The Role of Prior Causal Theories in Generalization

Michael Pazzani, Michael Dyer, Margot Flowers
Artificial Intelligence Laboratory
3531 Boelter Hall
UCLA
Los Angeles, CA 90024

Abstract

OCCAM is a program which organizes memories of events and learns by creating generalizations describing the reasons for the outcomes of the events. OCCAM integrates two sources of information when forming a generalization:

- Correlational information which reveals perceived regularities in events.
- Prior causal theories which explain regularities in events

The former has been extensively studied in machine learning. Recently, there has been interest in explanation-based learning in which the latter source of information is utilized. In OCCAM, prior causal theories are preferred to correlational information when forming generalizations. This strategy is supported by a number of empirical investigations. Generalization rules are used to suggest causal and intentional relational relationships. In familiar domains, these relationships are confirmed or denied by prior causal theories which differentiate the relevant and irrelevant features. In unfamiliar domains, the postulated causal and intentional relationships serve as a basis for the construction of causal theories.

Introduction

When learning the cause of a particular event, a person can utilize two sources of information. First, a person may detect a similarity between the event and previous events noticing that whenever C occurs, R also occurs. After noticing such a correlation, a person might induce that C causes R. Secondly, a person may use his prior causal theories.

In machine learning, the correlational techniques have been extensively studied (e.g. [4, 12, 23, 18, 32]). More recently, there has been interest in explanation-based learning [9, 22, 24, 25, 28] in which prior knowledge used to understand an example guides the generalization process [31]. For example, the first time a merchant asks the customer if he wants the carbon paper from a credit card purchase, the customer may wonder why the merchant is doing this. A clever person might be able to deduce that by taking the carbon paper, he can prevent a thief from retrieving the carbon paper from the merchant's garbage. Since the carbon paper contains his credit card number, the thief can make mail and phone purchases with the credit card number. Since this is a somewhat complicated inference, it would be advantageous to remember it, rather than rederive it the next time it is needed.

The topic of this paper is how these two sources of information, correlation of feature over a number of examples, and prior causal theories can be combined. There are a number of possibilities:

- Correlational information is used exclusively.
- Correlational information is preferred to prior causal theories.
- Prior causal theories are preferred to correlational information.
- Prior causal theories are used exclusively.

In the remainder of this paper, we first discuss some examples of learning programs. Next, we review a number of experiments which assess how correlational information is combined with prior causal knowledge. Finally, we present an overview of our theory as implemented in OCCAM, a program under development at UCLA which integrates correlational and explanation-based learning.

Related Work

Correlational Learning

In this section, we discuss IPP [18] a program which uses correlational information exclusively. IPP was selected to exemplify correlational learning because a recent extension [20] also adds explanation-based capabilities. IPP is a program that reads, remembers, and makes generalizations from newspaper stories about international terrorism.

IPP starts with a set of MOPs [30] which describe general situations such as extortion. After adding examples of events to its memory, it creates more specialized MOPs (spec-MOPs). Spec-MOPs are created by noticing the common features of a number of examples. Not all features of a generalization are treated equally. Some features are predictive; their presence allows IPP to infer the other features if they are not present. The predictive features are those that are unique to that generalization. The features that appear in a large number of generalizations are non-predictive. IPP keeps track of the number of times a feature is included in generalizations. The idea is that the predictive features are likely to be causes of the non-predictive features.

Since IPP makes no attempt to explain its generalizations, it may include irrelevant information which is coincidentally true in a generalization. A mechanism to identify and correct erroneous generalizations when further examples are added to memory was included in IPP. This mechanism was later extended in UNIMEM, the generalization and memory component of IPP [19].

Explanation-based Learning

GENESIS [25] is an example of a system which exclusively uses its prior causal theories in generalization. GENESIS accepts English language input of stories and produces a conceptual representation of the story. The conceptual representation contains a justification for the outcome of the story and the actions of the actors in terms of causal and intentional relationships. If an actor achieves a goal in a novel manner, the explanation for how the goal was achieved is generalized into a schema which can be used for understanding future events. This generalization process notes which parts of the conceptual representation are necessary for establishing the causal and intentional relationships. For example, consider how GENESIS learns about kidnapping. Part of this process is determining why the ransom is paid. Given only one example of a kidnapping in which a father pays the ransom for his daughter who is wearing blue jeans, GENESIS incorporates this fact into its schema: there must be a positive interpersonal theme [10] between the victim and the person who pays the ransom. This generalization is possible from just one example because the inference that explains why the ransom is paid contains the precondition that there be a positive interpersonal theme. In contrast, a correlational learner might have to see another kidnapping where the victim was not wearing blue jeans to determine that clothing is not relevant in kidnapping. Similarly, a correlational learner might have to see many examples before the father-daughter relationship could be generalized to any positive interpersonal theme.

