Journal of Experimental Psychology:
Learning, Memory, and Cognition
1991, Vol. 17, No. 3, 416–432

Influence of Prior Knowledge on Concept Acquisition: Experimental and Computational Results

Michael J. Pazzani
University of California, Irvine

The influence of the prior causal knowledge of subjects on the rate of learning, the categories formed, and the attributes attended to during learning is explored. Conjunctive concepts are thought to be easier for subjects to learn than disjunctive concepts. Conditions are reported under which the opposite occurs. In particular, it is demonstrated that prior knowledge can influence the rate of concept learning, and that the influence of prior causal knowledge can dominate the influence of the logical form. A computational model of this learning task is presented. To represent the prior knowledge of the subjects, an extension to explanation-based learning is developed to deal with imprecise domain knowledge.

It has been suggested (e.g., Murphy & Medin, 1985; Pazzani, Dyer, & Flowers, 1986; Schank, Collins, & Hunter, 1986) that a person's prior knowledge influences the rate or accuracy of learning. In this article, I explore the influence of prior causal knowledge on the number of trials to learn a concept, the concepts formed, and the selection of attributes used to form hypotheses.

In concept-identification tasks, it has been found that the logical form of a concept influences the number of trials required to learn a concept (Dennis, Hampton, & Lea, 1973; Shepard, Hovland, & Jenkins, 1961). In particular, conjunctive concepts require fewer trials to learn than disjunctive concepts (Bruner, Goodnow, & Austin, 1956). Here the interaction between the prior knowledge and the logical form of concepts is investigated. I hypothesize that the prior knowledge of the learner is as important an influence in concept learning as the logical form of the concept.

Context-dependent expectations facilitate cognition on many different tasks. For example, prior presentation of a semantically related word increases the speed with which words are distinguished from nonwords (Meyer & Schvaneveldt, 1971). Similarly, Palmer (1975) found that, in the context of a face, less detail was necessary to recognize drawings of facial parts than was necessary in isolation. In addition, it has been found that prior expectations influence the perception of covariation (Chapman & Chapman, 1967, 1969) and result in more robust judgments of covariation by reducing the impact of atypical data points (Wright & Murphy, 1984).

This research has two goals. First, if a form of prior knowledge can reverse the superiority of conjunctive concepts, it provides additional evidence for the importance of this often-ignored factor on concept acquisition. Second, the type of knowledge that subjects bring to bear on the learning task is analyzed, and it is shown that this knowledge cannot easily be represented as a set of inference rules with necessary and sufficient conditions. As such, this provides constraints on computational models of the concept-acquisition task.

There are two major computational approaches to learning. Empirical learning techniques (Michalski, 1983; Mitchell, 1982) operate by searching for similarities and differences between positive and negative examples of a concept. Current connectionist learning techniques (e.g., Rumelhart, Hinton, & Williams, 1986) are essentially empirical learning techniques. Explanation-based learning techniques (DeJong & Mooney, 1986; Mitchell, Kedar-Cabelli, & Keller, 1986) operate by forming a generalization from a single training example by proving that the training example is an instance of the concept. The proof is constructed by an inference process that makes use of a domain theory, a set of facts and logical implications. In explanation-based learning, a generalization is created by retaining only those attributes of a training example that are necessary to prove that the training example is an instance of the concept. Explanation-based learning is a general term for learning methods such as knowledge compilation (Anderson, 1989) and chunking (Laird, Newell, & Rosenbloom, 1987) that create new concepts that deductively follow from existing concepts.

Pure empirical-learning techniques do not make use of prior knowledge during concept acquisition. Therefore, a model of human learning that is purely empirical would predict that if two learning problems are syntactically isomorphic, the problems will be of equal difficulty for a human learner. A model of human learning that relied solely on explanation-based learning could not account for the fact that subjects are capable of learning concepts in the absence of any domain knowledge. In addition, current explanation-based learning methods assume that the domain theory is

Portions of the data in this article were reported at the meeting of the Cognitive Science Society, Ann Arbor, Michigan, August 1989. This research is supported by a faculty research grant from the Academic Senate Committee on Research of the University of California, Irvine, and National Science Foundation Grant IRI-8908260. Comments by David Schulenburg and Caroline Ehrlich on an earlier draft of this article were helpful in improving the presentation. Discussions with Lawrence Barsalou, James Hampton, Brian Ross, Tom Shultz, Glenn Silverstein, Edward Wisniewski, and Richard Doyle helped develop and clarify some of the ideas in the article. Clifford Brunk, Timothy Cain, Timothy Hume, Linda Palmer, and David Schulenburg assisted in the running of experiments.

Correspondence should be addressed to Michael J. Pazzani, ICS Department, University of California, Irvine, California 92717.

complete, correct, and consistent. This same assumption cannot be made about the prior knowledge of human subjects (Nisbett & Ross, 1978).

Many have argued (e.g., Flann & Dietterich, 1989; Lebowitz, 1986; Pazzani, 1990) that a complete model of concept learning must have both an empirical and an explanation-based component. Prior empirical studies (e.g., Barsalou, 1985; Nakamura, 1985; Wattenmaker, Dewey, Murphy, & Medin, 1986), together with the experiments reported here, provide constraints on how these learning methods may be combined. After the first experiment in this article, a novel model for combining the two learning methods is proposed. Next, simulations of the model are used to make predictions about the learning rates and biases. These predictions are then tested with experiments on human subjects. Where necessary, revisions to the model are proposed to account for differences between prediction and observations.

Nakamura (1985) investigated the role that prior knowledge has on the accuracy of classification learning. In particular, he analyzed the interaction between learning linearly separable and nonlinearly separable concepts and the type of instructions provided to subjects. One set of instructions was neutral in that it asked the subjects to correctly classify stimuli (descriptions of flowers). A second set of instructions gave subjects a background theory that helped with the task (e.g., one class of flowers attracts birds and the birds cannot see color and are active at night). The linearly separable task resulted in fewer errors during learning using theory instructions than under neutral instructions. This pattern was reversed for the nonlinearly separable task: Neutral instructions led to fewer errors than theory instructions. One explanation for this finding is that the concept with the fewest violations of prior knowledge is easier for subjects to learn. Such a violation occurs when a subject is given feedback that contradicts prior knowledge (e.g., a flower that blooms during the day only attracts a bird that is active at night). In this experiment, the linearly separable concept required fewer violations of the prior knowledge than the nonlinearly separable concept. This explanation is also supported by later studies (Pazzani & Silverstein, 1990; Wattenmaker et al., 1986) that suggest a nonlinearly separable concept consistent with prior knowledge is easier to learn than a linearly separable concept that violates prior knowledge.

In this article, I compare the learning rates of simple conjunctive and disjunctive concepts. Note that both of these classes of concepts are linearly separable. Therefore, the experiments will test whether the effect of prior knowledge is more pervasive than that suggested by previous work that studied the role of prior knowledge in learning linearly separable and nonlinearly separable concepts.

Experiment 1

All of the experiments in this article use a similar method to investigate the effect of prior knowledge on concept acquisition. One group of subjects performs a standard concept-acquisition experiment. This group of subjects must determine whether each stimuli is an example of an *alpha*. The stimuli are photographs of a person doing something with a balloon. The stimuli differ in terms of the color of the balloon (yellow or purple), the size of the balloon (small or large), the age of the person (adult or child), and the action the person is doing (stretching the balloon or dipping the balloon in water). Existing knowledge about inflating balloons is irrelevant for this group of subjects. Another group of subjects uses the same stimuli. However, the instructions indicate the subject must predict whether the balloon will be inflated when the person blows into it. In this condition, called the inflate condition, the subject's prior knowledge may provide expectations about likely hypotheses. The goal of the experiments is to determine conditions under which these expectations facilitate or hinder the concept-acquisition task.

The purpose of the first experiment was to investigate the interaction between prior knowledge and the acquisition of conjunctive and disjunctive concepts. The experiment follows a 2 (concept form [conjunctive vs. disjunctive]) × 2 (instruction set [alpha vs. inflate]) between-subjects design.

The conjunction to be learned was "size = small and color = yellow." The disjunction to be learned was "age = adult or action = stretching a balloon." Note that with the inflate instructions, the conjunctive concept is not implied by prior knowledge, whereas the disjunctive concept is implied by this knowledge. It is also important to stress that the prior background knowledge[1] (e.g., adults are stronger than children and stretching a balloon makes it easier to inflate) is not sufficient for subjects to deduce the correct relationship in the absence of any data. There are several possible consistent relationships including a conjunctive one (adults can inflate only balloons that have been stretched) and the disjunctive relationship tested in this experiment. Experiment 2 tests whether prior knowledge also facilitates a conjunctive concept consistent with prior knowledge.

