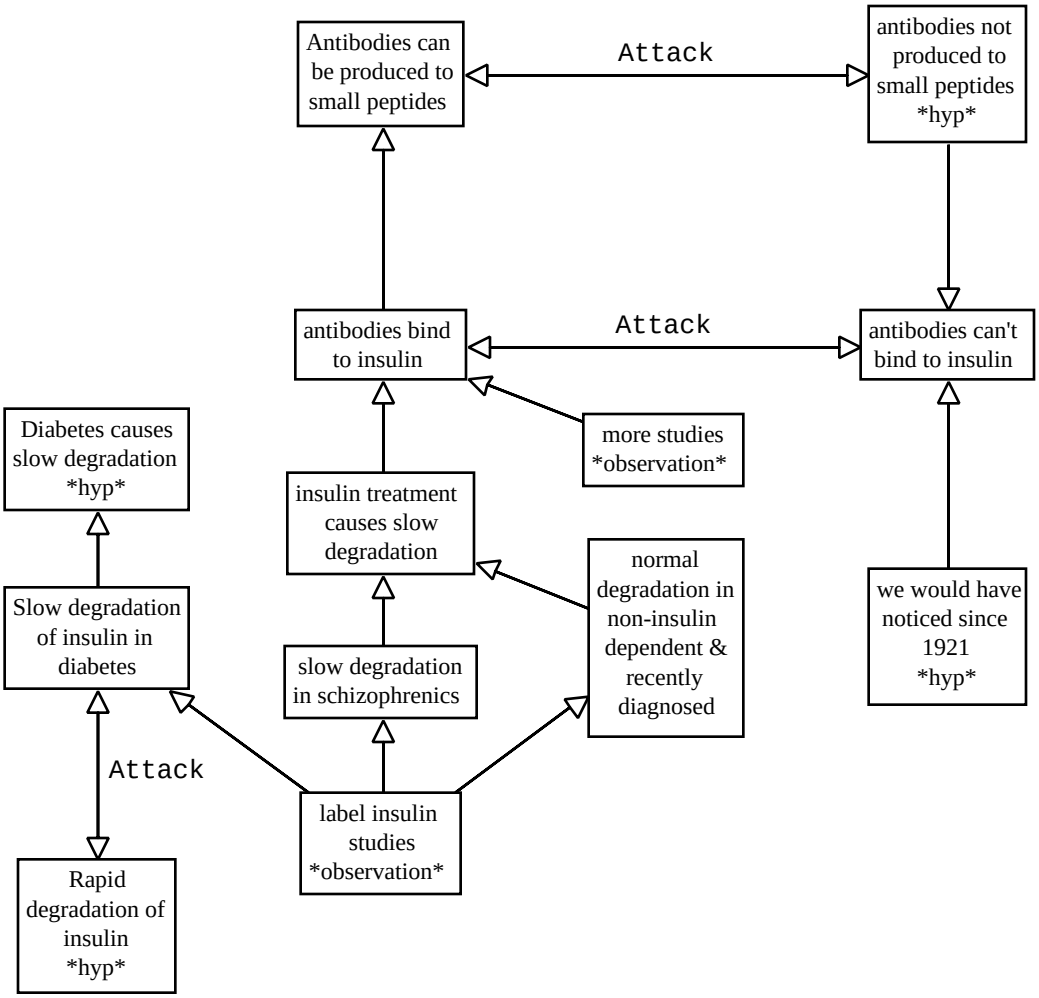

Figure 7: Argument Analysis for competeing theories of the disposition of insulin

insulin treatment as causes of slow insulin degradation, and whether or not antibodies can bind to small peptides (including insulin). Numerous other belief-oriented rhetorical steps were involved: assertion of new beliefs, finding support for beliefs, attacking beliefs of others, inferential analysis of beliefs, and rejection of beliefs that violate other beliefs, retraction of old beliefs shown to be incorrect. Viewing data and theories as beliefs, and applying argument heuristics to them, is a useful technique modeling the process of scientific research.

