

EXPLANATION AND GENERALIZATION BASED MEMORY

Michael J. Pazzani

**April 1986
CSD-860080**

Explanation and Generalization Based Memory

Michael J. Pazzani
Artificial Intelligence Laboratory
3731 Boelter Hall
University of California at Los Angeles
Los Angeles, CA 90024
and
The Aerospace Corporation
P.O. Box 92957
Los Angeles, CA 90009
(213)648-6644
Topics: Learning, Memory
Word Count: 4716

Abstract

A model of memory and learning is presented which indexes a new event by those features which are relevant in explaining why the event occurred. As events are added to memory, generalizations are created which describe and explain similarities and differences between events. The memory is organized so that when an event is added, events with similar features are noticed. An explanation process attempts to explain the similar features. If an explanation is found, a generalized event is created to organize the similar events and the explanation is stored with the generalized event.

Introduction

The goal of this research is to identify the role of explanation in a generalization based memory. A computer program, OCCAM, has been implemented which learns in two domains. In one domain, the program starts out with general knowledge about coercion represented as a meta-MOP [Schank 82]. After some examples, it creates a MOP 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 MOP which represent an inherent flaw in kidnapping: that the victim can testify against the kidnapper, since the kidnapper can be seen by the victim. This specialization is stored as a sub-MOP of the kidnapping MOP which is indexed by the result of the kidnapper going to jail which is a goal failure of the kidnapper. After some more examples, a similarity is noticed about the kidnapping of infants. This coincidence starts an explanation process which explains the choice of victim to avoid a possible goal failure, since infants cannot testify.

In the second domain, OCCAM models a 4-year old child trying to figure out why she can inflate

some balloons but not others. Figure 1 is a protocol of a conversation with L. After successfully blowing up a red balloon, and unsuccessfully blowing up a green balloon, it appears that L. believes that color is an important feature to predict whether or not the balloon can be inflated. After one more example, she abandons this belief, apparently in favor of the hypothesis that L. can blow up balloons after M. has.¹

1 M. is blowing up a red balloon.
2 L: "Let me blow it up."
3 M. lets the air out of the balloon and hands it to L.
4 L. blows up the red balloon.

5 L. picks up a green balloon and tries to inflate it.
6 L. cannot inflate the green balloon.
7 L. puts down the green balloon and looks around.
8 L: "How come they only gave us one red one?"
9 M: Why do you want a red one?
10 L: I can blow up the red ones.

11 M. picks up a green balloon and inflates it.
12 M. lets the air out of the green balloon and hands it to L.
13 M: "Try this one."
14 L. blows up the green balloon.
15 L. gives M. an uninflated blue balloon.
16 L: "Here, let's do this one."

Figure 1: Protocol of a 4 year-old trying to blow up balloons

In this second domain, there is much less explanation capability. However, it appears that L. is still trying to produce cause and effect relationships. Another difference between the two domains is that the kidnapping examples are adding causal and motivational explanations to specializations of a more general knowledge structure. This specialization is also a generalization of specific examples. In the balloon examples, the generalizations are built without specializing a particular knowledge structure. Without a good explanation capability and top down knowledge, it is not surprising that the initial generalizations turn out to be wrong.

There are a couple of interesting features of this type of learning:

- The explanation process eliminates the problem of including unrelated coincidences in generalized events. For example, all of the infants kidnapped in the events presented to OCCAM have blond hair. This feature is not used in the explanation, so it is not included in the generalized event.
- There is causal and motivational information associated with generalized events. This

¹Of course, the real reason is that the balloon is stretched after it has been inflated.

information states why various features are included in the generalized event.

- Generalized events in memory are indexed by inferred features, such as the goals and goal conflicts of the participants, in addition to the features of generalized event.
- The explanation process can make use of the generalized events in memory. Explanation consists of a rule based component similar to PAM [Wilensky 78] and a memory based explanation component. The rules state such things as that if someone says they are going to hurt a family member, this motivates a goal of preserving the health of the family member. There are no special rules about kidnapping. Therefore, it is not capable of explaining the kidnapping of infants until it has built a generalized event about the victim testifying against the kidnapper. The explanation process uses intentional links [Dyer 83] to specify the relationships between goals, plans and events.