The following three predictions were made about the outcome of this experiment. First, subjects in the alpha-conjunction category are predicted to take fewer trials than those in the alpha-disjunction category. In the absence of prior knowledge, it was anticipated that the data would replicate the finding that conjunctions are easier to learn than disjunctions. Second, subjects in the inflate-disjunction category are predicted to take fewer trials than those in the inflate-conjunction category. It is anticipated that the influence of prior knowledge would dominate the influence of logical form. Third, subjects in the inflate-disjunction category are predicted to take fewer trials than those in the alpha-disjunction category. Prior knowledge can be expected to facilitate learning only with the inflate instructions. The rationale here is that there are fewer

[1] As part of a previous experiment (Pazzani, in press), 80 University of California, Los Angeles, undergraduates were asked several true–false questions concerning what balloons are more likely to be inflated. All of the subjects indicated that stretching a balloon makes it easier to inflate, that adults can inflate balloons more easily than small children, and that the color of a balloon does not affect the ease of inflation. Seventy-two percent (58) of the subjects felt that the shape of the balloon influenced the ease of inflation. Of these subjects, 63% (37) felt that long balloons were harder to inflate than round balloons and the remainder felt that long balloons were easier. In the experiments in this article, all of the balloons were round balloons.

hypotheses consistent with both prior knowledge and the data than those consistent with the data alone. Therefore, it is anticipated that fewer trials would be needed to rule out alternatives when the prior knowledge of the subject is applicable in the learning task.

Method

Subjects. The subjects were 88 male and female undergraduates attending the University of California, Irvine, who participated in this experiment to receive extra credit in an introductory psychology course. Each subject was tested individually. Subjects were randomly assigned to one of the four conditions.

Stimuli. The stimuli consisted of pages from a photo album. Each page contained a close-up photograph of a balloon that varied in color (yellow or purple) and size (small or large) and a photograph of a person (either an adult or a 5-year-old child) doing something to the balloon (either dipping it in water or stretching it). For the inflate subjects, the back of the page of the photo album had a picture of the person with a balloon that had been inflated or a balloon that had not been inflated. For the alpha subjects, a card with the words *Alpha* or *Not Alpha* was on the reverse side of each page. Because there are four attributes that can take on two values, there are a total of 16 unique stimuli. Of these stimuli, 12 are positive examples of a disjunction of two attributes and 4 are positive examples of a conjunction of two attributes. Haygood and Bourne (1965) recommended duplicating stimuli to ensure roughly equal numbers of positive or negative examples because of the effect of the proportion of positive examples on learning rates (Hovland & Weiss, 1953). The four negative examples of the disjunction were duplicated in the disjunction conditions, and the four positive examples were duplicated in the conjunction conditions to produce a total of 20 stimuli in all conditions.

The set of stimuli used in the conjunction conditions followed the rule "size = small and color = yellow." In the conjunctive condition, one positive example was a photograph of a child stretching a small, yellow balloon. One negative example was a photograph of an adult stretching a large, yellow balloon. The stimuli in the disjunction conditions follow the rule "age = adult or action = stretching." In the disjunctive condition, one positive example was a photograph of a child stretching a large, yellow balloon. One negative example was a photograph of a child dipping a small, yellow balloon in water.

Procedures. Subjects read either the alpha or inflate instructions. Both sets of instructions mention that the photographs differed in only four aspects (the size and color of the balloon, the age of the actor, and the action the actor was performing). The alpha and inflate instructions differed only in one line ("predict whether the page is an example of an 'alpha'" as opposed to "predict whether the balloon will be inflated").

Subjects were shown a page from the photo album and asked to make a prediction. Then the page was turned over and the subject saw the correct prediction. Next, the subject was presented with another card. This process was repeated until the subjects were able to predict correctly on 6 consecutive trials. The number of the last trial on which the subject made an error was recorded. The pages were presented in a random order, subject to the constraint that the first page was always a positive example. If the subject exhausted all 20 pages, the pages were shuffled and the training was repeated until the subject responded properly on 6 consecutive trials or until 50 pages were presented. If the subject did not obtain the correct answer after 50 trials, the last error is considered to have been made on Trial 50.

Note that subjects in the alpha-disjunction and inflate-disjunction conditions see the exact same stimuli. The only difference is one line in the instructions and the nature of the feedback (the words *Alpha* or *Not Alpha* as opposed to a photograph of an inflated or uninflated balloon). Similarly, the subjects in the alpha-conjunction and inflate-conjunction conditions see the exact same stimuli.

Results

The results of this experiment (see Figure 1) confirmed the predictions. Figure 1 illustrates that the learning task is influenced by prior theory. This effect is so strong that it dominates the well-known finding that conjunctive concepts are easier to learn than disjunctive concepts. The interaction between the learning task and the logical form of the concept to be acquired is significant at the .01 level, $F(1, 84) = 22.07$, $MS_e = 264.0$. However, neither main effect is significant.

Analysis of the data with the Tukey honestly significant difference (HSD) test confirmed the three predictions. The results are significant at the .05 level (Critical difference [C.diff] = 11.8). First, subjects in the alpha-conjunction condition required significantly fewer trials than those in the alpha-disjunction category (18.0 vs. 30.8). Second, the inflate-disjunction subjects required significantly fewer trials than the inflate-conjunction subjects (9.4 vs. 29.1). Third, the inflate-disjunction subjects required significantly fewer trials than the alpha-disjunction subjects (9.4 vs. 30.8).

Discussion

The findings provide support for the hypothesis that concepts consistent with prior knowledge require fewer examples to learn accurately than concepts that are not consistent with prior knowledge. The result is especially important because it demonstrates that prior knowledge dominates the commonly accepted finding that disjunctive concepts are more difficult to learn than conjunctive concepts. Cue salience (Bower & Trabasso, 1968) cannot account for the finding that subjects who read the inflate instructions found disjunctions easier

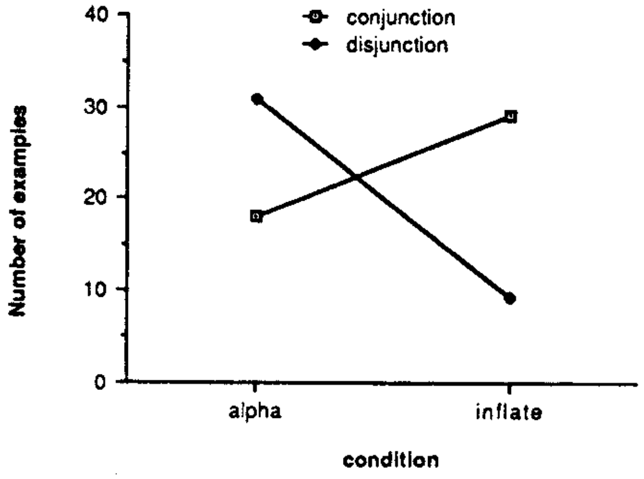


Figure 1. The ease of acquiring disjunctive and conjunctive concepts as a function of the instructions. (The disjunctive relationship is consistent with prior knowledge on the ease of inflating balloons, whereas the conjunctive relationship violates these beliefs.)

than conjunctions. Otherwise, subjects who read the alpha instructions would be expected to exhibit similar preferences. This experiment raises important issues for empirical learning methods, including neural network models (Rumelhart et al., 1986). The learning rules of purely empirical methods do not take into account the learner's prior knowledge. Any difference in learning rates between subjects who read the inflate instructions and those who read the alpha instructions must be accounted for by a difference in the nature of the prior knowledge that can be applied to the task.

The experiment also points out inadequacies of current explanation-based learning methods. Explanation-based learning assumes that the background theory is sufficiently strong to prove why a particular outcome occurred. Purely explanation-based approaches to learning predict that subjects would be capable of learning from a single example. This single-trial learning merely summarizes a deductive proof based on the background knowledge of the subjects. In contrast, it does not appear that the background knowledge of the subjects is sufficiently strong to create such a proof. Instead, the subjects' background knowledge seems to be able to identify what factors of the situation might influence the outcome of an attempt to inflate a balloon. However, subjects needed several examples to determine which of these factors were relevant and whether the factors were necessary or sufficient.

In the next section, a method of combining empirical and explanation-based learning that makes use of this weaker sort of domain knowledge represented as an influence theory is introduced. A simple computational model capable of explaining the learning rates observed in Experiment 1 is proposed. Next, additional simulations are run under a variety of different conditions. Additional experiments are described that test the predictions made by the model.

Explanation-Based Learning With an Influence Theory

To develop a computation model of the learning task, the assumption of explanation-based learning that the domain theory be complete and correct must be relaxed. The full, incomplete, and incorrect domain-theory problem in explanation-based learning (Rajamoney & DeJong, 1987) is not addressed. Instead, I consider an influence theory, a particular type of incomplete theory. In such a theory, the influence of several factors is known, but the domain theory does not specify a systematic means of combining the factors. In addition, it is not assumed that the domain theory identifies all of the influential factors. Loosening these constraints on the domain theory allows prior knowledge to be more widely applicable. In particular, it is necessary to relax these constraints to model the type of prior knowledge used by the subjects in Experiment 1.

PostHoc uses an influence theory to propose hypotheses that are then tested against further data. The influence theory is also used to revise hypotheses that fail to make accurate predictions. PostHoc is also capable of performing classification tasks for which its background knowledge is irrelevant.

Representation of Training and Test Examples

An example in PostHoc consists of a set of attributes and a classification. Each attribute is a pair of an attribute name (e.g., age) and an attribute value (e.g., adult). A classification can be thought of as an outcome (e.g., inflate) or category-membership information (e.g., alpha). For example, an adult successfully inflating a small, yellow balloon that had been stretched is represented as:

size = small color = yellow age = adult act = stretch $\in$ inflate.

A large, purple balloon that had been dipped in water by a child that is not an example of an alpha is represented as:

size = large color = purple age = child act = dip $\notin$ alpha.

