Inducing Causal and Social Theories: A Prerequisite for Explanation-based Learning

Michael J.Pazzani (pazzani@CS.UCLA.EDU)

Artificial Intelligence Laboratory, 3531 Boelter Hall
University of California, Los Angeles, CA 90024, USA

Abstract

We present an approach to learning to predict and explain the outcome of events which lies between similarity-based methods and explanation-based methods. In the approach to learning presented in this paper and implemented in a computer program called OCCAM, a theory of causality constrains the search for causal hypotheses. We present evidence that people possess a theory of causality which facilitates learning causal relationships when certain spatial and temporal relationships exist between a cause and an effect.

1. Introduction

We address the problem of learning causal and social knowledge by observing examples of events and their consequences. We wish to consider the acquisition of simple causal or social theories such as those which 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).

In explanation-based learning (DeJong, 1983, Mooney & DeJong, 1985, Mitchell, Kedar-Cabelli & Keller, 1986, DeJong, 1986), causal and social theories serve as background knowledge which explain the outcome of an example event. The example event is generalized by retaining only those features of the example which were necessary to produce the explanation. An important question is not addressed in explanation-based learning: how is this background knowledge acquired? In this paper, we address this issue in the context of a learning program called OCCAM (Pazzani, 1985, Pazzani, 1986) which is unique among explanation-based learning systems in that it also learns background causal and social knowledge by empirical techniques.

In this paper, we look at two sources of information which guide the search for a causal hypothesis:

- Inter-example relationships: Regularities among a number of examples reveal the conditions under which a cause produces an effect.
- Intra-example relationships: Temporal and spatial relationships between a cause and an effect which constrain the search for a causal hypothesis.

We concentrate on intra-example relationships in this paper for a number of reasons.

1. Intra-example relationships can facilitate learning by constraining the set of possible causes. For example, consider the following event sequence: First, two events occur at approximately the same time: a taxi is seen driving on the street outside the kitchen window and a brown bird flies into the window pane. Next, the window pane shatters. By ruling out the taxi as a potential cause for the window breaking, the search space for the problem of determining what causes windows to break can be reduced. Of course, inter-example knowledge is also needed to determine that the color of the object which strikes the window is not important, but the weight and velocity of the object are. However, resources such as memory can be utilized more effectively if they are not also required to correlate the color and velocities of cars passing by.

An analogy can be drawn between OCCAM's use of intra-example relationships and STAHL's (Langley et. al., 1986) heuristics which constrain the search for the components of a compound. For example, STAHL contains a heuristic which states that if a substance occurs in both sides of a chemical reaction, it does not enter into the reaction. It is conceivable that STAHL could still determine the components of compounds if it did not use this heuristic. However, the search space would be larger and more examples would be required to rule out alternatives. Similarly, in OCCAM, intra-example relationships reduce the search for potential causes.

2. There are a number of findings in developmental and social psychology which relate to this problem (e.g. (Kelley, 1983, Bullock, 1979, Michotte, 1963, Piaget, 1930, Shultz, 1986)). In addition, in conjunction with Professor Mort Friedman of the Psychology Department at UCLA, I am conducting a series of experiments investigating how prior knowledge or intra-example relationships facilitate learning.

3. The discovery of inter-example relationships is a fairly well understood problem in machine learning (Mitchell, 1982, Michalski, 1977, Holland et al., 1986). In contrast, intra-example relationships have largely been overlooked in systems which learn to predict the outcome of events (Lebowitz, 1980, Lebowitz, 1986a, Salzberg, 1985)[1].

In the remainder of this paper, we first discuss a number of studies which indicate what intra-example relationships people employ when attributing causality. Finally, we indicate how these intra-example relationships constrain the search for a causal and social hypotheses in OCCAM.

2. Constraints on Causal Relationships

There have been a number of studies investigating what relationships between an effect and a potential cause are required to attribute causality. These relationships include:

- Temporal order: Children as young as four require a potential cause to precede an effect (Shultz and Mendelson, 1975). Although this may seem like a trivial constraint, existing machine learning systems (Salzberg, 1985, Lebowitz, 1986b) which predict the outcome of events don't make use of temporal information.
- Temporal contiguity: In one experiment (Michotte, 1963), 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. 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 induce a causal relationship. From this experiment, it is clear that correlation alone is not enough to attribute causality.
- Mechanism: An important constraint on causal relationships is the existence of a mechanism which transmits a causal "force" to the effect. For example, when a ball rolled down a runway and hit a jack-in-the-box, and the jack popped up, both adults and children indicate that the ball caused the jack to pop up. However, if the ball is stopped before it reaches the jack-in-the-box and the jack pops up, neither adults nor preschool children attribute the ball as a cause (Bullock, 1979).