Related Work

Much early work on learning (e.g., [Winston 70], [Vere 75], and [Hayes-Roth 77]), centered on the acquisition of a concept from a number of examples. A characteristic description of a class of objects was built by inductive means by considering positive (and, in some instances, negative) examples. The work reported here differs from this work on concept acquisition in a number of ways. First, these programs were "told" what concepts to learn and examples were identified as positive or negative instances. In contrast, OCCAM is not told what to learn. Instead, OCCAM incrementally learns new concepts from examples as a natural consequence of organizing memory around similarities. Secondly, the generalized events built by OCCAM do not contain all features common to the examples. Its explanation process distinguishes between relevant and coincidental features.

DeJong presents a model of explanation based learning [DeJong 83] which learns schemata from a single example. His program constructs an explanation of relationships between various components of an event by a knowledge-intensive understanding process similar to PAM. The explanation and the event are then generalized by retaining only those parts used in the explanation. Our work differs from DeJong's in a number of aspects. First, OCCAM can learn without the explanation process, (e.g., the balloon examples). The price it pays for this learning is making mistakes, which are corrected when contradictory examples are seen. Secondly, OCCAM learns incrementally. It is difficult to imagine a system learning the specialized motivation for kidnapping infants from the first example of a kidnapping, since the explanation process can find an explanation for kidnapping any person. In OCCAM, after the basic kidnapping schema (or MOP) is learned, later examples focus OCCAM on explaining coincidences about the age of the victims. Finally, the explanation process makes use of other events or generalized events.

In IPP [Lebowitz 80], Lebowitz is concerned with making "factual" generalizations from natural

language texts about terrorism stories. No attempt is made to perform a causal or explanatory analysis. Therefore, no distinction is made between relevant or coincidental features. After a number of diverse examples, IPP can correct its generalizations to remove coincidences which are contradicted. Lebowitz [Lebowitz 84] has also suggested applying explanation based learning techniques to the generalized events to focus the learning process. In OCCAM, this suggestion is taken one step further, by postulating causality and intentionality when no explanation can be found.

CYRUS [Kolodner 84] is a program which organizes and searches a model of episodic memory. Like IPP, it does not produce an explanation of its generalizations. It avoids the problem of indexing on coincidentally similar features by an a priori set of relevant features.

OCCAM is inspired by Schank's dynamic memory model. However, in addition to "failure-driven" learning (i.e., learning when expectations are violated), OCCAM uses similarities to focus the learning and explanation process.

Learning and Memory in OCCAM

OCCAM makes a distinction between three types of generalized events. *Explanatory generalized events* are MOPs created as a specialization of a more general MOP. Associated with each explanatory generalized event are new causal and goal relationships in addition to those of the more general event. For example, the kidnapping of infants is an explanatory generalized event which includes the special motivation for selecting the hostage. *Tentative generalized events* are MOPs inductively generalized from examples without an appropriate more general event (i.e., a MOP or meta-MOP). They also contain causal relationships but these are not confirmed by an explanation process. For example, the generalization that L. can inflate red balloons but not green ones is a tentative generalized event since the explanation was not verified by causal principles. *Organizational generalized events* are MOPs also created as specialization of more general MOPs. However, they add no additional explanatory information. They correspond to the factual generalizations of IPP and serve mainly to organize the memory. An example organizational generalized event would be kidnappings where the hostages grandmother paid the ransom².

There are two parts to the incremental learning algorithm used by OCCAM. The first step is to find the appropriate place in memory to index a new events. The memory is organized so that a new event will be added to memory in the same place as similar events. The second step is to attempt to create a generalization.

²Unless, of course, some explanation could be found.