Representation and Use of Hypotheses

PostHoc maintains a single hypothesis consisting of a disjunctive normal form description (i.e., disjunction of conjunctions) of a concept and a prediction. For example, the following represents the hypothesis that a child can inflate a stretched balloon or an adult can inflate any balloon:

(age = child $\wedge$ act = stretch) $\vee$ age = adult $\rightarrow$ inflate

Note that to avoid confusion the symbol $\rightarrow$ is used in hypotheses, while $\in$ is used to denote that an instance is a member of a class.

Influence Theories

An influence theory consists of two components. First, it has a set of influences. An influence consists of an influence type (either easier or harder), an outcome (e.g., inflate), and a factor that influences the outcome (e.g., more elastic). Second, an influence theory has a set of inference rules that describe when an influence is present in an example.

To simulate the knowledge of subjects in the previous experiment, the two influences in Appendix A are used. These influences state that it is easier for a strong actor to inflate a balloon, and that it is easier to inflate a more elastic balloon.

The inference rules determine when an influence is present in a training example. The inference rules used to simulate the knowledge of the subjects are also shown in Appendix A. These rules state that stretching an object makes the object more elastic, that older actors tend to be stronger actors, and that adults are old.

Note that the attributes used to represent the training examples are the only attributes that are permitted in the hypotheses. The influence theory can be used to generate a hypothesis, but a factor of the influence theory cannot be used as an attribute in a hypothesis. Rather, the learning procedure may suggest including those attributes of training examples whose presence indicates the presence of a factor from the influence theory.

Learning Task

PostHoc is an incremental learning model that maintains a single hypothesis (Levine, 1966, 1967). The current hypothesis is revised only when it makes an incorrect classification. The learning task is summarized as follows:

Given: a set of training examples
an influence theory (optional)
Create: a hypothesis that classifies examples.

The influence theory is optional because the learning system must operate when there is no prior knowledge or when the prior knowledge does not apply to the current learning task.

PostHoc is intended to model the interaction between prior knowledge and logical form by accounting qualitatively for differences in human learning rates and differences in human hypothesis-selection biases on different tasks. The model is designed to predict that one learning task requires significantly more trials than another task as a function of the prior knowledge and logical form of the hypothesis. Although it does make quantitative predictions on the number of training examples, PostHoc is evaluated only on its ability to partially order the difficulty of learning tasks. PostHoc is intended as the simplest representative of a class of models that can account for how prior knowledge constrains the learning process. PostHoc is not intended as a complete model of the tasks because it does not make use of additional information that human learners have (e.g., perceptual salience of cues; Bower & Trabasso, 1968). Furthermore, each training example in PostHoc is represented as a set of potentially relevant attributes. Although the instructions in the experiments tell the subjects which attributes are potentially relevant, the subjects perform an additional task by determining the values of these attributes from the photographs. Because subjects perform this additional task, as well as perceive other tasks (e.g., perceive facial expressions of the actor in the photographs), PostHoc is not solving as complex a learning task as the subjects. Nonetheless, it is still possible for PostHoc to make predictions about the relative difficulty of learning tasks because these additional complications are held constant for each group of subjects.

PostHoc

PostHoc is an incremental, hill-climbing model of human learning of the type advocated by Langley, Gennari, and Iba (1987). PostHoc is implemented as a simple production system. When the current hypothesis makes an error (or there is no current hypothesis), a set of productions produces a new hypothesis. The productions examine the current hypothesis, the current training example, and the influence theory. There are three sets of productions. One set creates an initial hypothesis when the first positive example is encountered. The second production set deals with errors of omission in which a positive example is falsely classified as a negative example. This production set makes the hypothesis more general. The final production set deals with errors of commission in which a negative example is falsely classified as a positive example. This production set makes the hypothesis more specific. Within each production set, the productions are ordered by priority.

Initializing hypotheses. Two productions used to initialize a hypothesis are shown in Appendix B. The first production (I1) determines if there are attributes of the example that would indicate the presence of a factor that influences the outcome of a positive example. This is accomplished by chaining backward from the influence rules, which indicate that a certain outcome (e.g., inflating a balloon) is easier when a certain factor is present. The presence of a factor is verified by chaining backward to find attribute values that are indicative of an influential factor. For example, if the initial positive example is an adult successfully inflating a large, yellow balloon that had been stretched:

$$\text{color} = \text{yellow size} = \text{large act} = \text{stretch age} = \text{adult} \in \text{inflate},$$

PostHoc might try to establish that the strength of the actor is an influential factor. The fact that strength is an influential factor can be established by showing that the actor is strong. The fact that the actor is strong can be verified because the example indicates that the actor is adult. The initial hypothesis is that adults can inflate balloons:

$$\text{age} = \text{adult} \rightarrow \text{inflate}.$$

In this example, there is more than one influence present. When this occurs, one influence is selected at random from the set of applicable influences. Given the balloon-influence theory, an alternative hypothesis is that stretching the balloon results in the balloon being inflated. However, rather than keeping track of the alternative hypotheses, PostHoc selects one. If this selection turns out to be incorrect, later examples will cause errors of omission or errors of commission and force the revision of the hypothesis.

The second production (I2) in this set initializes the hypothesis to a conjunction of the attributes of the first positive example. This occurs if there are no influences present that would account for the outcome. This is true for modeling alpha-instructions subjects because there are no known factors that influence whether or not something is classified as an alpha.

Errors of omission. Three productions to correct errors of omission are also shown in Appendix B. The first production (O1) applies only if the current hypothesis is consistent with the influence theory and the attributes of the example indicate the presence of an additional factor. This additional factor is assumed to be a multiple sufficient cause (Kelley, 1971). The new hypothesis created is a disjunction of the old hypothesis and a conjunction of the attributes indicative of the additional factor.

For example, if the current hypothesis is:

$$\text{age} = \text{adult} \rightarrow \text{inflate},$$

and the current training example is:

$$\text{size} = \text{large act} = \text{stretch age} = \text{child color} = \text{yellow} \in \text{inflate},$$

then the current hypothesis will cause an error of omission because the hypothesis fails to predict the correct outcome. Because there are attributes indicative of an additional influence (more elastic), Production O1 will create a new hypothesis that represents a multiple sufficient cause:

$$\text{act} = \text{stretch} \lor \text{age} = \text{adult} \rightarrow \text{inflate}.$$

The second production (O2) is a variant of the wholist strategy in Bruner et al. (1956), which drops a single attribute that differs between the misclassified example and the hypothesis. In case of ties, one is selected at random. For example, if the current hypothesis is:

$$\text{color} = \text{yellow} \land \text{size} = \text{large} \land \text{act} = \text{dip} \land \text{age} = \text{adult} \rightarrow \text{inflate},$$

and the current training example is:

$$\text{color} = \text{yellow} \quad \text{size} = \text{small} \quad \text{act} = \text{dip} \quad \text{age} = \text{adult} \in \text{inflate},$$

then Production O2 will drop the attribute that differs between the example and the current hypothesis to form the new hypothesis:

$$\text{color} = \text{yellow} \land \text{act} = \text{dip} \land \text{age} = \text{adult} \rightarrow \text{inflate}.$$

The third error of omission production (O3) forms a disjunction of the current hypothesis and a random attribute of the example when the current hypothesis is consistent with background knowledge and when conjunctive hypotheses have been ruled out. For example, if the current hypothesis is:

$$(\text{age} = \text{child} \land \text{act} = \text{stretch}) \lor \text{size} = \text{small} \rightarrow \text{inflate},$$

and the current training example is:

$$\text{color} = \text{yellow} \quad \text{size} = \text{large} \quad \text{act} = \text{dip} \quad \text{age} = \text{child} \in \text{inflate},$$

then Production O3 will create a new disjunction of the current hypothesis and a randomly selected attribute of the current example:

$$\text{age} = \text{child} \lor (\text{age} = \text{child} \land \text{act} = \text{stretch}) \lor \text{size} = \text{small} \rightarrow \text{inflate}.$$

The simplification of the hypothesis affects the form of the hypothesis to make it more concise and understandable, but does not affect the accuracy of the hypothesis. It consists of several simplification rules, for example, $X \lor (X \land Y) \equiv X$. The hypothesis from the previous example may be simplified to:

$$\text{age} = \text{child} \lor \text{size} = \text{small} \rightarrow \text{inflate}.$$

Errors of commission. Two productions to revise the hypothesis when an error of commission is detected are shown in Appendix B. The first production (C1) adds a multiple necessary cause to the hypothesis (Kelley, 1971). For example, if the hypothesis is that all adults can inflate balloons:

$$\text{age} = \text{adult} \rightarrow \text{inflate},$$

an error will occur on an example of an adult not inflating a large, yellow balloon that has been dipped in water:

$$\text{size} = \text{large} \quad \text{color} = \text{yellow} \quad \text{act} = \text{dip} \quad \text{age} = \text{adult} \notin \text{inflate}.$$

The hypothesis is modified by finding an additional factor not present in the example that could affect the outcome (e.g., stretching the balloon) and asserting that the attributes indicative of this factor are necessary to inflate the balloon. The new hypothesis consists of a single conjunction representing the prediction that adults can inflate only balloons that have been stretched:

$$\text{act} = \text{stretch} \land \text{age} = \text{adult} \rightarrow \text{inflate}.$$

The second error of commission production (C2) specializes a hypothesis by adding additional attributes to each true conjunct. For example, if the current hypothesis is that yellow balloons or purple balloons that had been dipped in water can be inflated:

$$\text{color} = \text{yellow} \lor (\text{color} = \text{purple} \land \text{act} = \text{dip}) \rightarrow \text{inflate},$$

and the following example is encountered:

$$\text{size} = \text{small} \quad \text{color} = \text{yellow} \quad \text{age} = \text{child} \quad \text{act} = \text{dip} \notin \text{inflate},$$

then an incorrect prediction will be made because color = yellow is true. This hypothesis is modified by finding the inverse of an attribute of the example (e.g., size) and asserting that this is necessary when the color is yellow:

$$(\text{color} = \text{yellow} \land \text{size} = \text{large}) \lor$$
$$(\text{color} = \text{purple} \land \text{act} = \text{dip}) \rightarrow \text{inflate}.$$

If this change turns out to be incorrect, later examples will force further refinement of the hypothesis.