At first, explanation-based learning may seem confusing. After all, isn't it just learning what is already known? For example, the schemata which GENESIS constructs encodes the same information which is in the inference rules used to understand the story. However, explanation-based learning serves an important role. Understanding using inference rules can be combinatorially explosive. Consider understanding the following story if there were no kidnapping schema:

The teenage girl who was abducted on her way to school Monday morning was released today after her father left $50,000 in a trash can in a men's room at the bus station.

If there were no kidnapping schema, a very complex chain of inference is necessary to determine why the father put money in a trash can and why the teenage girl was released. In contrast, understanding this story with a kidnapping schema is simplier since the kidnapping schema records the inferences necessary to understand the relationship between the kidnapper's goal of possessing money and the father's goal of preserving the health of his child. Explanation-based learning produces generalizations which simplify understanding by storing the novel interaction between a number of inference rules. In this respect, the goals (but not the mechanism) of explanation-based learning are similar to those of knowledge compilation [1] and chunking [17].

Of course, in some domains an understander might not have enough knowledge to produce a detailed causal and intentional justification for an example. In such cases, explanation-based learning is not applicable and an understander might have to rely solely on correlational techniques.

UNIMEM

UNIMEM [20] is an extension to IPP which integrates correlational and explanation-based learning. UNIMEM operates by applying explanation-based learning techniques to correlational generalizations rather than instances. Hence, UNIMEM prefers correlation information to prior causal theories. UNIMEM first builds a generalization and identifies the predictive and non-predictive features. Then, it treats the predictive features as potential causes and the non-predictive features as potential results. Backward-chaining production rules representing domain knowledge are utilized to produce an explanation of how the predictive features cause the non-predictive. If no explanation is found for a non-predictive feature it is considered a coincidence and dropped from the generalization. There are two possible reasons that a predictive feature might not be used to explain non-predictive features: either it is irrelevant to the generalization (and should be dropped from the generalization) or the feature may in fact appear to be cause (i.e., predictive) due to a small number of examples but in fact be a result. To test the later case, UNIMEM tries to explain this potential result in terms of the verified predictive features.

The rational behind using correlational techniques to discover potential causal relationships which are then confirmed or denied by domain knowledge is to control the explanation process. It could be expensive or impractical to use brute force techniques to produce an explanation. Since the predictive features are likely to be causes, UNIMEM's explanation process is more focused. However, since UNIMEM keeps track of the predictability of individual features rather than combinations of features it can miss some causal relationships. This occurs when no one feature is predictive of another but a conjunction of features is. For example, in kidnapping when the ransom is demanded from a rich person who has a positive interpersonal relationship with the hostage one could predict that the ransom would be paid. Of course, if the ransom were demanded from a poor relative or rich stranger, the prediction should not be made.

Experimental Data

How do people combine correlational information with prior causal theories? There have been a number of experiments in social psychology which assess the ability to learn causal relationships. Some of these are motivated by Kelley's attribution theory [15, 14]. Kelley proposed that the average person makes causal inferences in a manner analogous to a trained scientist. Kelley's covariation principle is similar to Lebowitz's notion of predictability*:

The covariation principle is based on the assumption that effects covary over time with their causes. The "with" in this statement conceals the important and little-studied problem of the exact temporal relations between cause and effect. The effect must not, of course, precede a possible cause... [14 page 7]

However, this view is not without criticism:

There is no assumption as critical to contemporary attribution theory (or to any theory that assumes the layperson's general adequacy as an intuitive scientist) as the assumption that people can detect covariation among events, estimate its magnitude from some satisfactory metric, and draw appropriate inferences based on such estimates. There is mounting evidence that people are extremely poor at performing such covariation assessment tasks. In particular, it appears that a priori theories or expectations may be more important to the perception of covariation than are the actually observed data configurations. That is, if the layperson has a plausible theory that predicts covariation between two events, then a substantial degree of covariation will be perceived, even if it is present only to a very slight degree or even if it is totally absent. Conversely, even powerful empirical relationships are apt not to be detected or to be radically underestimated if the layperson is not led to expect such a covariation. [26 page 10]

Some work in developmental psychology is also relevant to determining how prior causal theories are used. In particular, since younger children have less knowledge about the world, they are less likely to have prior causal theories.