The search for an appropriate place to insert an event in memory starts at the most general MOP in memory which represents that class of events (e.g., coercion). The features of the new event are used as indices to traverse the memory from the most general MOP to the most specific MOP(s).³ Once a specific MOP is found, if it is a tentative generalized event, an analysis is performed to see if the tentative causal relationships apply to this new event. If they do not (as in the balloon examples), the tentative generalization is considered erroneous.⁴

After the most specific applicable MOP is found, similar events are found by using the features of the new event as indices. Next, generalization is attempted by a number of rules which postulate causal or intentional relationships. For example, one rule states *If an action always precedes a state, postulate the action causes the state.* Other rules which postulate goal relationships will be discussed in the next section. An explanation process is then used to verify the postulated causal or intentional relationships. This explanation process marks all features necessary for establishing the relationships. The explanation process used here is more focused than that used in DeJong's work. Rather than asking general questions such as "Why did this happen?", more specific questions are used such as "Was there an action which motivates a goal before this action which this action achieves." A new MOP may be created depending on the result of the generalization:

- If the explanation is successful, then an explanatory generalized event is built and indexed under the most specialized MOP by the new features establishing the explanation. The new event and any similar events are organized under this new generalization, indexed by the features not used in the generalized event.
- If the explanation process is unsuccessful, and the most specific MOP is a tentative generalized event, a new more specific tentative generalized event is created. That is, the causal relationships postulated by the generalization rules are assumed to hold until they are contradicted.
- If the explanation process is unsuccessful, and the most specific MOP is an explanatory generalized event, a default rule is used to attempt to form an organizational generalized event. This notes that there appears to be a coincidental relationship but does store any justification.⁵

³This traversal can find an organizational generalized events which are predicted to be similar to the new event, but are not. When this occurs, the erroneous generalized event is corrected in a manner similar to [Lebowitz 82].

⁴Currently, an erroneous generalization is marked as erroneous and not found in the the normal traversal process, but not discarded.

⁵This approach allows for a organizational generalized event to become an explanatory generalized event after enough confirming examples as described in [Lebowitz 82]. At this conversion, the explanation postulated by the generalization rules would be stored with the generalized event.

Examples

This section presents two examples of OCCAM learning. In the first, the creation of an explanatory generalized event is illustrated in the domain of kidnapping. In the second, OCCAM creates tentative generalized events to describe the class of balloons which L. can blow up.

Kidnapping

The meta-MOP for coercion involves a PREParation, a THREAT, a DEMAND, and several RESULT scenes. It usually involves at least three roles: an ACTOR, who performs the PREParation, and says he will carry out the THREAT unless his DEMAND is met; an OBJECT which is the object of the PREParation and the THREAT (i.e., in kidnapping the hostage is the OBJECT); and the VICTIM which receives the THREAT, and usually performs one of the RESULTS.⁶ The coercion meta-MOP is intended to be very general and account for many situations from kidnappings to playground arguments (e.g., "If you don't let me pitch, I'm gonna take my ball and go home"). Figure 2 illustrates an example of kidnapping which is a kind of coercion. In this Figure, the notation "the(FEATURE)" indicates the actual value of the feature is the same as the value of that feature. OCCAM makes use of this relationship in its generalizations.

The initial state of the memory of OCCAM contains only the coercion meta-MOP (mm-COERCE). K1, the example in Figure 2, is then added to memory. It is indexed under mm-COERCE by all of its features (i.e., its scenes and roles). The next example, K2, is similar to K1, except the AGE of the OBJECT is an infant, some minor difference in the features of the VICTIM and the ACTOR, and there is no trial in which the ACTOR goes to jail. Figure 3 gives an edited transcript of the creation of a MOP which describes the kidnapping of a family member for a monetary ransom. The similarities between K1 and K2 are noted. Then, a rule, GENERALIZE-RESULTS, is used to postulate an explanation for this similarity. This rule states *Look for an action before the RESULT which motivates a goal which the RESULT achieves or an action before the RESULT which is part of plan which the RESULT realizes*. In this example, a goal of preserving the health of the OBJECT by the VICTIM is inferred and paying the ransom achieves this goal. Additionally, the ACTOR is performing the plan of keeping a bargain when he gives the OBJECT back. In general, a better explanation utilizes the goals rather than the plans. However, in this case, it's not possible to infer why the kidnapper releases the hostage. This new MOP (MOP.327) is indexed under mm-COERCE by the relevant features, and the inferred goal as shown in Figure 4.