An example of PostHoc acquiring a predictive rule will help to clarify how hypotheses are formed and revised. Here I consider how PostHoc operates with an incomplete theory. In this incomplete theory, there is only one influence present:

(easier more elastic inflate),

and the data presented to PostHoc are consistent with the rule that adults can inflate any balloon or anyone can inflate a balloon that has been stretched:

$$\text{age} = \text{adult} \lor \text{act} = \text{stretch} \rightarrow \text{inflate}.$$

This example illustrates how both the analytical and empirical components cooperate to create a hypothesis. The first example is of a balloon being inflated:

$$\text{color} = \text{purple} \quad \text{size} = \text{small} \quad \text{act} = \text{stretch} \quad \text{age} = \text{child} \in \text{inflate}.$$

Production I1 finds an influence present, and the initial hypothesis is that all balloons that have been stretched can be inflated:

$$\text{act} = \text{stretch} \rightarrow \text{inflate}.$$

This hypothesis is consistent with several more examples. Finally, an error of omission occurs when PostHoc predicts that a balloon will not be inflated, but it is:

color = yellow size = large act = dip age = adult ∈ inflate.

Production O3 randomly selects one attribute and makes a new disjunction of the old hypothesis and the attribute. This attribute is dipping the balloon in water. The new hypothesis states that stretching a balloon or dipping a balloon in water are predictive of the balloon being inflated:

$$\text{act} = \text{stretch} \vee \text{act} = \text{dip} \rightarrow \text{inflate}.$$

This hypothesis causes an error of commission when an example is erroneously predicted to result in a successful inflation of a balloon:

color = yellow size = small act = dip age = child ∉ inflate.

Production C2 specializes the term of the disjunction that indicates that dipping a balloon in water is predictive of the balloon being inflated. The inverse of the age is selected as an additional necessary condition for this conjunct. The new hypothesis is:

$$(\text{age} = \text{adult} \wedge \text{act} = \text{dip}) \vee \text{act} = \text{stretch} \rightarrow \text{inflate}.$$

This hypothesis is consistent with the rest of the training set because only two kinds of actions are present in these data.

Simulation 1

Procedure

PostHoc was run on each of the four conditions from Experiment 1. Like Experiment 1, this simulation follows a 2 (concept form [conjunction vs. disjunctive]) × 2 (instruction set [alpha vs. inflate]) between-subjects design. The simulations were run in Common Lisp on an Apple Macintosh II computer. The stimuli and procedures described for Experiment 1 were adapted as necessary to account for the difference between a computer and human "subjects." Training examples were prepared by defining four attributes for each page of the photo album. The balloon-influence theory displayed in Appendix A is used to represent the prior knowledge of the subjects who read the inflate instructions. No influence theory was used when modeling the alpha conditions. No change to PostHoc is necessary to model the alpha conditions. However, because the information needed by the productions that make use of the influence is not present, none of these productions will be used.

PostHoc was run 200 times on different random orders of training examples for each of the four conditions. As in Experiment 1, PostHoc was run until six consecutive examples were classified correctly. The last trial on which PostHoc made an error was recorded for each simulation. Both the ordering of examples and the alternative attributes randomly selected by the productions account for differences in training times in different simulations of the same condition.

Results

The results of this simulation are similar to those of Experiment 1. The interaction between the learning task and the logical form of the concept to be acquired is significant at the .01 level, $F(1, 793) = 132.9$, $MS_e = 78.4$. Analysis of the data with the Tukey HSD test confirms the same three predictions from Experiment 1. The results are significant at the .01 level (C.diff = 2.7). First, the alpha-conjunction category required significantly fewer trials than the alpha-disjunction category (6.85 vs. 18.80). Second, the inflate-disjunction category required significantly fewer trials than the inflate-conjunction category (3.97 vs. 16.52). Third, the inflate-disjunction category required significantly fewer trials than the alpha-disjunction category (3.97 vs. 18.80).

Discussion

Inconsistent conjunctive concepts (e.g., the inflate-conjunction condition) are more difficult for PostHoc to acquire than conjunctive concepts without an influence theory (e.g., the alpha-conjunction condition) because the initial hypothesis typically includes irrelevant attributes (e.g., age = adult) predicted to be relevant by the influence theory. These irrelevant attributes must be dropped from the hypothesis when they cause errors.

Simulation 1 demonstrates that PostHoc can account for the differences in learning rates in Experiment 1 as a function of the logical form of the concept and the existence of relevant prior knowledge. Next, four more simulations are presented that make predictions about learning rates of human subjects, the type of stimulus information that subjects process during learning, the types of hypotheses that subjects create, and the effect of incomplete and incorrect knowledge on learning rates. These simulations are followed by experiments in which the predictions of PostHoc are tested on human subjects.

Simulation 2

Procedure

Experiment 1 and Simulation 1 demonstrate that relevant background knowledge makes a consistent disjunctive concept easier to learn than the same disjunctive concept when no background knowledge is relevant. In this next simulation, the learning rate of PostHoc on a consistent conjunctive concept is compared with the learning rate on the identical conjunctive concept when no background knowledge is relevant. The stimuli in the experiment follow the rule "age = adult and action = stretching" (i.e., adults can inflate only balloons that have been stretched). As in Simulation 1, there are 16 unique stimuli, and a total of 20 training examples were constructed by duplicating the 4 positive examples. PostHoc is run with the influence theory in Table 1 and with no influence theory. One hundred random orders of training examples were simulated in each of the conditions. As in Simulation 1, PostHoc was run until 6 consecutive examples were classified correctly and the last trial on which an error was made was recorded.

Results and Discussion

In this simulation, the conjunctive concept was learned more quickly when the relevant influence theory is present (3.6 vs. 5.5), $t(198) = 8.42$, $SE = 0.328$, $p < .01$. This simulation clearly demonstrates that prior knowledge can facilitate the learning of more than one logical form. Furthermore, the fact that the same influence theory was used in both simulations shows that more than one concept can be consistent with the same influence theory. In Simulation 1, the balloon-influence theory in Appendix A was shown to facilitate learning the disjunctive rule "age = adult or action = stretching." Here this same knowledge facilitated learning the conjunctive rule "age = adult and action = stretching." The domain theory used by prior work in explanation-based learning cannot exhibit this flexibility because both of these concepts cannot be in the deductive closure of the same domain theory. However, the influence theory used by PostHoc allows it to use prior knowledge to facilitate learning either concept.

Note that in Simulation 1, the influence theory hindered learning a conjunctive concept, whereas the same influence theory facilitated learning a conjunctive concept in Simulation 2. The difference is that in Simulation 1 the conjunctive concept "size = small and color = yellow" was not consistent with the influence theory, but in Simulation 2 the conjunctive concept "age = adult and action = stretching" was consistent.

In PostHoc, there are three types of relationships between a concept and the background knowledge. First, the background knowledge can be neutral in that it does not provide support for any hypothesis. This occurs when the influence theory contains no influences for the concept being learned (e.g., the alpha conditions of Simulation 1). In this case, PostHoc uses only those productions that do not require an influence theory (I2, O2, O3, and C2). Second, the concept to be learned may be consistent with the background knowledge. In this case, PostHoc uses only those productions that refer to the influence theory (I1, O1, C1). Finally, the concept to be learned may be inconsistent with the background knowledge. When this is true, an initial subset of the training examples may be consistent. Therefore, PostHoc may start to use productions that make use of the influence theory. The hypotheses formed by these productions will be inconsistent with later examples, and PostHoc will eventually resort to those productions that do not reference the influence theory.

Figure 2 shows the average number of times (N = 100) each production was used when learning a disjunctive and conjunctive concept for each of the three relationships. In each case, the concept had two relevant attributes and two irrelevant attributes. Note that the neutral and consistent cases use only a subset of the productions, whereas the inconsistent case requires all productions.