[1] Salzberg's HANDICAPPER program does contain some heuristics for controlling the size of the search space. However, these heuristics focus mainly on selecting features of a potential cause independent of the effect (e.g., select an unusual feature or the feature you know the least about).

Mechanism appears to be the dominant constraint on causal relationships. When presented with potential causes which violate the other intra-example relationships, subjects prefer selecting a cause which obeys the mechanism constraint (Shultz, 1982). An understanding of the mechanism is also important in identifying whether a new situation which is slightly different will produce the same effect. For example, in one experiment (Bullock, Gelman & Baillargeon, 1982), children were shown an apparatus which consisted of a box containing a row of blocks. When a long orange rod was inserted in one end, the blocks fell down and pushed a rabbit out of the box. Next, they were shown different situations, such as a different color rod, or a short rod which could not reach the blocks and asked to predict the outcome. Children as young as three were able to distinguish a modification which interfered with the mechanism from one which did not.

I am collaborating with Professor Mort Friedman of the Psychology Department at UCLA, on a series of experiments investigating how knowledge of a mechanism facilitates learning. In the first experiment, we investigated how existing causal knowledge affects the number of trials before learning to make accurate predictions. In the second experiment, we investigated how general knowledge about causality affects the number of trials before learning to make accurate predictions.

2.1. Experiment 1- Existing causal knowledge

The subjects in this experiment were 120 undergraduates fulfilling a requirement for an introductory psychology course. Each subject was shown a number of cards in random order which contained a photograph of a child performing an action on a balloon (either stretching or measuring the balloon). In addition, on different cards the balloons varied in shape (long or round) and color (blue or yellow). Subjects were divided into two conditions:

- Inflate: those who had to predict whether the child would be able to inflate the balloon.
- Alpha: those who had to predict whether the card belonged to an artificial category called alpha.

Subjects were presented with a card, asked to predict, and then informed of the correct answer. Trials continued until the subject was able to predict correctly on every card. We recorded the number of the last trial on which the subject made an error. In the "Inflate" condition, we predicted that subjects would be able to use their existing causal knowledge about balloons (i.e., stretching a balloon makes it easier to inflate the balloon). We predicted that existing causal knowledge would facilitate learning to make the correct prediction when the data was consistent with existing knowledge and hinder learning when the data was not consistent with existing knowledge. In the "Alpha" condition, knowledge of what makes a balloon easier to inflate should neither facilitate or hinder classifying the cards.

Subjects in each condition were subdivided into groups who had to predict based upon the action performed on the balloon. (In the "Inflate" condition, some subjects saw examples that indicated the child could only inflate a balloon if she stretched it; others saw examples that indicated the child could only inflate a balloon if she measured it. In the "Alpha" condition, some subjects saw examples that indicated that alpha is a child stretching a balloon; others saw examples that indicated that alpha is a child measuring a balloon.)

There are two major findings of this experiment (The results are significant at the .05 level $F(9,110)= 3.41$):

- Subjects required less trials to learn to predict that a balloon which had been stretched could be inflated (2.1 trials) than to predict that a balloon which had been measured could be inflated (6.1 trials). This finding indicates that knowledge of an existing causal relationship facilitates learning. Note that there are a small number of hypotheses consistent with the existing causal knowledge (i.e., the child can inflate all balloons, the child can inflate no balloons, and the child can inflate stretched balloons.) Subjects required a small number of examples to determine which hypothesis is correct. On the other hand, if the correct answer is inconsistent with prior causal knowledge, many more hypotheses are possible (e.g., the child can inflate only blue balloons, the child can only inflate measured balloons, etc.). In this situation, more examples are required before finding the correct hypothesis.