The next event added to memory is K3, which is similar to K1 in that the kidnapper goes to jail

⁶The VICTIM in kidnapping is not the hostage but the person who pays the ransom.


```

-----
K1: COERCION
ACTOR human  NAME Joe K.
              HEIGHT tall
              AGE 30s
              HAIR brown
OBJECT human  NAME John V.
              HEIGHT short
              AGE teens
              HAIR blond
              RELATION family TYPE son
                                OF the(VICTIM)
VICTIM human  NAME Dad V.
              HEIGHT tall
              AGE 40s
              HAIR blond
              RELATION family TYPE father
                                OF the(OBJECT)
PREP atrans  ACTOR the(ACTOR)
              TO the(ACTOR)
              OBJECT the(OBJECT)
DEMAND poss-by ACTOR the(ACTOR)
              OBJECT money AMOUNT 50000
THREAT health OF the(OBJECT)
              VAL -10
RESULT atrans TO the(VICTIM)
              OBJECT the(OBJECT)
              ACTOR the(ACTOR)
              FROM the(ACTOR)
RESULT atrans TO the(ACTOR)
              OBJECT money AMOUNT 50000
              ACTOR the(VICTIM)
              FROM the(VICTIM)
RESULT $trial SENTENCE 15
              VERDICT guilty
              WITNESS the(OBJECT)
              CRIMINAL the(ACTOR)
-----

```

Figure 2: An Example of Coercion: A Kidnapping

after the hostage testifies. There are minor differences in the features of the participants. MOP.327 the kidnapping MOP is the most specific MOP which is not contradicted by K3. It has an additional RESULT which is similar to a result of K1. A rule which states *If there is a RESULT which thwarts a goal, look for an action before the RESULT which enables the RESULT* finds an inherent flaw in kidnapping: the hostage sees the kidnapper when he is abducted and can testify against the kidnapper. A new MOP, MOP.332 (jailed-kidnapper) is created and indexed under MOP.327 (kidnap) by the indices of the RESULT, the goal failure, and the PREPARation which enables the goal RESULT which thwarts the goal as shown in Figure 4. K1 and K3 are indexed under this new MOP, while k2 remains indexed under the kidnapping MOP.

 Looking for conflicts between event k2 and MOP mm-COERCE... none found.
 Looking for similar events under mm-COERCE... found (K1).

Similarities:

COERCION

ACTOR human HEIGHT tall
 AGE 30s
 HAIR brown

OBJECT human RELATION family TYPE son
 OF the(VICTIM)

VICTIM human RELATION family TYPE father
 OF the(OBJECT)

PREP atrans ACTOR the(ACTOR)
 TO the(ACTOR)
 OBJECT the(OBJECT)

DEMAND poss-by ACTOR the(ACTOR)
 OBJECT money

THREAT health OF the(OBJ)
 VAL -10

RESULT atrans TO the(VICTIM)
 OBJECT the(OBJECT)
 ACTOR the(ACTOR)
 FROM the(ACTOR)

RESULT atrans TO the(ACTOR)
 OBJECT money
 ACTOR the(VICTIM)
 FROM the(VICTIM)

Running generalization rule GENERALIZE-RESULTS.