Simulation 3

Procedure

Simulation 3 is designed to help to explain why PostHoc requires fewer examples to learn when hypotheses are consistent with an

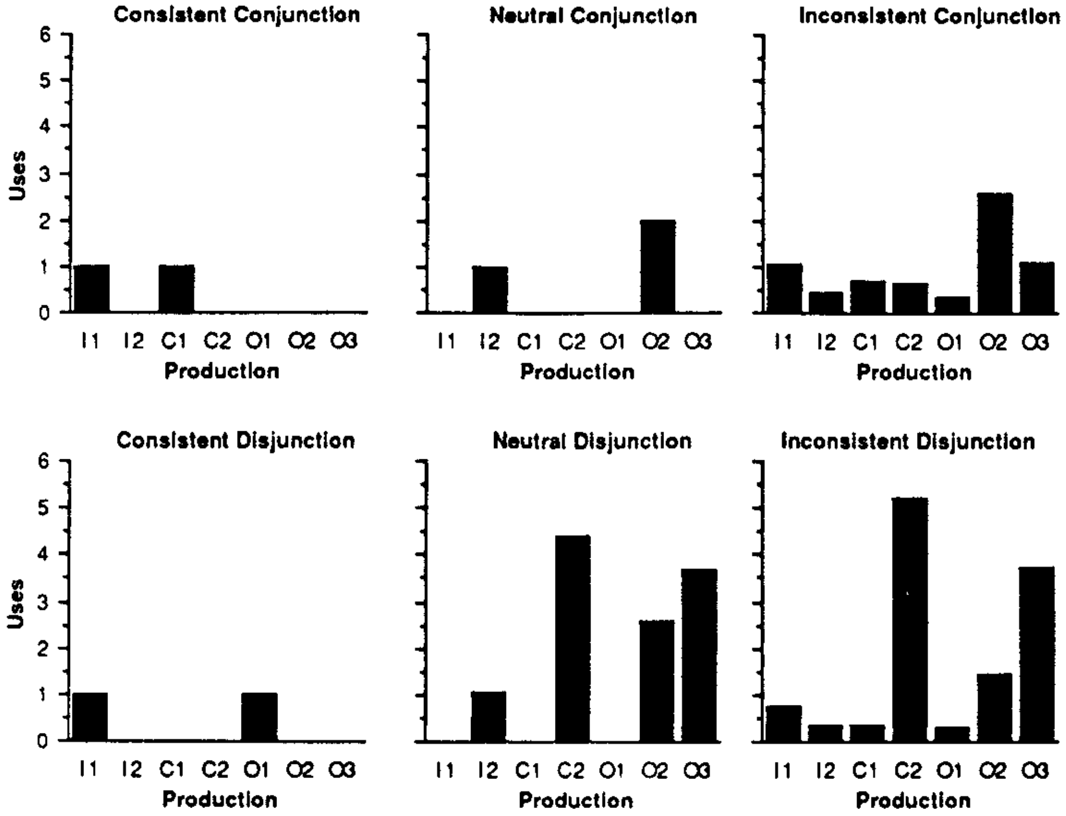


Figure 2. The productions used by PostHoc learning disjunctive and conjunctive concepts that are consistent, neutral, or inconsistent with prior knowledge.

influence theory. The claim tested is that when there is correct prior knowledge a smaller hypothesis space is searched. In this situation, PostHoc can ignore irrelevant attributes (i.e., attributes not indicative of known influences). Simulation 3 investigates which attributes are considered when categorizing examples. It is assumed that every attribute that is in the current hypothesis is considered during categorization. In addition, there is a sampling probability that an attribute not in the current hypothesis will be considered. It is also assumed that after a categorization error is made all attributes are considered when forming a new hypothesis. There are two reasons for considering attributes not part of the hypothesis during categorization. First, in a similar experiment using human subjects (Experiment 3), subjects reported examining some attributes out of curiosity. Second, occasionally considering attributes not in the hypothesis introduces some variability in PostHoc and enables analyses of the data. In three simulations, values of 0.1, 0.5, and 0.9 were used as the probability that an attribute not in the hypothesis will be considered by PostHoc during classification.

The hypothesis tested is that PostHoc will consider fewer irrelevant attributes when there is an influence theory (and when the data are consistent with the influence) than when there is no influence theory. On each trial, starting with the second trial, the proportion of irrelevant attributes considered is recorded. This proportion is calculated by dividing the number of irrelevant attributes considered by the total number of attributes considered. Note that with a correct influence theory an irrelevant attribute is considered only because there is a sampling probability that an attribute not in the hypothesis is considered. Without an influence theory, an irrelevant attribute can be considered because it appears in a hypothesis or because the attribute is sampled randomly.

The disjunctive concept "age = adult or action = stretching" is tested both with and without an influence theory. One hundred trials of each condition are run for each level of sampling (i.e., 0.1, 0.5, and 0.9). Twenty training examples are generated in the same manner as the disjunctive conditions of Simulation 1. However, in the current simulations, the same randomly selected order is used for every simulation to eliminate the effect of the ordering of training examples on the creation and revision of hypotheses. The exact same order of examples was also used in Experiment 3.

Each simulation follows a 2 × 19 mixed design with one between-subjects factor (knowledge state [no influence theory vs. consistent influence theory]) and one within-subjects factor (trial [number of the learning trial ranging from 2 to 20]). On each trial, the proportion of irrelevant attributes was measured.

Results and Discussion

Figure 3 illustrates the results of the three simulations; one panel shows the result for each sampling probability. When there was a consistent influence theory, the proportion of irrelevant attributes was less than or equal to the proportion of irrelevant attributes with no influence theory. When there was no influence theory, the mean proportion of irrelevant attributes always started at 0.5 and over the 20 trials declined to varying degrees depending on the sampling probability. With the influence theory (i.e., the inflate instructions), no irrelevant attributes are in any hypothesis, and the proportion of irrelevant attributes considered is equal to the proportion of irrelevant attributes selected randomly. Without the influence theory (i.e., the alpha instructions), the proportion of irrelevant features considered (Alpha Total) is the sum of the irrelevant features in the hypothesis (Alpha Hyp) and the irrelevant features selected randomly (Alpha Rand).

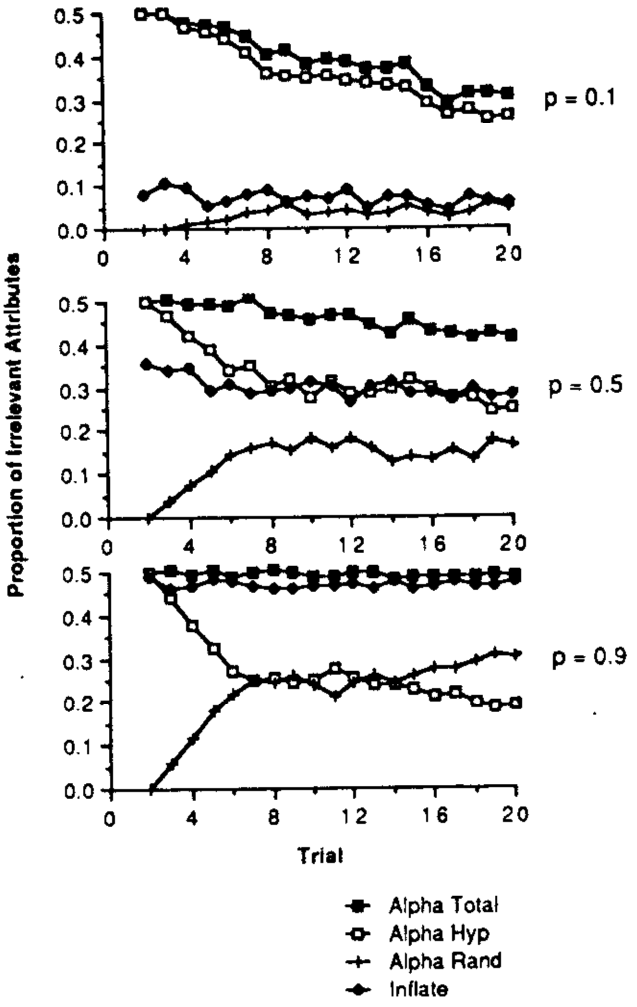


Figure 3. The mean proportion of irrelevant attributes selected by PostHoc simulating the inflate and alpha instructions. (This proportion is calculated by dividing the number of irrelevant attributes considered by the total number of attributes considered. The three graphs plot the data when the probability of randomly considering an attribute was 0.1 [upper], 0.5 [middle], and 0.9 [lower]. Inflate is the total proportion of irrelevant attributes considered in the alpha condition and is identical to the proportion of irrelevant attributes selected randomly. Alpha Rand is the proportion of irrelevant features selected randomly in the alpha condition. Alpha Hyp is the proportion of irrelevant features in the hypothesis. Alpha Total is the total proportion of irrelevant attributes considered in the alpha condition.)

An arcsine transformation was applied to the proportion data and an analysis of variance shows that, as expected, there is a main effect for knowledge state, $F(1, 198) = 288.22$, $MS_e = 179.4$, $p < .0001$. In addition, the main effect of trial was significant, $F(18, 3{,}564) = 2.19$, $MS_e = 0.65$, $p < .01$.

Figure 3 also plots data when the probability of considering a feature not in the hypothesis is 0.1 (upper) and 0.9 (lower). As this probability approaches 1, the total proportion of irrelevant attributes that PostHoc considers when it simulates inflate instructions approaches the proportion of irrelevant attributes that it considers when it simulates alpha instructions. However, even when this probability is 0.9, there is a main effect for knowledge state, $F(1, 198) = 59.08$, $MS_e = 2.60$, $p < .0001$.