Perceiving Causality

One of the earliest inquiries into causality was conducted by Michotte [21]. He conducted a series of experiments to determine when people perceive causality. In one experiment, subjects observed images of discs moving on a screen. When the image of one disc bumped a stationary disc and the stationary disc immediately began to move, subjects would state that the bumping caused the stationary disc to move. Michotte called this the Launching Effect.

However, if the stationary disc starts moving one fifth of a second after it is bumped, subjects no longer indicate that the bumping caused the motion. Here we have an example of a perfect correlation in which people do not perceive causality. From this experiment, it is clear that correlation alone is not enough to induce causality. A similar finding was reported by Bullock [5]. Children as young as five will not report causality if there is a spatial separation between the potential cause of motion and the potential result.

Illusory Correlation

Chapman and Chapman performed a series of tests to determine why practicing clinical psychologists believe that certain tests with no empirical validity are reliable predictors of personality traits. For example, in one study [6], clinical psychologists were asked about their experience with the Draw-a-Person Test (DAP). In this test, a patient draws a picture of a person which is analyzed by the psychologist. Although the test has repeatedly been proved to have no diagnostic value, 80% of the psychologists reported that men worried about their manliness draw a person with broad shoulders and 82% stated that persons worried about their intelligence draw an enlarged head. In the second experiment in this study, the Chapmans asked subjects (college undergraduates) to look at 45 DAP tests paired with the personality trait of the person who (supposedly) drew them. The subjects were asked to judge what sort of picture a person with certain personality traits did draw. Although the Chapmans paired the pictures with traits so that there would be no correlation, 76% of the subjects rediscovered the invalid diagnostic sign that men worried about their manliness were likely to draw a person with broad shoulders and 55% stated that persons worried about their intelligence drew an enlarged head. In the next experiment, the Chapmans asked another set of subjects about the strength of the tendency for a personality trait to call to mind a body part. For example, subjects reported a strong association between shoulders and manliness, but a weak association between ears and manliness. For four of the six personality traits studied, the body part which was the strongest associate was the one most commonly reported as having diagnostic value by clinical psychologists and subjects. In a final experiment, subjects were presented DAP Tests which were negatively correlated with their strong associates. In this study, subjects still found a correlation between personality traits and their strong associates but to a lesser degree (e.g., 50% rather than 76% reported that broad shoulders was

*The primary difference between predictability and covariation is that covariation implies that there is a unique cause: whenever the result is present, the cause is present and whenever the cause is present, the result is present. In contrast, predictability only requires that whenever the cause is present, the result is present.

a sign of worrying about manliness). The Chapmans labeled this phenomenon "illusory correlation".

A similar finding was found for particular Rorschach cards which have no validity [7]. These experiments clearly demonstrate that covariation may be noticed when it is not actually present if there is a reason to suspect covariation. Conversely, actual covariation may go undetected if it unexpected. Due to the phenomenon of illusory correlation, Kelley has qualified the covariation principle to apply to perceived rather than actual covariation.

Developmental Differences

Ausubel and Schiff [3] asked kindergarten, third and sixth grade students to learn to predict which side of a tetter-totter would fall when the correct side was indicated by a relevant feature (length) or an irrelevant feature (color). They found that the kindergarten children learned to predict on the basis of relevant or irrelevant features at the same rate. However, the older children required one third the number of trials to predict on the basis of a relevant feature than an irrelevant one.

Presumedly, the older children had a prior causal theory which facilitated their learning: a tetter-totter falls on the heavier side and the longer side is likely to be the heavier side. The younger children had to rely solely on correlation. Their performance on learning in the relevant and irrelevant conditions were comparable to the older children in the irrelevant condition.

Correlation in Animals

In classical conditioning, it is claimed that animals make use of correlations between any conditioned stimulus and an unconditioned stimulus. However, Garcia and his coworkers [11] have shown that there are limits to what rats will accept as an unconditioned stimulus. In this experiment, Garcia paired two different conditioned stimuli (size of food pellet and flavor of food pellet) with two different unconditioned stimuli (immediate pain via electric shock, delayed illness by exposure to x-ray). The rats were able to make an association between flavor and delayed illness, but not between size and delayed illness. Additionally, the rats learned the association between size and immediate pain, but not between flavor and immediate pain.