6.0 Conclusions

Two immediate points have been made in this chapter. First, models of scientific discovery should incorporate a number of kinds of knowledge: empirical laws, understanding of process, specification of intermediate state, causal constraints, and argument structure and heuristics. Second, models of scientific discovery need to incorporate everyday "layperson" techniques of reasoning and creativity. What is the potential benefit of taking such an approach?

Over a decade ago, Lenat (1982) reported on AM, a model of discovery which used heuristics of "interestingness" as a focusing technique in the analysis of data. We claim that additional focusing techniques, such as those provided by prior knowledge play a major role in discovery in the intuitive scientist and the practicing scientist. Yet much subsequent work in scientific discovery has concentrated primarily on what to do with given experimental data, rather than on techniques for deciding which experimental data to focus upon. The first step of hypothesis and theory formation is not explanation of data, but noticing or selection of what particular data is worth looking at to begin with. In this paper we have identified some kinds of knowledge and techniques that embody some of the "biases" people use to help in this focusing task: causal analysis, use of prior knowledge and generalizations, and techniques of belief formation and analysis of argumentation structure. In an integrated, ecologically valid model of scientific discovery, decisions must explicitly be made about the meaning, import, and disposition of raw data and developed theories. The features we have discussed contribute to making such

decisions for productive scientific reasoning.

References

Abelson, R. P. (1973). The Structure of Belief Systems. In *Computer Models of Thought and Language*. Edited by R.C. Schank and K.M. Colby. San Francisco: Freeman.

Alvarado, S., , & Flowers, M. (1985). Memory Representation and Retrieval for Editorial Comprehension. *Seventh Annual Conference of the Cognitive Science Society* (University of California, Irvine), 228-235.

Alvarado, S., Dyer, M., & Flowers, M. (1986). Editorial Comprehension in OpEd through Argument Units. *Proceedings of the Fifth National Conference on Artificial Intelligence* (University of Pennsylvania, Philadelphia), 250-256.

August, S & Dyer, M. (1985). Understanding Analogies in Editorials. *Proceedings Ninth International Joint Conference for Artificial Intelligence* (IJCAI-85) (Los Angeles, CA).

Birnbaum, L., Flowers, M., and McGuire, R. (1980). Towards an AI Model of Argumentation. *Proceedings of the First Annual National Conference on Artificial Intelligence* (Stanford, CA), 313-315.

Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). *A Study of Thinking.* New York: Wiley.

Bullock, M., Gelman, R. & Baillargeon, R. (1982). The development of causal reasoning. In Friedman, W. (Ed.), *The developmental psychology of time*. New York, NY: Academic Press.

Chapman, L.J., & Chapman, J.P. (1969). Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs. Journal of Abnormal Psychology, 74, 271-280.

Cerf, C. and V. Navasky. (1984). *The Experts Speak: The Definitive Compendium of Authorative Misinformation.* Pantheon.

Cohen, L. (1981). Can Human Irrationality be Experimentally Demonstrated? *Behavior & Brain Sciences* 4 : 317-331.

DeJong, G. and Mooney, R. (1986). *Explanation-based learning: An alternate view*. Machine Learning, Vol. 1(2).

Dietterich, T. (1986). *Learning at the knowledge level.* Machine Learning, 1(3), 287-315.

Eddy, D. (1982). Probabilistic Reasoning in Clinical Medicine: Problems and Opportunities. In *Judgment under Uncertainty: Heuristics and Biases*. Edited by D. Kahneman, P. Slovic, and A. Tversky. 249-267. New York: Cambridge University Press.

Elstein, A., L. Shulman, and S. Sprafka. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning.* Cambridge, MA: Harvard University Press.

Faust, D. (1984). *The Limits of Scientific Reasoning.* Minneapolis: University of Minnesota Press.