This simulation shows that PostHoc can ignore irrelevant attributes when hypotheses are consistent with the influence theory. The ability of human learners to ignore irrelevant attributes will be tested in Experiment 3.

Simulation 4

Procedure

Simulations 1 and 2 show that PostHoc learns more rapidly when hypotheses are consistent with its influence theory. Simulation 3 helps to explain this finding by demonstrating that PostHoc need not attend to some attributes during learning when hypotheses are consistent with its influence theory.

In this simulation, I elaborate on this finding by demonstrating that the prior knowledge of PostHoc also affects the hypotheses it forms. In particular, whenever possible, a hypothesis will only include attributes indicative of factors in the influence theory. This simulation is a variation of the redundant, relevant cue experiments (Bower & Trabasso, 1968). In a redundant, relevant cue experiment, at training time, a subject performs a classification task in which the data are consistent with more than one hypothesis. For example, in this simulation the balloon that is dipped in water is always purple and a balloon that is stretched is always yellow. The yellow, stretched balloons are the balloons that receive positive feedback (i.e., can be inflated or are an instance of alpha). There are multiple hypotheses consistent with the data (e.g., all yellow balloons are instances of alpha or all stretched balloons are instances of alpha). A total of eight such training examples are constructed. After the system is able to perform accurately on six consecutive examples, the hypothesis created by the system is recorded. The hypotheses created by PostHoc with the influence theory in Appendix A are compared with those produced by PostHoc on the same training set without an influence theory. Two hundred simulations of each condition were run with training examples in randomly selected orders.

Results and Discussion

An analysis of PostHoc productions indicates that with an influence theory PostHoc will always create the hypothesis "act = stretch → inflate." Without the influence theory, the hypothesis "(color = yellow ∧ act = stretch) → alpha" will always be created. This analysis was substantiated by the simulation in which it was found that, with an influence theory, the only relevant variable used was the action. Without an influence theory, both color and action are in the final hypothesis.

Both hypotheses created by PostHoc are consistent with the data. However, the hypothesis will also be consistent with the influence theory if one is applicable. The results of the simulation without an influence theory differ from the findings of redundant, relevant cue experiments on human subjects. Most subjects in the Bower and Trabasso (1968) experiment favored one-attribute discriminations (i.e., either "color = yellow" or "action = stretched") to conjunctions.

An examination of PostHoc's productions reveals that the only means of learning one-attribute discriminations is by dropping an attribute from a conjunction of two attributes. An extension to PostHoc to more faithfully model the empirical findings would contain an additional initialization production to create one-attribute discriminations. This extension has not yet been implemented because the focus of PostHoc has been to account for the influence of a particular type of prior knowledge on learning.

Simulation 5

Procedure

Here I explore the influence of the completeness and correctness of the influence theory on the learning rate. PostHoc was run with five variations of the balloon-influence theory: consistent (the complete and correct influence theory consisting of two influences), incomplete (one of the two influences was deleted from the complete theory), neutral (the entire influence theory was deleted), partially inconsistent (the influence theory consisted of one correct and one incorrect influence [yellow balloons are easier to inflate]), and inconsistent (two incorrect influences were used). The goal of learning in each condition is to acquire the rule that adults can inflate any balloon or anyone can inflate a balloon that has been stretched. Each condition was run 128 times, and the number of the last trial on which PostHoc misclassified an example was recorded.

Results and Discussion

In order, the mean number of trials required to converge on an accurate hypothesis for the consistent, incomplete, neutral, partially inconsistent, and inconsistent conditions are 3.9, 12.8, 18.4, 15.1, and 20.1, respectively. The quality of the domain theory has a significant effect on the number of trials required to acquire an accurate concept, $F(4, 635) = 72.2$, $MS_e = 100.7$.

When there is no influence theory, only the empirical productions of PostHoc are used to form a hypothesis. When there is an incomplete influence theory, the empirical and analytical productions cooperate to produce a hypothesis. When there is a complete influence theory, only the analytical productions are used. With the incorrect influence theory, the analytical productions usually create incorrect hypotheses that are revised by the empirical hypotheses. The analytical productions are not used if the initial examples are not consistent with the incorrect influence theory. For example, if the first positive training example contains a large, purple balloon, then there is no initial hypothesis consistent with the influence theory, and an empirical production initializes the hypotheses. The inconsistent theory is the most difficult, because the initial hypothesis often involves irrelevant features that must later be deleted.

The most interesting result of this simulation is that PostHoc with the partially inconsistent theory ($M = 15.1$) takes fewer trials than PostHoc with no theory ($M = 18.4$). Analysis of the data with the Tukey HSD reveals that the difference in learning rates is significant (C.diff = 3.2, $p < .01$). The difference between these two conditions is partially accounted for by the fact that it is more likely that the correct rather than incorrect influence will be chosen to initialize the hypothesis because the correct influence is present in more of the positive examples (100%) than the incorrect influence (50%). As a result, in 75% of the cases, the hypothesis will be initialized correctly with an inconsistent theory.

Experiment 2

In Simulations 2, 3, and 4, several emergent properties of PostHoc's learning algorithm were described. The following three experiments assess whether similar phenomenon are true of human learning. The effects of prior knowledge on learning rates, relevance of attributes, and hypothesis selection are measured in Experiments 2, 3, and 4, respectively.

One result of Experiment 1 is that consistent disjunctive concepts (the inflate-disjunction condition) required fewer trials to learn than neutral disjunctive concepts (the alpha-disjunction condition). A second result of Experiment 1 was that inconsistent conjunctive concepts (the inflate-conjunction condition) required more trials than neutral conjunctive concepts (the alpha-conjunctive condition).

In Experiment 2, the conjunctive concept tested is consistent with the subjects' prior knowledge (age = adult and action = stretching). The learning rate of this consistent conjunctive concept is compared with that of the same conjunctive concept with neutral instructions. The design of Experiment 2 parallels the design of Simulation 2.

Experiment 2 has several goals. First, a prediction made by PostHoc is tested. In particular, Simulation 2 showed that consistent conjunctive concepts require fewer trials than neutral conjunctive concepts. Second, it is hoped that Experiment 2 will show that the subjects' background knowledge provides weak constraints on learning similar to those provided by PostHoc's influence theory. In particular, Experiment 1 assumes the disjunctive concept (age = adult or action = stretching) is consistent with background knowledge, whereas Experiment 2 assumes that the conjunctive concept (age = adult and action = stretching) is consistent. Clearly, these both cannot be deduced from the type of background knowledge required by explanation-based learning. However, both are consistent with the influence theory of explanation-based learning.

Experiment 2 will also serve to rule out an alternative explanation[2] for the results of Experiment 1. In particular, it is possible that there is something about the inflate instructions (but not the alpha instructions) that leads subjects to predict when a balloon will not be inflated. The interpretation of the results of Experiment 1 assumed that the hypothesis learned by subjects can be represented as "if the actor is an adult or the action is stretching a balloon, then the balloon will be inflated." However, this rule is logically equivalent to "if the actor is a child and the action is dipping a balloon in water, then the balloon will not be inflated." If this is the case, then prior knowledge is irrelevant, and the results of Experiment 1 simply indicate that disjunctive concepts are harder to learn than conjunctive concepts. However, if this is the case, one would expect to find that subjects reading the inflate instructions and learning a consistent conjunction (age = adult and act = stretching) would require more trials than those reading the alpha instructions and learning a neutral conjunction. This would occur because the inflate instructions would presumedly lead subjects to learn when a balloon was not inflated. In this case, the rule indicating that a balloon is not inflated is a disjunction: (age = child or act = dipping).

Method

Subjects. The subjects were 54 undergraduates attending the University of California, Irvine, who participated in this experiment to receive extra credit in an introductory psychology course. Subjects were randomly assigned to one of two conditions (alpha or inflate).

Stimuli. The stimuli consisted of pages from a photo album identical to those of Experiment 1. Each page contained a close-up photograph of a balloon that varied in color and size and a photograph of a person doing something to the balloon. However, subjects now received positive feedback if the actor is an adult and the action is stretching a balloon. The stimuli used in the conjunction conditions follow the rule "size = small and color = yellow." One positive example was a photograph of a child stretching a small, yellow balloon. One negative example was a photograph of an adult stretching a large, yellow balloon. Twenty stimuli were constructed by duplicating the four positive examples.

Procedures. The procedure was identical to that in Experiment 1. The instructions read by the two groups differed in only one line ("predict whether the page is an example of an 'alpha'" as opposed to "predict whether the balloon will be inflated"). The number of the last trial on which the subject made an error was recorded.

Results and Discussion

Subjects in the inflate condition learned the concept more rapidly than those in the alpha condition (8.9 vs. 13.8), $t(52) = 2.09$, $SE = 2.39$, $p < .05$.

This experiment provides additional support for the hypotheses that consistency with prior knowledge is a significant influence on the rate of concept acquisition. The experiment also points out that the prior knowledge of the subjects can be used to facilitate the learning of several different hypotheses. This demonstrates that the prior knowledge of subjects is more flexible than the domain theory used by explanation-based learning. Two different hypotheses cannot be deduced from the domain theory of explanation-based learning, but can be consistent with the influence theory of PostHoc. For example, the same influence theory enables PostHoc to model the relative difficulty of learning in Experiments 1 and 2.