While it may not be a causal theory in the strongest sense, it does appear that rats have an innate mechanism to relate illness to the flavor of food:

Since flavor is closely related to chemical composition, natural selection would favor associative mechanisms relating flavor to the aftereffects of ingestion. [11 page 795]

Analysis

There is considerable evidence that in people, prior causal theories are preferred to correlational information. Why should this be so? Should we design computer learning programs to do the same? There are a number of advantages to preferring causal theories:

- As demonstrated by Ausubel and Schiff, prior knowledge can facilitate the learning process. Fewer examples are necessary to arrive at the correct generalization.
- Detecting correlation among many events with a large set of features places great demands on memory. Many previous examples must be recalled and compared in correlational learning. Explaining an example with prior causal theories is less demanding.

Of course, correlational information is quite important in an unfamiliar domain. The ideal combination would be to use prior causal theories when possible, but to use correlation information to learn the causal theories in an unfamiliar domain. This is the strategy used by OCCAM.

OCCAM

OCCAM is a program which maintains a memory of events in several domains. As new events are added to memory, generalizations are created which describe and explain similarities and differences between events. This paper focuses on the generalization process. Details of the memory organization are found in [27].

In one domain, OCCAM starts with general knowledge about coercion and family relationships. After some examples, it creates a generalization which describes a kind of kidnapping (along with the explanation that a family member of the victim's family pays the ransom to achieve the goal of preserving the victim's health). Further examples create a specialization of this generalization which represents an inherent flaw in kidnapping: the victim can testify against the kidnapper, since the kidnapper can be seen by the victim. After some more examples, a similarity is noticed about the kidnapping of blond infants. This coincidence starts an explanation process which explains the choice of victim to avoid a possible goal failure, since infants cannot testify. Because the hair color of the victim was not needed to explain the choice of victim, it is not included in the generalization. Since there is a lot of background knowledge, OCCAM can use explanation-based techniques in this domain. Some of the knowledge of family relationships (e.g., parents have a goal of preserving the health of their children) was learned using correlational techniques.

In another domain, OCCAM starts with no background knowledge and is presented with data from a protocol of a 4-year old child trying to figure out why she can inflate some balloons but not others. Since there are no prior causal theories in this domain, OCCAM uses correlation techniques to build a causal theory.

Generalization in OCCAM

In this section we present the generalization strategy used by OCCAM. When a new event is added to memory the following generalization process occurs:

1. **Find the most specific generalization applicable to the new event.**
2. **Recall previous events which are similar to the new event.**

Individual events in OCCAM's memory are organized by generalizations. An individual event is indexed by its features which differ from the norm of the generalization [16]. Events similar to the new event may be found by following indices indicated by the features of the new event. After similar events are found, a decision must be made to determine if the new event and other similar events are worth generalizing. DeJong [9] gives a number of criteria to determine if a single event is worth generalizing (e.g., Does the event achieve a goal in a novel manner)...To his list, we add *an event should be generalized if a similar event has been seen before.* The idea here is that if two similar events have been seen, it is possible that more similar events will occur in the future. It is advantageous to create a generalization to facilitate understanding of future events. If the new event is not generalized, it is indexed under the most specific generalization found in Step 1. Otherwise, generalization is attempted:**

3. **Postulate an explanation for the similarities among the events.**

Generalization rules postulate causal or intentional relationships. Typically a generalization rule suggests a causal explanation for a temporal relationship. For example, the simplest generalization rule is "If an action always precedes a state, postulate the action causes the state". Generalization rules serve the same purpose that predictability serves in UNIMEM: to focus the explanation process. However, the experimental evidence reviewed earlier seems to cast doubt on the assertion that people use predictability as the sole indicator of causality. Instead, OCCAM uses rules which focus on answering two important questions:

- Why do people do the things they do?
- What caused the outcome to occur?