- Subjects required approximately the same number of trials to determine that a balloon being stretched is an alpha (3.9 trials) or to determine that a balloon being measured was an alpha (3.0 trials). Subjects in the "Alpha" condition are presented with the same data as the "Inflate" condition. Since existing knowledge cannot help in the "Alpha" condition, there is no significant difference between the group that learned that stretching is an alpha and the group that learned that measuring is an alpha. The alpha group serves as a control group. Differences in the "Inflate" condition cannot be explained by factors such as greater perceptual salience of stretching as opposed to measuring. Otherwise, these same differences would appear in the "Alpha" condition.

In this experiment, we have demonstrated that the process of learning to predict outcomes is not simply comparing and contrasting examples. If this were true, then the results in the "Inflate" condition would not differ from the "Alpha" condition. Instead, existing causal knowledge facilitates learning so that fewer examples are needed to arrive at the correct hypothesis.

2.2. Experiment 2- General theories of causality

The second experiment is similar in design to the first. The first experiment addressed the issue of how existing causal knowledge facilitates learning. In the second experiment, we investigated a general relationship between the cause and the effect which facilitate learning.

The subjects in this experiment were 80 undergraduates fulfilling a requirement for an introductory psychology course. Each subject was shown a number of videotapes of a child picking up a balloon and then doing something (dipping the balloon in a glass of water, putting a necklace on or snapping her fingers). In addition, on different tapes the balloons varied in size (small or large) and color (orange or yellow). Subjects were divided into "Inflate" and "Alpha" conditions. Subjects were shown a tape, asked to predict whether the child could inflate the balloon (or whether the tape was an alpha), and then informed of the correct answer. Trials continued until the subject was able to predict correctly on 6 tapes in a row. We recorded the number of the last trial on which the subject made an error. Subjects in each condition were subdivided into groups who had to predict based upon the action performed (either snapping her fingers or dipping the balloon in water).

We predicted that subjects in the "Inflate" condition would find it easier to learn that the child could only inflate a balloon which had been dipped in water than to learn that the child could only inflate a balloon after she snapped her fingers. The difference here is not any specific prior causal knowledge.[2] Instead, knowledge about causal relationships in general is applicable. The sequence *dipping a balloon in water followed by blowing air into the balloon, followed by the balloon*

[2] In fact, we asked 60 subjects in the "Alpha" condition of Experiment 1 if they felt that dipping a balloon in water would make it harder or easier to inflate the balloon. 41 subjects replied it wouldn't matter, 18 weren't sure, and 1 thought it would be harder because the balloon would be slippery.

changing fits a pattern for a causal relationship: An action on an object (dipping the balloon in water) results in a state change for the object which enables a subsequent action (blowing into the balloon) to produce a state change. In contrast, snapping fingers before blowing into the balloon does not fit this general pattern. An important constraint is violated. For an action to result in a state change for an object, the action has to operate on the object.

Of course, in the "Alpha" condition, general knowledge of causality should not facilitate or hinder learning. Therefore, we anticipated that it would take the same number of trails to identify the alpha tapes whether the child was snapping her fingers or dipping a balloon in water.

The results of this experiment confirmed our predictions. (The results are significant at the .05 level $F(3,76) = 8.88$):

- Subjects required less trials to learn to predict that a balloon which had been dipped in water could be inflated (3.5 trials) than to predict that a balloon could be inflated after the child snapped her fingers (7.6 trials). This finding indicates that general knowledge of existing causal relationships facilitates learning.
- Subjects required approximately the same number of trials to determine that a balloon being dipped in water is an alpha (5.7 trials) and to determine that the child snapping her fingers is an alpha (5.9 trials).

In the second experiment, we have demonstrated that general theories of causality can facilitate learning to predict the outcome of an event. In particular, when there is a relationship between a cause and an effect which suggest a causal mechanism (e.g., a set of intermediate states), then fewer examples are required to identify a causal relationship. The results of this experiment indicate that the process of learning to predict the outcome of an event is not simply empirically associating two events which have occurred in succession.[3]

3. OCCAM

OCCAM is a program which learns to predict and explain the outcome of events. In this paper, we concentrate on learning causal and social theories in the absence of existing knowledge. OCCAM contains a number of *generalization rules* which postulate explanations for similarities and differences between events. The simplest generalization rule (*If an action on an object always precedes a state change for the object, then the action results in the state change*) is displayed in Figure 1.

```
(def-gen-rule
  ?state-1 = (state type ?stype                    ;potential effect
                    object ?object)
  after                                            ;temporal relation
  ?act-1 = (act type ?atype                        ;potential cause
                object ?object)
  ((?act-1 result ?state-1))                       ;causal relation
  )
```

Figure 1: An exceptionless generalization rule (variables are preceded by "?"): If an action on an object always precedes a state change for the object, then the action results in the state change.