Inferring RESULT REALIZES PLAN (KEEP-BARGAIN)

Inferring RESULT ACHIEVES GOAL (P-HEALTH)

Making sub MOP MOP.327 {kidnap} of mm-COERCE from (K2 K1)

Used in explanation:

COERCION

ACTOR human

OBJECT human

VICTIM human RELATION family TYPE father
 OF the(OBJECT)

DEMAND poss-by ACTOR the(ACTOR)
 OBJECT money

THREAT health OF the(OBJECT)
 VAL -10

RESULT atrans TO the(VICTIM)
 OBJECT the(OBJECT)
 ACTOR the(ACTOR)
 FROM the(ACTOR)

RESULT atrans TO the(ACTOR)
 OBJECT money
 ACTOR the(VICTIM)
 FROM the(VICTIM)

Figure 3: Forming an Explanatory Generalized Event

K4, another kidnapping of a blond infant in which the kidnapper was not caught, is added to

memory next. MOP.327 (kidnap) is found to be the most specific MOP which describes K4. A similarity is noticed between k4 and k2, the OBJECTs are both blond infants. An applicable generalization rule states *If the PREParation is performed on the an object, look for other MOPs which have a goal failure. Check if the PREParation avoids the goal failure, if it does postulate the ACTOR performed the PREParation to avoid the goal failure.* In this example, the goal of preserving freedom of the kidnapper cannot be thwarted by the infant testifying. A new MOP is created indexed by the AGE of the OBJECT (and not the hair color) as shown in Figure 4.