If a potential explanation (in terms of human motivation or physical causality) for the similarity among a number of events is found, the next step is to verify the explanation:

4. **Postulated causal and intentional explanations are confirmed or denied using prior causal theories.**

If prior causal theories confirm the explanation, a new generalization is created. This type of generalization is called an explanatory generalization. As in explanation-based learning, the features of the new generalization are those which are necessary to establish the causal relationship. The relevant features depend on the prior causal theories. For example, some person's causal theories could explain the high crime rate in certain areas due to the

**It is important to note that the generalization algorithm can operate on a single event. In this case, the "similar" features are simply all the features, and "always precedes" is interpreted as "precedes".

racial make-up of the area. Other's causal theories will place the blame on the high unemployment in the area.

We distinguish another kind of generalization: a tentative generalization. A tentative generalization is one whose causal relationship is proposed by generalization rules but not confirmed by prior causal theories.*** In a tentative generalization, the relationships postulated by the generalization rules are assumed to hold until they are contradicted by later examples. The verification of a tentative generalization occurs after Step 1.

1.5 If the most specific generalization is tentative, compare the new event to the prediction made by the generalization.

The primary difference between a tentative generalization and an explanatory generalization is how they are treated when a new event contradicts the generalization. In this case, a tentative generalization will be abandoned. However, if an explanatory generalization is contradicted an attempt will be made to explain why the new event differs from previous events. For example, OCCAM constructs a explanatory generalization which states that in kidnapping the kidnapper releases the victim to keep his end of the trade demanded in the ransom note. After this generalization is made, it is presented with an kidnapping example in which the victim is murdered. Rather than abandoning the previous generalization it finds an explanation for murdering the victim: to prevent the victim from testifying against the kidnapper. In contrast, consider what happens if a tentative generalization were built by correlational means describing the release of the hostage: if a hostage were killed in a later kidnapping, the tentative generalization would be contradicted and abandoned. The perseverance of explained generalizations is supported by a study by Anderson et al. [2] in which subjects who were requested to explain a relationship showed a greater degree of perseverance after additional information than those who were not requested.

Later in this paper, we will describe the mechanisms used by OCCAM to confirm tentative generalizations. An example of OCCAM learning with and without prior causal theories should help to clarify how generalization rules and causal theories interact to create explanatory and tentative generalizations.

Inflating Balloons

In this example, the initial memory is essentially empty. The examples are input as conceptual dependency representations [29] of the events taking place in Figure 1.

First L. successfully blowing up a red balloon is added to memory. Next, the event which describes L. unsuccessfully blowing up a green balloon is added to memory and similarities are noticed. DIFFERENT-FEATURES is an applicable generalization rule:

DIFFERENT-FEATURES

If two actions have different results, and they are performed on similar objects with some different features, assume the differing features enable the action to produce the result.

This generalization rule produces a question for the explanation process: "Does the state of a balloon being red enable the balloon to be inflated when L. blows air into it?". This cannot be confirmed, but it is saved as a tentative generalization. Associated with this generalization is the explanation which describes the difference in results as enabled by the color of the balloon.

A green balloon successfully blown up by L. after M. deflated it is added to memory. It contradicts the previously created tentative generalization, which is removed from the memory. Next, a generalization rule is applied:

PREVIOUS-ACTION

If ACTION-1 always precedes ACTION-2 which results in STATE-2, assume ACTION-1 results in STATE-1 which enables ACTION-2 to produce STATE-2.

In this case, ACTION-1 is M. deflating the balloon, ACTION-2 is L. blowing into the balloon and STATE-2 is that the balloon is inflated. The confirmation process attempts to verify that deflating a balloon results in a state that enables L. to inflate the balloon. However, this fails and a new tentative generalized event is created which saves the postulated explanation. Note that if there existed a proper causal theory, an explanatory generalization could be created which would save the information that STATE-1 is that the balloon is stretched.

***OCCAM does not always create a tentative generalization if it cannot create an explanatory one. See [27] for the details.

Mike is blowing up a red balloon.
Lynn: "Let me blow it up."
Mike lets the air out of the balloon and hands it to Lynn.
Lynn blows up the red balloon.

Lynn picks up a green balloon and tries to inflate it.
Lynn cannot inflate the green balloon.
Lynn puts down the green balloon and looks around.
Lynn: "How come they only gave us one red one?"
Mike: "Why do you want a red one?"
Lynn: "I can blow up the red ones."

Mike picks up a green balloon and inflates it.
Mike lets the air out of the green balloon; hands it to Lynn.
Mike: "Try this one."
Lynn blows up the green balloon.
Lynn gives Mike an uninflated blue balloon.
Lynn: "Here, let's do this one."

Figure 1: Protocol of Lynn (age 4) trying to blow up balloons.