Experiment 3

In Simulation 3, it was shown that with a correct influence theory, PostHoc ignores irrelevant attributes. However, when the influence theory is missing, PostHoc initially forms a hypothesis that includes these irrelevant attributes, and then later revises the hypothesis by removing the irrelevant attributes when examples are misclassified.

The goal of this experiment is to test the hypothesis that subjects learning a concept consistent with their background knowledge will attend to a smaller proportion of irrelevant

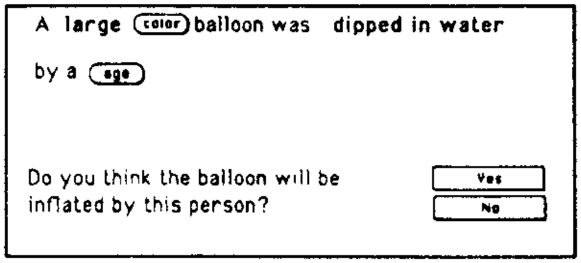
[2] We thank Richard Doyle for pointing out this explanation.

attributes than those learning the identical concept in a context in which their prior knowledge is irrelevant.

It is a relatively simple matter to determine which attributes PostHoc is ignoring during learning. To test this hypothesis on human subjects, different stimuli and procedures were used than in the earlier experiments. Experiment 3 uses verbal descriptions of actions instead of photographs for the stimuli. A program was constructed for an Apple Macintosh II computer to display the verbal descriptions. Each training example presented on the computer screen consisted of a verbal description of an action and a question. Subjects in the inflate condition saw the question "Do you think that the balloon will be inflated by this person?" Subjects in the alpha condition saw the question "Do you think this is an example of an Alpha?"

Each verbal description initially appears as "A ⟨SIZE⟩ ⟨COLOR⟩ balloon was ⟨ACTION⟩ by a ⟨AGE⟩." A subject could request to see a value for any of the attributes by moving a pointer to the attribute name and pressing a button on the mouse. When this was done, the value for the attribute name replaced the attribute name in the verbal stimuli. For example, a subject might point at ⟨COLOR⟩ and then press the mouse button. The effect of this action might be to change the stimuli to "A ⟨SIZE⟩ red balloon was ⟨ACTION⟩ by a ⟨AGE⟩." Next, the subject might point at ⟨ACTION⟩ and click, changing the description to "A ⟨SIZE⟩ red balloon was dipped in water by a ⟨AGE⟩." Figure 4 shows a sample display with the values of two attributes filled in. The attributes selected by the subject were recorded on each trial. Subjects were allowed to select as few or as many attributes on each trial. However, to discourage simply selecting all attributes, subjects had to hit an extra key to confirm that they wanted to see the third and fourth attribute.

A pilot study revealed some interesting information. Subjects in the inflate group asked to see the size attribute much more often than expected. When asked about the need to see this information, a common reply was that large balloons were easier to inflate than small balloons. Subjects in pilot studies for Experiments 1 did not mention size as a possible relevant factor. The difference in stimuli may account for this difference. In Experiment 3, verbal descriptions of actions were used. In Experiments 1, photographs were used. The small balloon in Experiment 1 is a 9-in balloon and the large balloon is a 13-in balloon. One subject in the pilot study of Experiment 3 was later shown the photographs used in Experiment 1, and reported that the small balloons in the photographs were actually medium-sized balloons. Therefore, in the analysis of Experiment 3, the attribute color is considered to be the only irrelevant attribute, and the attribute size, action, and age are considered to be potentially relevant.



A large color balloon was dipped in water

by a age

Do you think the balloon will be
inflated by this person?

Figure 4. An example of the stimuli used in Experiment 3.

The experiment follows a 2 × 20 trial mixed design with one between-subjects factor (instructions [inflate vs. alpha]) and one within-subjects factor (trial [number of the learning trial ranging from 1 to 20]). On each trial, starting with the first trial, the proportion of irrelevant attributes considered is recorded. As in Simulation 3, this proportion is calculated by dividing the number of irrelevant attributes considered by the total number of attributes considered.

Method

Subjects. The subjects were 34 undergraduates attending the University of California, Irvine, who participated in this experiment to receive extra credit in an introductory psychology course. Subjects were randomly assigned to one of the two conditions (alpha or inflate). Seventeen subjects in each condition were tested simultaneously in a room equipped with Apple Macintosh II computers. Each subject worked individually on a separate computer.

Stimuli. The stimuli were verbal descriptions of an action. Twenty stimuli were constructed and shown in the randomly selected order used for Simulation 3. The descriptions varied according to the color of the balloon (red or blue), the size of the balloon (small or large), the age of the actor (adult or child), and the action the actor is performing (either dipping it in water or stretching it). The descriptions were displayed on the screen of an Apple Macintosh II computer in a fixed order. Subjects could interact with the display by asking the computer to show the value of any (or all) attributes. Positive feedback was given for those stimuli whose age is adult or whose action is stretching.

Procedures. The instructions read by the two groups differed in only one line ("determine if this example is an 'alpha'" as opposed to "determine whether the balloon will be inflated successfully by this person"). After reading the instructions, subjects were given the opportunity to practice using a mouse to move the pointer and to press the mouse button to indicate a selection. Next, the subjects repeated a cycle of seeing a template action, asking to view some or all attributes of the action, indicating a prediction by moving a pointer to the word *Yes* or to the word *No*, and pressing a button to answer the question. In the subject selected the correct answer, the computer simply displayed a message to this effect. However, if the subject selected the wrong answer, the computer replaced all attributes with their values, and informed the subject that the answer was incorrect. When the subject finished studying the screen, a button was pressed to go on to the next example. This cycle was repeated 20 times for each subject.

Results and Discussion

Figure 5 displays the mean proportion of irrelevant attributes selected on each trial for the inflate and alpha conditions. The subjects in the inflate condition were less likely to request to see the color attribute than those in the alpha condition. An arcsine transformation was applied to the proportion data, and an analysis of variance revealed that there was a main effect for instructions, $F(1, 32) = 9.89$, $MS_e = 20.73$, $p < .01$. The interaction between instructions and trial and the main effect of trial were not significant.

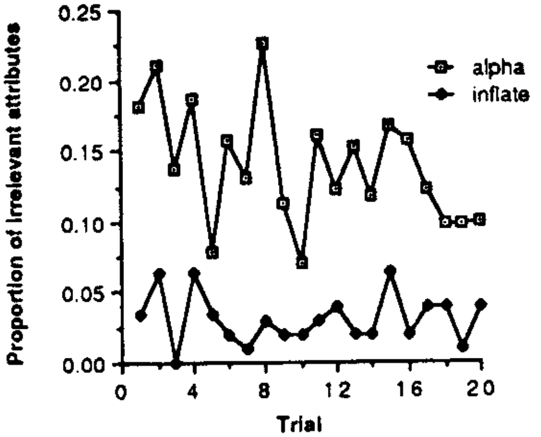


Figure 5. The mean proportion of irrelevant attributes selected by subjects reading the inflate and alpha instructions.

A simple manipulation in the instructions influenced the attributes the subjects attended to during learning. In a classification task with neutral instructions, the subjects have no reason to initially ignore color or any other attribute. However, when the same stimuli are used to make predictions about inflating balloons, subjects are more likely to ignore the color of the balloon. Subjects favored attributes that prior knowledge indicates are likely to influence the ease of inflating a balloon.

Experiment 4

Experiments 1 and 2 suggest that human subjects learn more rapidly when hypotheses are consistent with their prior background knowledge. One explanation for this finding is that hypotheses not consistent with prior knowledge are not considered unless hypotheses consistent with prior knowledge are ruled out. Experiment 4 tests this idea using a redundant relevant cue experiment modeled after Simulation 4. As with Simulation 4, both the action and the color are equally consistent with the feedback on the training data.

An implication of the computational model is that the number of subjects who predict on the basis of the action attribute for the inflate task will be greater than the number of subjects who classify on the basis of this attribute for the alpha task. The reason for this prediction is that the stretching is a factor that is known to influence the inflation of a balloon.

Method

Subjects. The subjects were 54 undergraduates drawn from the same population as those in Experiment 1. Subjects were randomly assigned to one of the two conditions (alpha or inflate).

Stimuli. The stimuli consisted of pages from a photo album identical to those of Experiment 1. Each page contained a close-up photograph of a balloon that varied in color and size and a photograph of a person doing something to the balloon. However, the pages were now constructed so that for the training material the color yellow was paired with stretching and the color purple was paired with dipping in water. In the test, these pairings were reversed so that purple was associated with stretching and yellow with dipping in water. A total of eight training examples and eight test examples were constructed. Subjects received positive feedback only on pages showing balloons that had been stretched.

Procedures. Subjects in the two groups read instructions that differed in only one line ("predict whether the page is an example of an 'alpha'" as opposed to "predict whether the balloon will be inflated"). The training data were presented to subjects in random orders. The subjects were trained on the training set until they were able to accurately classify six pages in a row. Subjects received positive feedback on the photographs that included a person of any age stretching a yellow balloon of any size. Then the subjects entered a test phase in which they predicted the category of test examples without feedback.