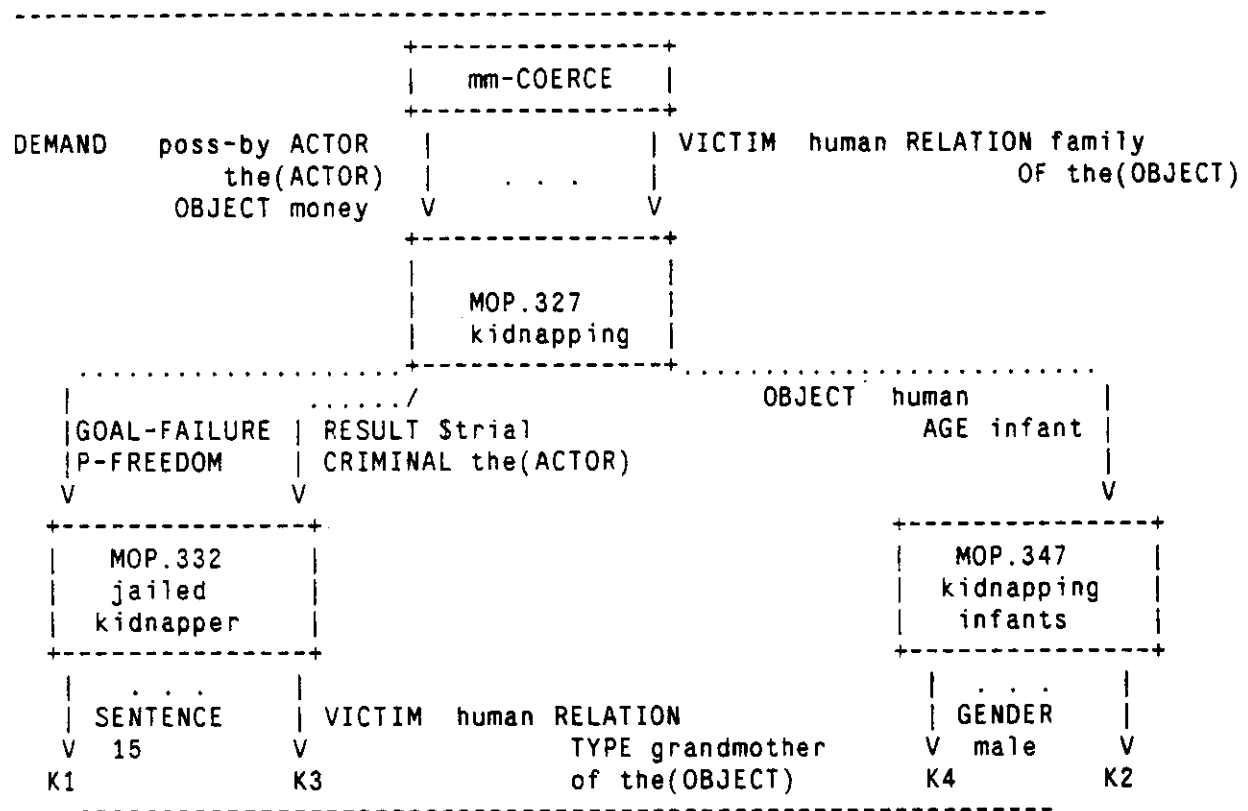


Figure 4: Memory after creating 3 specializations of mm-COERCE

These examples illustrate the process of creating an explanatory generalized event. In a more realistic set of examples, several organizational generalized events would also be created and each MOP would index a greater number of events and sub-MOPs. The point of creating a explanatory generalizations is to create specialized explanations for situations. With only mm-COERCE in memory, the explanation of kidnapping an infant would be "The ACTOR wants the VICTIM to do something". After MOP.327 (kidnap) is created, the explanation would be "The ACTOR wants a member of the OBJECT's family to give him money". After MOP.347 (kidnapping infants) is created the explanation would be "The ACTOR wants a member of the OBJECT's family to give him money and the ACTOR wants to avoid being convicted, so he's kidnapping an infant since infants can't

testify”.

Inflating Balloons

In this second example, the initial memory is essentially empty.⁷ The examples are conceptual dependency representations of the events taking place in Figure 1.

B1 which describes L. successfully blowing up a red balloon is first added to memory. Next, B2 which describes L unsuccessfully blowing up a green balloon is added to memory and similarities are noticed. An applicable generalization rule is *If two actions have different results, and they are performed on different object, assume the differing features enable the action to produce the result.* This produces a question for the explanation process "Does that the state of a balloon being red enables the balloon to be inflated when L. blows air into it?". This cannot be confirmed⁸, but it is saved as a tentative generalized event (MOP.348). Associated with this is the explanation which describes the difference in results as enabled by the color of the balloon. B1 and B2 are indexed under this MOP. B1 is indexed by the features COLOR = RED and the RESULT successfully inflated. B2 is indexed by the COLOR = GREEN and the RESULT unsuccessfully inflated.

Next, B3, a green balloon successfully blown up by L. after M. deflated it is added to memory and MOP.348 is the most specific MOP which describes B3. Since MOP.348 is tentative, its explanation is reconsidered to see if B3 contradicts it. It is contradicted since, MOP.348 predicts green balloons cannot be blown by L. This MOP is marked as erroneous. Now an applicable generalization rule states *If an action always precedes an action which results in a state, assume that the initial action results in a state which enables the subsequent action to produce the result.* In this case, the action which precedes L. successfully blowing a balloons is M. deflating the balloon. This action does not occur when L. cannot blow up the balloon. A new tentative generalized event (MOP.349) is created which saves the postulated explanation. B2 is indexed off this relation by the unsuccessful RESULT and the COLOR = GREEN. Another tentative generalized event (MOP.350) with a more specialized explanation is created and indexed under the successful RESULT. B3 and B1 are indexed off this MOP by the COLOR = GREEN and COLOR = RED, respectively. See Figure 5.

Before adding B3 to memory, OCCAM did not consider the event preceding the attempt to inflate the balloon to be significant. After postulating that the preceding event is needed to explain a result it has created a MOP describing a preparation, an action, and a result.

⁷ Actually, it contains mm-COERCE, which is of no help in blowing up balloons.

⁸ If it were confirmed, an explanatory generalized event would be created.