Economic Sanctions

OCCAM is provided with a large amount of background knowledge in its newest domain of economic sanctions. In this section, we illustrate how a specific generalization rule is useful both in the previous example as well as in explaining the effects of economic sanctions. Due to space limitations we must ignore the memory issues. We assume the initial memory contains the following example summarized from [13]:

```
In 1980, the US refused to sell grain
     to the USSR unless the USSR
     withdrew troops from Afghanistan.
The USSR paid a higher price to buy grain
     from Argentina.
```

When the following event is added to memory, the generalization process is initiated when a similarity is noticed between the new event and a previous event:

```
In 1983, Australia refused to sell uranium
     to France, unless France ceased
     nuclear testing in the South Pacific
France paid a higher price to buy uranium
     from South Africa.
```

Here, PREVIOUS-ACTION suggests an explanation for the similarities. In this case, ACTION-1 is identified as the US or Australia refusing to sell a product, ACTION-2 is identified as USSR or France buying the product from another country, and STATE-2 is the USSR or France possessing the product. PREVIOUS-ACTION postulates that ACTION-1 (refusing to sell the product) resulted in STATE-1 which enabled ACTION-2 (purchasing the product for more money from a different country) to result in STATE-2 (possessing the product). OCCAM's causal theories in the economic domain identify STATE-1 as the willingness to pay more money for the product. Therefore, OCCAM constructs the explanatory generalization in Figure 2 from these two examples.

In this generalization country-1 is generalized from the US and Australia. A purely correlational approach could find a number of features in common between these two countries (e.g., both have a native language of English, both have a democratic government etc.). However, the explanation-based approach finds relevant only that both countries are suppliers of a product-1. Similarly, country-3 is generalized from Argentina and South Africa but only two of their common features are relevant: that they supply product-1 and that they have a business relationship with country-2.

```
coerce
 ACTOR country-1
 VICTIM country-2
 DEMAND goal-1
 THREAT sell
         ACTOR  (country-1
                   A-SUPPLIER-OF product-1)
         TO     country-2
         OBJECT product-1
         AMOUNT amount-1
         MODE   neg
 RESULT sell
         ACTOR (country-3
                  A-SUPPLIER-OF product-1
                  BUSINESS-REL country-2)
         TO     country-2
         OBJECT product-1
         AMOUNT amount-2
 RESULT goal-failure GOAL  goal-1
                     ACTOR country-1
```

Figure 2: Explanatory generalization created by OCCAM to explain one possible outcome of economic sanctions. This generalization indicates that country-1 refusing to sell a product to country-2 to achieve goal-1 will fail to achieve this goal if there is a country-3 which supplies the product and has a business relationship with country-2.

Confirming Tentative Generalizations

In many respects, tentative generalizations are treated in the same manner as explanatory generalizations. Both can be used to predict or explain the outcomes of other events. For example, after several examples of parents helping their children and some strangers not assisting children OCCAM builds a tentative generalization which describes the fact that parents have the goal of preserving the health of their children. This tentative generalization is used as a causal theory to explain why a parent pays the ransom in a kidnapping.

However, as stated earlier tentative generalizations are treated differently when new evidence contradicts the generalizations. When a tentative generalization is confirmed, it becomes an explanatory generalization. There are a number of strategies which are useful in confirming a tentative generalization.

- **Increase confidence with new examples.** When new examples conform to the prediction made by a generalization, the confidence in the generalization is increased. When the confidence exceeds a threshold, the generalization is confirmed. This strategy was the mechanism utilized by IPP [18].
- **Increase confidence when a tentative generalization is used as an explanation for another generalization.** When the explanation stored with a tentative generalization is confirms a postulated causal relationship the confidence in the tentative generalization is increased. For example, when OCCAM uses the tentative generalization that parents have a goal of preserving the health of their children to explain why the parent pays the ransom in kidnapping, the confidence in the tentative generalization is increased.
- **Search for competing hypotheses.** If no competing hypothesis can be found to explain the regularities in the data, the confidence of the generalization is increased. In OCCAM searching for competing hypotheses consists of trying other generalization rules. If no other generalization rules are applicable, the confidence in the tentative generalization is increased.