[3]It is interesting to note that the subjects found it easier to associate dipping a balloon in water with inflation even though this may not be consistent with their specific world knowledge. We would like to run a similar experiment with a physical cause unfamilar to most college freshman, or run the same experiment with preschool children.

Although the rule in Figure 1 appears to be very simple, it encodes many assumptions about causal relationships which drastically reduce the search space. First, it encodes the intra-example constraint that the action must operate on the object whose state has changed. This would rule out a taxi driving past a window as a potential cause for the window breaking. Secondly, it indicates that the only important features of the action are the type of action (e.g., an application of force) and the object. The actor who performs the action, the time the action is performed and any instrument with which the action is performed are not relevant. For example, if a ball was kicked into a window pane and the window shattered, OCCAM would not consider the actor (or the actor's hair color) to be features of the potential cause. Finally, the generalization rule in Figure 1 also contains a temporal constraint. Therefore, if a ladybug flies through the broken window after a ball has hit it, OCCAM would not produce the explanation that the window broke because [sic] the ladybug will fly through it.

```
(def-gen-rule
  ?state-1 = (state type ?stype                    ;potential effect
                    object ?object)
  after                                            ;temporal relation
  ?act-1 = (act type ?atype                        ;potential cause
                object ?object)
  ((?act-1 result ?state-1))                       ;causal relation
  (:difference ?act-1 actor)                       ;exceptions
)
```

Figure 2: A dispositional generalization rule: If two similar actions performed on an object have different results, and they are performed by different actors, the differing features of the actor are responsible for the different result.

The generalization rule in Figure 1 is called an exceptionless generalization rule because it applies when there are only positive examples. Other generalization rules focus on reasons that similar actions have different results. For example, in addition to the above generalization rule, OCCAM contains a rule which blames the actor for different results of actions whose actors differ (see Figure 2). This type of generalization rule is called a dispositional generalization rule because it attributes a different result to differing properties (or disposition) of actors or objects. This rule would focus the search for an explanation on the different features of the actor. For example, there are a number of actions which adults can successfully perform but children cannot (e.g., a heavy object might move if an adult pushes it, but not a child). Of course, without prior knowledge, OCCAM must correlate features of the actor with outcomes over a number of examples to discover that age rather than hair color is relevant.[4] Generalization rules in OCCAM are ordered by simplicity.[5] The rule in Figure 2 would only apply if the rule in Figure 1 was not able to make accurate predictions without considering differences in the actor.

The final type of generalization rule in OCCAM is called a historical generalization rule because it attributes different results of similar actions to different histories of the objects involved. For example, one rule (displayed in Figure 3) attributes the difference in a result to the existence of a state of an object which enables the result. This rule would reveal relationships such as removing the top from a bottle results in a state which enables the contents to come out if the bottle is overturned or that stretching a balloon results in a state which enables blowing air into the balloon to inflate the balloon. Recall that simpler generalization rules are tried before more complex ones. This rule

[4]Distinctions which have proven useful are recorded (indexed by the generalization rule and type of action) and influence the order of generating future hypotheses (Pazzani, 1987).

[5]This is why we call the system OCCAM. The simplest generalization rule produces the simplest hypothesis.

would only apply if the effect is not adequately explained by simpler rules such as the ones in Figure 1 or Figure 2.

```
(def-gen-rule
 ?state-2 = (state type ?stype                              ;potential effect
                   object ?object)
  before                                                    ;temporal relation
  ?act-2 =  (act type ?atype-2                              ;potential cause
                 object ?object)
  ((?act-2 result ?state-2)                                 ;causal mechanism
   (?act-1 result ?state-1 = (state object ?object))
   (?state-1 enables ?act-2))
  (:link ?act-2 before ?act-1 = (act type ?atype-1          ;constraint
                                     object ?object))
)
```

Figure 3: A historical generalization rule: If an initial action (?act-1) on an object is always present when a subsequent action (?act-2) precedes a state change (?state-2) for the object, then ?act-1 results in a state (?state-1) which enables ?act-2 to result in the state change (?state-2).