Feifer, R G., Dyer, M., and Flowers, M. (1986). Teaching Inferencing Strategies. *Annual Meeting of American Educational Research Associations* (San Francisco), 1-12. also UCLA-AI-85-15.

Fiefer, R. (1989). *An Intelligent Tutoring System for Graphic Mapping Strategies.* Ph.D., University of California, Los Angeles.

Flowers, M, McGuire, R, & Birnbaum, L. (1982). Adversary Arguments and the Logic of Personal Attacks. In *Strategies for Natural Language Understanding.* Edited by W.G. Lehnert and M.G. Ringle. Hillsdale NJ: Lawrence Erlbaum.

Greenwald, A., A. Pratkanis, M. Lieppe, and M. Baumgardner. (1986). Under What Conditions Does Theory Obstruct Research Progress? *Psychological Review* 93 : 216-229.

Gorman, M. E. (1986). How the Possibility of Error Affects Falsification on a Task that Models Scientific Problem Solving. *British Journal of Psychology* 77 : 85-96.

Hadamard, J. (1954). *The Psychology of Invention in the Mathematical Field.* New York: Dover.

Hovey, D. (1962). *Experience and Insight: Experiments on Problem-Solving and a Case Study of Scientific Thinking.* dissertation, University of Colorado, Boulder. (Analysis of Franklin's work.)

Jenkins, R. V. (1983). Elements of Style: Continuities in Edison's Thinking. *Annals of the New York Academy of Sciences* (424) : 149-162.

Jenkins, R. V. & Jeffrey, T. E.. (1984). Worth a Thousand Words: Nonverbal Documents in Editing. *Documentary Editing* 6 : 1-8.

Jennings, D., Amabile, T. & Ross, L. (1980). Informal covariation assessment: data-based vs. theory-based judgments. In A. Tversky, D. Kahneman & P. Slovic (Eds.) *Judgments under uncertainty: Heuristics and Biases.* New York: Cambridge University Press.

Jorpes, J. E.(1966). *Jac. Berzelius: His Life and Work.* Almquist &Wi.skell, Stockholm

Kelley, H. (1973). The process of causal attribution. *American Psychologist*, 28, pp. 107-128.

Klahr, D. & Dunbar, K. (1988). Dual Space Search During Scientific Reasoning. *Cognitive Science* 12 : 1-48.

Kuhn, D., Amsel,E., & O'Loughlin, M. (1988). *The Development of Scientific Thinking Skills.* Edited by H. Beilin. Developmental Psychology Series. San Diego: Academic Press.

Langley, P. Bradshaw G. & Simon, H. (1983). Rediscovering chemistry with the BACON system. In Michalski, R., Carbonell, J., and Mitchell, T. (Ed.), *Machine Learning: An Artificial Intelligence Approach.* Palo Alto, Ca.: Tioga Publishing Co.

Langley, P., Zytkow, J., Simon, H., and Bradshaw, G. (1986). The search for regularity: Four aspects of scientific discovery. In Michalski, R., Carbonell, J., & Mitchell, T. (Ed.), *Machine Learning, An artificial Intelligence Approach, Vol 2.* Los Altos, Ca.: Morgan Kaufmann Publishers.

Lebowitz, M. (1986). Concept Learning in an rich input domain: Generalization-based Memory. In Michalski, R., Carbonell, J., & Mitchell, T. (Ed.), *Machine Learning, An artificial Intelligence Approach, Vol 2.* Los Altos, Ca.: Morgan Kaufman Publishers.

Lenat, D. (1982). AM: An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search. In *Knowledge-based Systems in Artificial Intelligence.* Edited by R. Davis and D.B. Lenat. New York: McGraw-Hill.

Lord, C., Ross, L., & Lepper, M. (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology* 37 : 2098-2109.

Mahoney, M. and T. Kimper. (1976). From Ethics to Logic: A Survey of Scientists. In *Scientist as Subject: The Psychological Imperative.* Edited by M. Mahoney. 187-194. Cambridge, MA: Ballinger..