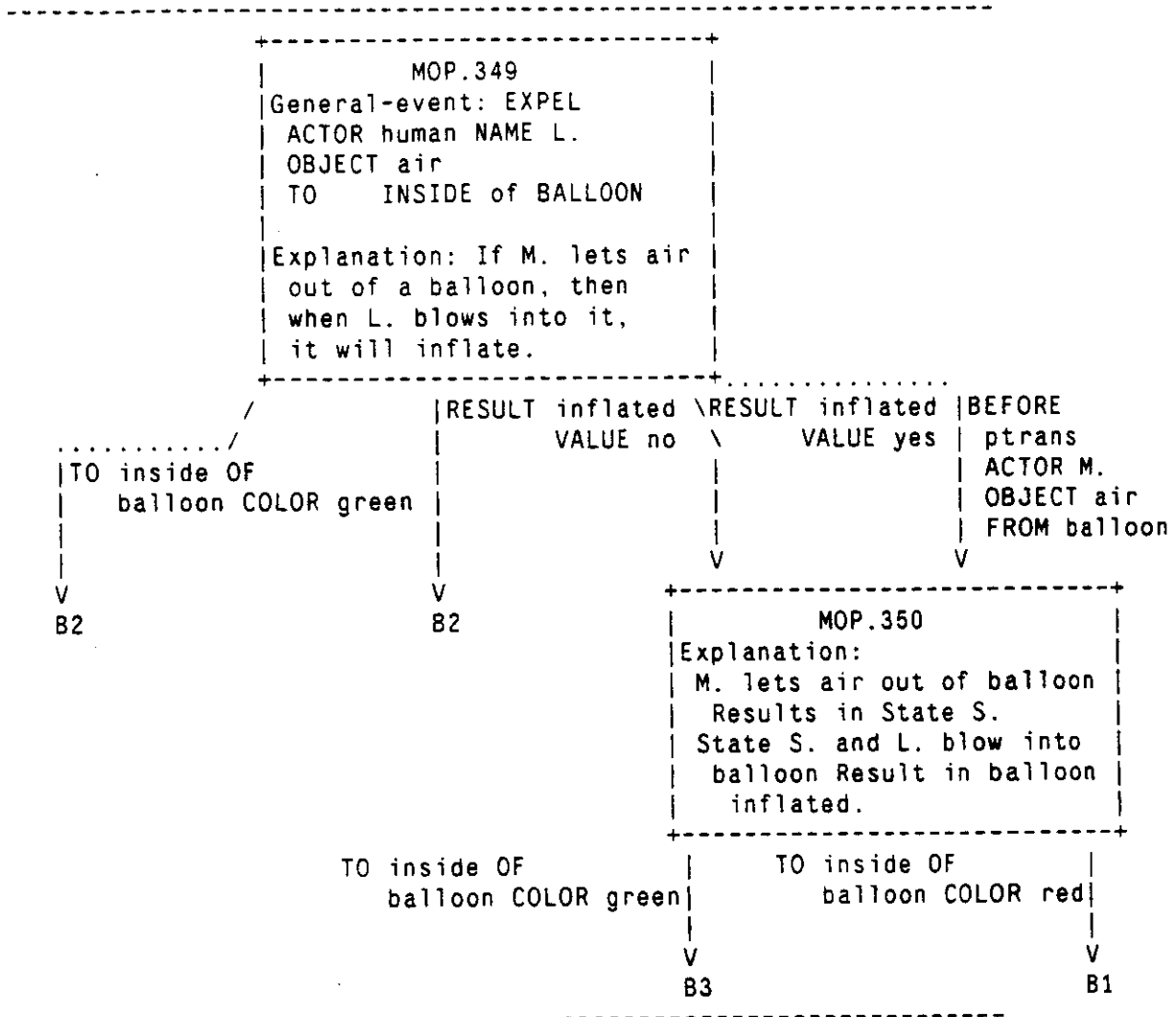


Figure 5: Memory after 3 examples of inflating balloons .