The experiment discussed earlier when subjects found it easier to learn that a child could inflate a balloon after she had dipped it in water than to learn that the child could inflate a balloon after she snapped her fingers was to demonstrate the necessity of the rule in Figure 3. In particular, the `object` of `?act-1` is the same as the `object` of `?act-2` when the balloon is dipped in water before blowing air into it. Therefore, the generalization rule in Figure 3 would apply and suggest that dipping the balloon in water results in some (unspecified) state which enables blowing into the balloon to make the balloon larger. In contrast, when the child could only inflate the balloon after she snaps her fingers the generalization rule in figure 3 does not apply because the `object` of the two actions (snapping fingers and blowing air into the balloon) differ. In fact, OCCAM contains no generalization rules which match this situation and is forced to rely on empirical methods alone by comparing and contrasting features of positive and negative examples.

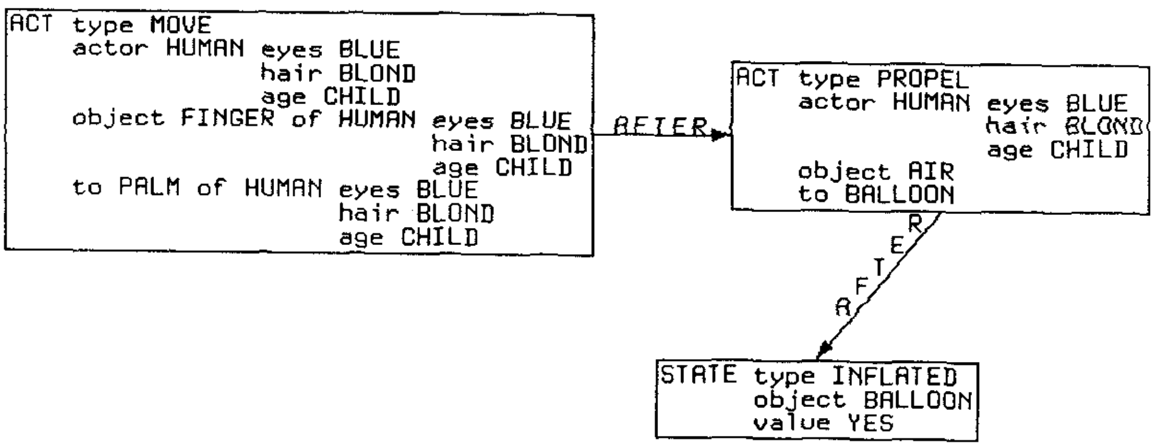

Figure 4: A generalization created by empirical methods describing the situation when a balloon is inflated only after the child snaps her fingers.

The rule in Figure 3 also illustrates a primary difference between OCCAM and UNIMEM or HANDICAPPER. In addition to predicting the outcome of an action, OCCAM also constructs a causal mechanism which accounts for how the action brings about outcome. Therefore, it should be able to predict the outcome of similar events by determining if they interfere with the postulated mechanism more accurately than approaches which rely entirely on similarities and differences between prior examples. Figure 4 illustrates the generalization that OCCAM constructs to describe the situation when a balloon is inflated after a child snaps her fingers. It is constructed entirely by empirical methods and contains temporal links but no causal links. In contrast, Figure 5 illustrates the generalization that OCCAM constructs to describe the situation when the child can inflate the balloon only after it has been dipped in water. A generalization rule similar to the one in Figure 3 suggests a causal mechanism which results in the balloon being inflated.

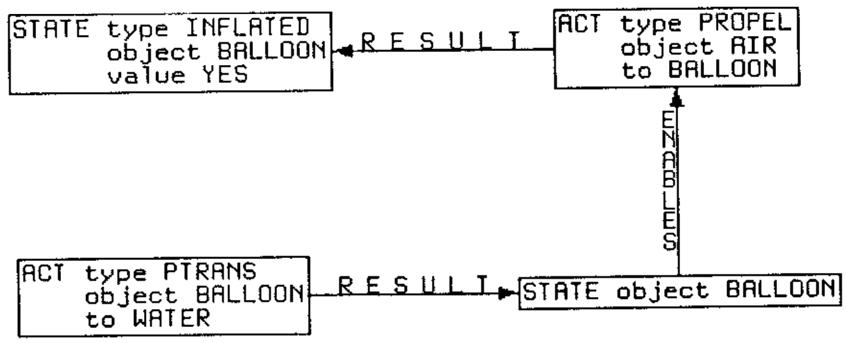

Figure 5: A generalization containing a causal mechanism is created when a situation matches a known causal pattern. This generalization describes the situation when the child can inflate the balloon only after it has been dipped in water.