McGuire, R., Birnbaum., L, & Flowers, M. (1981). Opportunistic Processing in Arguments. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence* (Vancouver, BC).

Mitchell, T., Kedar-Cabelli, S. & Keller, R. (1986). *Explanation-based learning: A unifying view.* Machine Learning, Vol. 1(1).

Nisbett, Richard and Lee Ross. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment.* Edited by J.J. Jenkins, W. Mischel, and W.W. Hartup. The Century Psychology Series. Englewood Cliffs, NJ: Prentice-Hall, Inc.

O'Brien, D. P., Costa, G, & Overton, C,. (1986). Evaluations of Causal and Conditional Hypotheses. *Quarterly Journal of Experimental Psychology* 38A : 493-512.

Osowski, J. V. (1986). *Metaphor and Creativity: A Case Study of William James.* dissertation, Rutgers University.

Oswalt, T, Hintzen, P., Liebert, J. & Sion, E. (1988). *L151-81A/B: A Unique white dwarf binary with DB and DA components.* The Astrophysical Journal Vol 333, Number 2, Part 2, L87-L89

Pazzani, M., Dyer, M. & Flowers, M. (1986). *The role of prior causal knowledge in generalization.* Proceedings of the National Conference on Artificial Intelligence. American Association for Artificial Intelligence, Morgan-Kaufmann.

Pazzani, M. & Schulenberg, D. (1989). *The role of prior knowledge in the acquisition of conjunctive and disjunctive concepts.* Proceedings of the Eleventh Meeting of the Cognitive Science Society.

Pazzani, M. (in press) *Creating a Memory of Causal Relationships: An integration of empirical and explanation-based learning methods.* Lawrence Erlbuam Associates, Hillsdale, NJ .

Quilici, A., Dyer, M., and Flowers, M. (1986). AQUA: An Intelligent UNIX Advisor. *Proceedings of Eurpean Conference on Artificial Intelligence* (ECAI-86) (Brighton, UK).

Quilici, A., Dyer, M., and Flowers, M. (1989). Recognizing and Responding to Plan-Oriented Misconceptions. *Computational Linguistics* .

Rouche, B. (1984), *The Medical detectives, Volume II.* Washington Square Press.

Schank, R. C. (1973). Identification of Conceptualizations Underlying Natural Language. In *Computer Models of Thought and Language.* Edited by R.C. Schank and K.M. Colby. San Francisco: Freeman.

Schank, R. C. (1975). *Conceptual Information Processing*. Amsterdam: North-Holland.

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

Shultz, T. & Kestenbaum, N. (1985). *Causal reasoning in children*. Annals of Child Development, 195-249.

Stitch, S. and Nisbett., (1980). Justification and the Psychology of Human Reasoning. *Philosophy of Science* 47 : 188-202.

Summerlin, L. & Ealy, J. (1985) *Chemical Demonstrations: A source book for teachers.* American Chemical society.

Thagard, P. and Greg Nowak (1989). *Conceptual structure of the the geological revolution.* This volume.

Toulmin, S. (1958). *The Uses of Argument*. Cambridge, MA: Cambridge University Press.

Toulmin, S., R. Reike, and A. Janik. (1979). *An Introduction to Reasoning.* New York: Macmillan.

Tulving, E. (1972). Episodic and Semantic Memory. In *Organization of Memory.* Edited by E. Tulving and W. Donaldson. New York: Academic Press.

Tweeney (1989). *Issues arising from the study of Faraday's notebooks.* This volume.

Valiant, L. (1984). *A theory of the learnable.* Communications of the Association of Computing Machinery, 27(11), 1134-1142.

Wertheimer, M. (1959). *Productive Thinking*. enlarged edition ed. New York: Harper & Row. Original work published 1945.

Yalow, R. S. (1981). Biomedical Investigation. In *The Joys of Research*. Edited by W. Shropshire. Smithsonian.