CONCLUSION

OCCAM is a program which organizes memories of events and learns by creating explanatory and tentative generalized events. It addresses the issue of deciding which features are relevant in producing a generalization. It answers this question by proposing those features which are relevant in producing a explanation. This has some implications for expert systems which operate by recalling similar experiences. Should a medical expert system index a case by the patients weight, height, clothing or jewelry? The answer proposed here is to use these as indices in explanatory generalized events only if they are relevant in the pathological explanation. Organizational generalized events describe those situations where a coincidence is noted but there is no explanation. These coincidences might initiate and focus the search for new pathological knowledge.

OCCAM also raises some questions that we have not yet addressed. How is a tentative

generalization confirmed? Lebowitz [Lebowitz 82] addresses this problem for generalizations without causal explanations. Is a separate explanation process necessary? OCCAM uses the memory (e.g., to find examples of failed goals), but it also has a rule based component. The inductively produced tentative generalizations can produce a rule-like conditional explanation (see Figure 5). If a tentative generalized event were used in the explanation process, what happens when the tentative generalization is later contradicted? Some psychological experimentation is necessary to see how much bookkeeping people actually do. In addition to trying to explain proposed causal relations, is there general mechanism to refute an proposed explanation? OCCAM currently looks for counter examples, but people seem to be able to say "There is just no way that could cause that." without coming up with a counterexample. How are meta-MOPs created? Is it possible to inductively generalize mm-COERCE from playground disputes and then specialize it with kidnapping examples? The ultimate goal of this research is to combine these two types of learning. As a "child", OCCAM should inductively build generalizations abouts causality and intentionality, creating MOPs and meta-MOPs (e.g., coercion) from childhood experiences with friends, siblings and parents. As an "adult", OCCAM should deductively specialize MOPs and meta-MOPs as required for understanding experiences such as kidnappings, business deals, and political disputes and rely on inductive generalizations to learn in novel areas.

Acknowledgements

Discussions with Mike Dyer and Margot Flowers helped in the evolution of the ideas in this paper. Communications with Michael Lebowitz were also fruitful.

References

- [DeJong 83] DeJong, G.
Acquiring Schemata Through Understanding and Generalizing Plans.
In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. Karlsruhe, West Germany, 1983.
- [Dyer 83] Dyer, M.
In Depth Understanding.
MIT Press, 1983.
- [Hayes-Roth 77] Hayes-Roth, F. and McDermott, J.
Knowledge Acquisition from Structural Descriptions.
In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*.
Cambridge, Mass., 1977.
- [Kolodner 84] Kolodner, J.
Retrieval and Organizational Strategies in Conceptual Memory: A Computer Model.
Lawrence Erlbaum Associates, Hillsdale, NJ., 1984.

- [Lebowitz 80] Lebowitz, M.
Generalization and Memory in an Integrated Understanding System.
Computer Science Research Report 186. Yale University, 1980.
- [Lebowitz 82] Lebowitz, M.
Correcting Erroneous Generalizations.
Cognition and Brain Theory 5(4), 1982.
- [Lebowitz 84] Lebowitz, M.
Interest and Predictability: Deciding What to Learn, When to Learn.
Technical Report. Columbia University, 1984.
- [Schank 82] Schank, R.
Dynamic Memory: A Theory of Reminding and Learning in Computers and People.
Cambridge University Press, 1982.
- [Vere 75] Vere, S.
Induction of Concepts in the Predicate Calculus.
In *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*. Tbilisi, USSR, 1975.
- [Wilensky 78] Wilensky, R.
Understanding Goal Based Stories.
Computer Science Research Report 140, Yale University, 1978.
- [Winston 70] Winston, P.
Learning Structural Descriptions from Examples.
Technical Report AI-TR-231, MIT, 1970.

List of Figures

Figure 1: Protocol of a 4 year-old trying to blow up balloons	1
Figure 2: An Example of Coercion: A Kidnapping	6
Figure 3: Forming an Explanatory Generalized Event	7
Figure 4: Memory after creating 3 specializations of mm-COERCE	8
Figure 5: Memory after 3 examples of inflating balloons	10