3.1. An example

In OCCAM, learning occurs when new events are added to memory. The memory in OCCAM is organized in a manner similar to IPP (Lebowitz, 1980) and CYRUS (Kolodner, 1984). The major difference between OCCAM and these programs is the approach to learning. IPP and CYRUS utilized similarity-based learning methods. In OCCAM, the memory update process consists of:

1. If an existing schema accounts for the outcome of an event, then index the event in memory by features which elaborate on the existing schema.[6]
2. Otherwise, if existing schemata can be combined to explain the outcome of an event, construct a new schema using explanation-based learning.
3. Otherwise, if there is a cue for a known causal pattern (i.e., a generalization rule which matches the new example), construct a schema by theory-driven learning. The variables in the generalization rule are instantiated to their values in the example.[7]
4. Otherwise, if there are several similar examples (currently 3), then a schema is created

[6]An existing schema may also predict an opposite outcome for an event. If this is the case, the existing schema is weakened or abandoned if there is little support for the schema (Lebowitz, 1986b). Support for a schema is increased each time a new example is explained by the schema. However, instead of abandoning a schema when a single counter-example is encountered, OCCAM keeps track of the exceptions. If the ratio of explained examples to exceptions is less then a threshold (currently .9), then the erroneous schema is deleted.

[7]One might question why OCCAM prefers explanation-based learning over theory-driven learning. We are using the general strategy of preferring the most specific knowledge. The utility of this strategy is best illustrated when a generalization rule would suggest a causal mechanism which is not consistent with specific knowledge of mechanisms. For example, if someone died of cancer several seconds after the nuclear accident at Three Mile Island, specific knowledge of the time scale of death by cancer should rule out the accident as the cause of the cancer.

by similarity-based methods[8].

5. Otherwise, the new event is simply added to memory, indexed by the features which differ from the most specific schemata.

The example that we consider in this section only makes use of the third step in this process. The input to OCCAM is the Conceptual Dependency (CD) representation (Schank, 1977) of the situations described in Figure 6. The data correspond to the experiment in which the child could only inflate a balloon after it has been dipped in water. The representation of the first example is illustrated in Figure 7.

1. **The child dips a small yellow balloon in water, blows air into the balloon and the balloon is inflated.**
2. **The child snaps her fingers, blows air into a large yellow balloon and the balloon is not inflated.**
3. **The child puts on a necklace, blows air into a large orange balloon and the balloon is not inflated.**
4. **The child dips a large yellow balloon in water, blows air into the balloon and the balloon is inflated.**

Figure 6: Input to OCCAM describing the situation when the child can only inflate a balloon after it has been dipped in water.

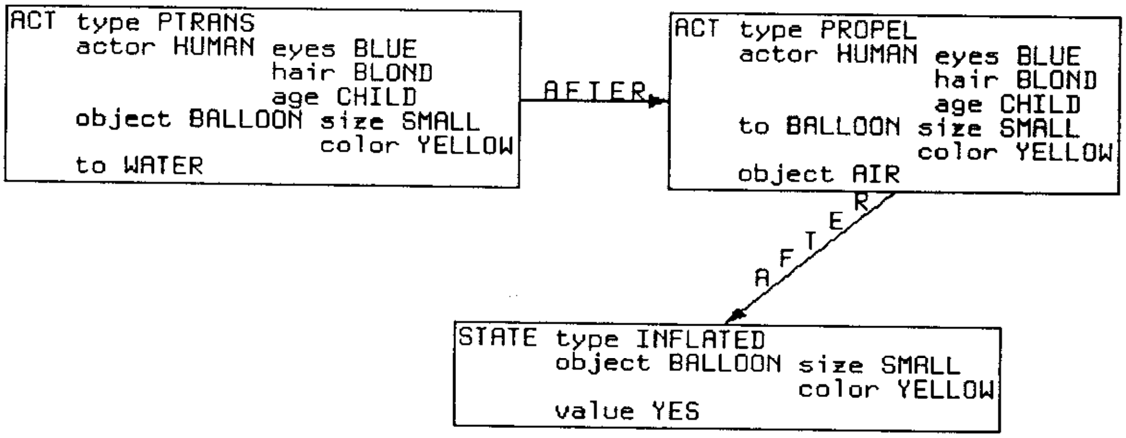

Figure 7: CD representation for "the child dips a small yellow balloon in water, blows air into the balloon and the balloon is inflated."