The above strategies all increase the confidence in a tentative generalization. We have experimented with different values for the increment of confidence and the threshold. More research needs to be done in this area to determine reasonable values for these parameters. There is some evidence [26] that these parameters are not constants but a function of the vividness of the new information. The following strategies can confirm a tentative generalization with just one additional example:

- **Specify intermediate states or goals.** Typically, a tentative generalization has intermediate states or goals which are not identified. For example, the tentative generalization describing which balloons can be inflated by L. contains an intermediate state which results from deflating the balloon which enables L. to inflate the balloon. If this intermediate state were identified in future examples, the tentative generalization could be confirmed. Similarly, cold weather was identified as a tentative cause for the Space Shuttle accident. Identifying a particular component whose performance is affected by cold weather and whose failure would account for the accident would confirm the cause.

Finally, we have identified but not yet implemented another strategy to confirm tentative generalizations:

- **Ask an authority.** Many children are constantly asking for explanations. There are two types of questions asked: verification (e.g., "Does X cause Y?") which corresponds to confirming a tentative generalization in OCCAM and generation (e.g., "Why X?") which corresponds to generating and confirming an explanation.

Future Directions

Currently, OCCAM is a passive learning program which learns as it adds new observations to its memory. We are in the process of making OCCAM play a more active role in the learning process. There are number of ways that OCCAM can initiative:

- **Ask Questions.** As discussed previously, a tentative generalization may be confirmed by asking a authority (e.g., a parent). However, the explanation provided by the authority may report the cause of an event without illustrating the justification used by the authority to attribute causality. Recall that in explanation-based learning the preconditions of the inference rules used to deduce causal relationships determine what features are relevant (i.e., should be included in a generalization). When the explanation is provided by another person it may not include these preconditions. We intend to make use of similarities and differences between examples to induce these preconditions.
- **Suggest Experiments.** Recall that one mechanism to confirm a tentative generalization is to search for other explanations. If there are two (or more) possible explanations, they may make different predictions. We plan to extend OCCAM to suggest an experiment which would distinguish between the competing explanations [8].

Conclusion

OCCAM is a program which integrates two sources of information to build generalizations describing the causes or motivations of events. The design of OCCAM was influenced by a number of studies which indicate that prior causal theories are more influential than correlational information in attributing causality. The combination of explanation-based and correlational learning techniques used by OCCAM improves on previous learning programs in the follow manners:

- Purely correlational learning programs such as IPP [18] require a large number of examples to determine which similarities among the examples are relevant and which are coincidental. In contrast, OCCAM builds explanatory generalizations which describe novel interactions among its causal theories. These causal theories indicate what features are relevant.
- Explanation-based learning programs such as GENESIS [25] must have a complete causal theory to generalize. In contrast, OCCAM's generalization rules enable the learning of causal theories. In addition, these generalization roles serve to focus the explanation process.
- UNIMEM [20] integrates correlational and explanation-based learning by using a strategy which prefers correlational information to prior causal theories: explanation-based learning to rules out coincidental similarities in correlational generalizations. Empirical evidence indicates that in people, the causal theories are preferred to correlational information. One reason for this bias is that correlating features over a number of examples may exceed the limitations of a person's memory. Due to the limitations of computer memory, UNIMEM does not perform correlation for combinations of features. Therefore, it cannot learn that a conjunction of features results in an outcome. In contrast, OCCAM's learning strategy naturally discovers when a conjunction of features results in an outcome. This occurs

when it forms an explanatory generalization by recording the interaction between two or more inference rules. If the preconditions of these inference rules rely on different features of the same entity, then the conjunction of these features is relevant.

In the generalization theory implemented in OCCAM, prior causal theories are used to infer causality. Generalizations are built to record novel chains of inference. Correlational information has a role similiar to temporal information: to suggest or confirm causal theories.

References

[1] Anderson, J.R. Knowledge Compilation: The General Learning Mechanism. In *Proceedings of the International Machine Learning Workshop*. Monticello, Illinois, 1983.

[2] Anderson, C.A., Lepper, M.R., & Ross, L. The perseverance of social theories: The role of explanation in the persistance of discredited information. *Journal of Personality and Social Psychology* 39:1037-1049, 1980.

[3] Ausubel, D.M. and Schiff, H. M. The Effect of Incidental and Experimentally Induced Experience on the Learning of Relevant and Irrelevant Causal Relationships by Children. *Journal of Genetic Psychology* 84:109-123, 1954.

[4] Bruner, J.S., Goodnow, J.J., & Austin, G.A. *A Study of Thinking*. Wiley, New York, 1956.