[8] One might wonder why OCCAM (or people) need similarity-based methods. There are many examples of causal relationships where the causal mechanism is not directly observable. For example, the connection between a light switch and a light is hidden. Similarly, the causal mechanism of the remote control for a television is not directly observable. Small children easily learn how to operate these devices. However, we would claim that it would be easier for a child to learn that a switch on a television controls the television rather than a remote control in an environment where there are several actions which simultaneously precede the operation of television.

When presented with the first example (the child inflating a small yellow balloon after dipping it in water), the situation matches a generalization rule similar to the one in Figure 1[9]. OCCAM constructs a generalization which indicates that blowing air into a balloon results in the balloon being inflated. This changes the temporal link "after" to a causal link "result".

The next example is the child not inflating a large yellow balloon after snapping her fingers. OCCAM first discards its current hypothesis, since it predicts that all balloons will be inflated after air is blown into them. A generalization rule suggests a difference in the size attributes of the balloons is responsible for the difference in the result: small balloon can be inflated and large balloons cannot. (A number of subjects in the experiment reported afterwards that they entertained this hypothesis.)

A third example is consistent with the current hypothesis (the child not inflating a large orange balloon after putting a necklace on), so OCCAM retains its current hypothesis.

The next example (the child inflating a large yellow balloon after dipping it in water) does not agree with the prediction of the current hypothesis. It is discarded. A generalization rule similar to the one in Figure 3 suggests that the action before blowing air into the balloon (dipping the balloon in water) results in a state which enables blowing air into the balloon to inflate the balloon. The generalization which describes this situation is illustrated in Figure 5.[10]

3.2. Learning social theories

In addition to generalization rules which guide the search for causal laws, OCCAM also contains generalization rules for social causation. These generalization rules postulate intentional relationships (Dyer, 1983) between goals, plans and events. There is less empirical support for these generalization rules. However, they fit the same pattern and serve the same purpose as the causal generalization rules. A generalization rule is a template for a causal or social relationship. For example, one social generalization rule is: *If an event (?e) motivates a goal (?g) for someone (?p1), and someone else (?p2) observes the event (?e) and performs an action (?a) which achieves the goal (?g) for ?p1, then the event (?e) motivates the goal (?g) for ?p2.*

This generalization rule is applicable in a number of situations. For example, we recently purchased a new kitten as a companion for our older cat. When the kitten misbehaved by climbing onto the table when we were eating, we closed it in another room. When the older cat saw us do this, he pushed the door open to let the kitten out. This generalization rule would enable one to hypothesize from the older cat's actions that the older cat wanted to let the kitten out. OCCAM uses this generalization rule when it is given a number of examples of parents helping their children and strangers not assisting a child. OCCAM hypothesizes that parents have a goal of preserving the health

---

[9]The generalization rule differs slightly because in Conceptual Dependency, the destination of an action can also change as a consequence of the action. In the representation of "blowing air into a balloon", the type of action is a PROPEL, the object is "air" and the destination is "balloon".

[10]Note that in this example, OCCAM contains no specific world knowledge about inflating balloons. In OCCAM, after a hypothesis is proposed, it is evaluated to determine if it is consistent with existing knowledge. An interesting problem is how one might represent knowledge that X doesn't cause Y. For example, my four year old daughter was watching a television show on the natives of New Guinea who believe that burning the feather of the loudest bird of paradise in a drum will make the drum sound louder when it is hit. She objected strongly to this belief. What sort of reasoning would enable her to arrive at this conclusion? One strategy would be to believe that your knowledge of the world is complete. For example, I know alot about what happens when things are burned, and alot about what makes noises loud. If burning a feather made a drum louder, I would already know this.

of their children. (Of course, before being ruled out by additional examples, OCCAM also entertained a number of incorrect hypotheses such as persons with brown hair have a goal of preserving the health of children).

Once OCCAM has constructed the rule that parents have a goal of preserving the health of their children, it can use it as background knowledge for explanation-based learning. This particular rule was useful in explaining why a parent pays the ransom in a kidnapping episode. A kidnapping schema was created by retaining only those features of the kidnapping episode which were necessary to produce the explanation. This kidnapping schema contains the knowledge that the ransom typically goes to a relative of the hostage because they may be willing to pay money to preserve the health of the hostage.

4. Conclusion

We have argued that learning causal or social theories is facilitated by intra-example relationships which constrain the possible causal relationships. OCCAM makes use of these relationships to learn social and causal theories which serve as background knowledge for explanation-based learning. In OCCAM, generalization rules exploit intra-example relationships to focus the search for causal relationships. An interesting area for future research is to determine whether these generalization rules can be learned from examples to facilitate learning in future situations.

Acknowledgements

Discussions with Michael Dyer helped develop the ideas expressed in this paper. Mort Friedman provided facilitates and encouragement for the experiments in Section 2. Discussions with Keith Holyoak were useful in designing the experiments. I'd like to thank Lynn and Karen for starring in the videotape for the experiments and Chris for directing. Research on OCCAM was supported in part by the UCLA-RAND Artificial Intelligence Fellowship.

References

Bullock, M. (1979). *Aspects of the young child's theory of causality.* Doctoral dissertation, University of Pennsylvania.

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.

DeJong, G. (1983). Acquiring schemata through understanding and generalizing plans. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence.* Karlsruhe, West Germany: Morgan- Kaufmann.

DeJong, G. (1986). An approach to learning from observation. In Michalski, R., Carbonell, J., & Mitchell, T. (Eds.), *Machine Learning, Vol 2.* Los Altos, CA: Morgan Kaufmann Publishers.

Dyer, M. (1983). *In depth understanding.* MIT Press.

Holland, J., Holyoak, K., Nisbett, R., & Thagard, P. (1986). *Induction.* MIT Press.

Kelley, H. (1983). The process of causal attribution. *America Psychologist*, pp. 107-128.

Kolodner, J. (1984). *Retrieval and organizational strategies in conceptual memory: A computer model.* Hillsdale, NJ.: Lawrence Erlbaum Associates.

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

Lebowitz, M. (1980). *Generalization and memory in an integrated understanding system* (Computer Science Research Report 186). Yale University.

Lebowitz, M. (1986). Not the path to perdition: The utility of similarity-based learning. In *Proceedings of the National Conference on Artificial Intelligence*. American Association for Artificial Intelligence.

Lebowitz, M. (1986). Concept learning in an rich input domain: Generalization-based Memory. In Michalski, R., Carbonell, J., & Mitchell, T. (Eds.), *Machine Learning, Vol 2*. Los Altos, CA: Morgan Kaufmann Publishers.

Michalski, R. (1977). A system of programs for computer aided induction: A summary. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. Cambridge, Mass.: Morgan-Kaufmann.

Michotte, A. (1963). *The perception of causality*. New York: Basic Books, Inc.

Mitchell, T. (1982). Generalization as search. *Artificial Intelligence*, Vol. 18(2).

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

Mooney, R. & DeJong, G. (1985). Learning schemata for natural language processing. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*. Los Angeles, CA: Morgan-Kaufmann.

Pazzani, M. (1985). Explanation and generalization-based memory. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*. Irvine, CA.

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

Pazzani, M., Dyer, M. & Flowers, M. (1987). Using prior learning to facilitate the learning of causal theories. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*. Milan, Italy: Morgan-Kaufmann.

Piaget, J. (1930). *The child's conception of physical causality*. London: Kegan Paul.

Salzberg, S. (1985). Heuristics for inductive learning. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*. Los Angeles, CA: Morgan-Kaufmann.

Schank, R.C. & Abelson, R.P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Shultz, T. (1982). Rules of causal attribution. Monographs of the Society for Research in *Child Development*, Vol. 47.

Shultz, T., Fisher, G., Pratt, C., & Rulf, S. (1986). Selection of causal rules. *Child Development*, 57, 143-152.

Shultz, T. & Mendelson, R. (1975). The use of covariation as a principle of causal analysis. *Child Development*, 46, 394-399.