[5] Bullock, Merry. *Aspects of the Young Child's Theory of Causality*. PhD thesis, University of Pennsylvania, 1979.

[6] Chapman, L.J., & Chapman, J.P. Genesis of Popular but Erroneous Diagnostic Observations. *Journal of Abnormal Psychology* 72:193-204, 1967.

[7] Chapman, L.J., & Chapman, J.P. Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs. *Journal of Abnormal Psychology* 74:271-280, 1969.

[8] Cohen, Paul. *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Appraoch*. Pitman Publishing Inc., Marshfield, Mass., 1985.

[9] DeJong, G. Acquiring Schemata Through Understanding and Generalizing Plans. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. Karlsruhe, West Germany, 1983.

[10] Dyer, M. *In Depth Understanding*. MIT Press, 1983.

[11] Garcia, J., McGowan, B., Ervin, F.R., & Koelling, R.A. Cues: Their relative effectiveness as reinforcers. *Science* 160:794-795, 1968.

[12] Granger, R. & Schlimmer, J. Combining Numeric and Symbolic Learning Techniques. In *Proceedings of the Third International Machine Learning Workshop*. Skytop, PA, 1985.

[13] Hufbauer, G.C., & Schott, J.J. *Economic Sanctions Reconsiderd: History and Current Policy*. Institute For International Economics, Washington, D.C., 1985.

[14] Kelley, Harold H. Causal Schemata and the Attribution Process. In Jones, Edward E., Kanouse, David E., Kelley, Harold H., Nisbett, Richard E., Valins, Staurt & Weiner, Bernard (editor), *Attribution: Perceiving the Causes of Behavior*, pages 151-174. General Learning Press, Morristown, NJ, 1971.

[15] Kelley, Harold H. The Process of Causal Attribution. *America Psychologist* :107-128, February, 1983.

[16] Kolodner, J. *Retrieval and Organizational Strategies in Conceptual Memory: A Computer Model*. Lawrence Erlbaum Associates, Hillsdale, NJ., 1984.

[17] Laird, J., Rosenbloom, P., and Newell, A. Towards Chunking as a General Learning Mechanism. In *Proceedings of the National Conference on Artificial Intelligence*. American Association for Artificial Intelligence, Austin, Texas, 1984.

[18] Lebowitz, M. *Generalization and Memory in an Integrated Understanding System*. Computer Science Research Report 186, Yale University, 1980.

[19] Lebowitz, M. Correcting Erroneous Generalizations. *Cognition and Brain Theory* 5(4), 1982.

[20] Lebowitz, M. Integrated Learning: Controlling Explanation. *Cognitive Science* , 1986.

[21] Michotte, A. *The Perception of Causality*. Basic Books, Inc., New York, 1963.

[22] Minton, S. Constraint-based Generalization: Learning Game-Playing Plans from Single Examples. In *Proceedings of the National Conference on Artificial Intelligence*. Austin, TX, 1984.

[23] Mitchell, T. Generalization as Search. *Artificial Intelligence* 18(2), 1982.

[24] Mitchell, T., Kedar-Cabelli, S. & Keller, R. *A Unifying Framework for Explanation-based Learning*. Technical Report, Rutgers University, 1985.

[25] Mooney, R. & DeJong, G. Learning Schemata for Natural Language Processing. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*. Los Angeles, CA, 1985.

[26] Nisbett, Richard & Ross, Lee. *Human Inference: Strategies and Shortcomings of Social Judgements*. Prentiss-Hall, Inc., Engelwood Cliffs, NJ, 1978.

[27] Pazzani, Michael. Explanation and Generalization-based Memory. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*. Irvine, CA, 1985.

[28] Pazzani, M. Refining the Knowledge Base of a Diagnostic Expert System: An Application of Failure-Driven Learning. In *Proceedings of the National Conference on Artificial Intelligence*. American Association for Artificial Intelligence, 1986.

[29] Schank, R.C. & Abelson, R.P. *Scripts, Plans, Goals, and Understanding*. Lawrence Erlbaum Associates, Hillsdale, NJ., 1977.

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

[31] Soloway E. *Learning = Interpretation + Generalization: A Case Study in Knowledge-directed Learning*. PhD thesis, University of Massachusetts at Amherst, 1978.

[32] Vere, S. Induction of Concepts in the Predicate Calculus. In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*. Tbilisi, USSR, 1975.