Results and Discussion

In this experiment, in the inflate condition, 26 subjects formed hypotheses using only the action attribute, no subjects used only the color attribute, and 2 subjects used a combination of attributes. In the alpha condition, the corresponding numbers were 13, 8, and 7, respectively. Analysis of the data indicates that the hypothesis-selection biases of the subjects in the inflate condition differed from those in the alpha condition, $\chi^2(2, N = 54) = 15.11, p < .01$. The results of this experiment indicate that human subjects favor hypotheses consistent with the data and prior knowledge over those hypotheses consistent with the data but not consistent with prior knowledge.

The hypothesis produced by PostHoc with an influence theory is that most commonly formed by subjects in the inflate condition of Experiment 4. In its current form, PostHoc cannot account for those subjects who produce alternative hypotheses in this condition. In addition, PostHoc does not adequately model the finding that, in the absence of prior knowledge, one-attribute discriminations are preferred to conjunctive descriptions in a redundant, relevant cue experiment (Bower & Trabasso, 1968).

Hypothesizing in Concept Acquisition

To more fully validate PostHoc, it would be desirable to compare the intermediate hypotheses generated by PostHoc with the hypotheses of subjects before converging on the correct hypothesis. Several previous studies have investigated the role of verbal reports of intermediate hypotheses on concept acquisition (e.g., Byers & Davidson, 1967; Dominowski, 1973; Indow, Dewa, & Tadokoro, 1974; Indow & Suzuki, 1972). Strong correlations were found between verbal reports of hypotheses and the subjects' true hypotheses. Furthermore, the verbal reports did not affect factors such as the learning rate or number of errors made.

A variety of verbal-report studies were run in an attempt to have subjects give verbal reports (either oral or written) of their intermediate hypotheses following a methodology similar to these previous studies. However, one modification was necessary to the instructions. In the previous studies of other researchers, the instructions informed the subject of the logical form of the concept to be learned (e.g., a conjunction of two attributes). In our verbal report studies, and all other experiments in this article, the instructions did not include information of the logical form of the concept to be learned. Including this information would affect the observed learning rate (Haygood & Bourne, 1965) interfering with the dependent variable measured (learning rate of conjunctive vs. disjunctive concepts). In these verbal-report studies, requiring verbal reports appeared to make the problem more difficult. As a result, very few subjects were able to complete the learning task. For the subjects who did complete the task, the mean learning rates differed substantially from the earlier experiments reported here. Furthermore, subjects' reports of their hypotheses did not always agree with the prediction made on the next example. For example, requiring verbal reports increased the mean number of trials for learning conjunctive concepts with the alpha instructions to over 30 compared with 13.8 in Experiment 2. Because the results are not interpretable, they are not reported here.

The difference in instructions appears to be responsible for the discrepancy between the verbal-report studies and the earlier findings. If some aspect of the learning process, such as the detection of covariation, is unconscious to some extent, asking for a verbal report may change the nature of the task. Forcing subjects to become more conscious of the processes may make the task more difficult. This hypothesis is consistent with findings by Reber and Lewis (1977), who presented evidence that subjects can learn some rules without having conscious access to the rules. Lewicki (1986) refined this finding by showing that subjects detect correlations and make classifications based on these correlations without being able to verbally report on the correlation. Furthermore, Reber (1976) showed that asking subjects to search for regularities in the data adversely affects the learning rate and accuracy.

Nisbett and Wilson (1977) reported that for some tasks verbal reports on decision-making criteria differ from the criteria that subjects are using. The discrepancies in the verbal-report studies between subjects' reports of their hypotheses and their classifications on subsequent trials appear to be another example of this phenomenon. Other researchers have refined the conditions under which verbal reports of decision-making criteria are likely to be accurate (Ericsson & Simon, 1984; Kraut & Lewis, 1982; Wright & Rip, 1981). More empirical research is needed to clarify the effects of verbal reports on concept learning. One tentative hypothesis is that either requiring a verbal rule or informing subjects that the concept to be learned can be represented as a logical rule of a certain form increases conscious awareness of the learning process and hinders the unconscious detection of covariation.

Brooks (1978) shed some light on the conditions under which verbal reports hinder concept learning. Brooks demonstrated that instructions to form an abstract rule may interfere with the storage of individual instances. This interference with memory storage hinders making future classifications by analogy to stored instances. Although I agree with Brooks that this form of analogical reasoning is common, accounting for the experimental findings in this article with an analogical reasoning model would require explaining how prior knowledge affects the analogical reasoning process along with the storage and retrieval of analogous instances.

Discussion

Experiments 1 and 2 demonstrated that human subjects learn more rapidly when hypotheses are consistent with prior knowledge. PostHoc also learns more rapidly when hypotheses are consistent with an influence theory. In PostHoc, the explanation for the faster learning rate is that it is searching a smaller space of hypotheses (i.e., those consistent with the data and the influence theory). Simulation 3 and Experiment 3 demonstrate that the hypothesis space is reduced by ignoring those attributes deemed irrelevant by prior knowledge. Simulation 4 and Experiment 4 demonstrated the reduced hypothesis space by investigating the types of hypotheses produced when there are multiple hypotheses consistent with the data.

In this article, the prior knowledge of a subject has been shown to influence the learning of predictive relations for actions and their effects. There is some evidence that the influence of prior knowledge is not restricted to this situation. In particular, when subjects are aware of the function of an object, it has been shown that they attend more to attributes of the object that are related to the object's function than to attributes that are predictive of class membership but not related to functionality (Wisniewski, 1989). In addition, Barsalou (1985) showed that the graded structure of goal-oriented categories (e.g., foods not to eat on a diet) is influenced by prior knowledge of ideals (e.g., zero calories).

Currently, PostHoc is limited in several ways. First, it deals only with positive influences. In addition, the influence language does not include information on the potency of each influence. Zelano and Shultz (1989) argued that subjects make use of such information when learning causal relationships.

A second limitation of PostHoc is the inability to learn new influences. A hypothesis that is not supported by an influence theory can be learned, but the influence theory is not currently updated. If the influence theory were updated, then PostHoc could use the knowledge it has acquired in one task to facilitate learning on another task.

A third limitation is that PostHoc does not account for some fundamental categorization effects. For example, PostHoc does not model phenomena such as the effects of typicality (Barsalou, 1985), basic level effects (Corter, Gluck, & Bower, 1988), or the acquisition of concepts that cannot be specified as a collection of necessary and sufficient features (Smith & Medin, 1981). However, background knowledge plays a role in these processes. For example, several experiments have shown (Barsalou, 1985; Murphy & Wisniewski, 1989) that the prior knowledge of a subject affects typicality judgments, but no detailed process has been proposed to account for these findings. Brown (1958) suggested that the knowledge of the learner plays a role in determining the basic level. It would be interesting to explore the role of background knowledge in computational models of these processes.

The simulations and experiments also point out a shortcoming of models of human learning based on the prior work on purely explanation-based methods. It is not likely that the prior knowledge of human subjects can be represented as a set of necessary and sufficient conditions capable of supporting a deductive proof of why particular balloons were inflated.

Rather, the prior knowledge can be applied more flexibly to allow for several concepts to be considered consistent. The influence theory used by PostHoc provides one means of making explanation-based learning more flexible.

In spite of its limitations, the construction of PostHoc has been useful in developing hypotheses about the influence of prior knowledge on human learning. Predictions resulting from simulations have led to experimental findings on human learning. Given the current domain of inflating balloons, it was not possible to test predictions of Simulation 5 concerning the relationship between the quality of the domain knowledge and the learning rate. Testing this hypothesis will first require training subjects on new domain knowledge before the classification task. In contrast, the current experiments rely on knowledge brought to the experiment by the subject.

Conclusion

I have presented experimental evidence that prior knowledge influences the ease of concept acquisition and biases the selection of hypotheses in human learners. Although often overlooked or controlled for, the prior knowledge of the learner may be as influential as the informational structure of the environment in concept learning.

A computational model of this learning task was developed that qualitatively accounts for differences in human learning rates and for hypothesis-selection biases. Predictions of the computational model were tested in additional experiments, and the model's ability to learn with incorrect and incomplete background theories was evaluated.

The ability of human learners to learn relatively quickly and accurately in a wide variety of circumstances is in sharp contrast to current machine-learning algorithms. I hypothesize that this versatility comes from the ability to apply relevant background knowledge to the learning task and the ability to fall back on weaker methods in the absence of this background knowledge. In PostHoc, I have shown how the empirical and analytical learning methods can cooperate in a single framework to learn accurate, predictive relationships. PostHoc learns most quickly with a complete and correct influence theory, but is still able to make use of background knowledge when conditions diverge from this ideal.

References

Anderson, J. R. (1989). A theory of the origins of human knowledge. *Artificial Intelligence, 40,* 313–351.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629–654.

Bower, G., & Trabasso, T. (1968). *Attention in learning: Theory and research.* New York: Wiley.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Brown, R. (1958). How shall a thing be called? *Psychological Review, 65,* 14–21.

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

Byers, J., & Davidson, R. (1967). The role of hypothesizing in concept attainment. *Journal of Verbal Learning and Verbal Behavior, 6,* 595–600.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology, 72,* 193–204.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271–280.

Corter, J. E., Gluck, M. A., & Bower, G. H. (1988). Basic levels in hierarchically structured categories. In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 118–124). Montreal, Canada: Erlbaum.

DeJong, G., & Mooney, R. (1986). Explanation-based learning: An alternate view. *Machine Learning, 1,* 145–176.

Dennis, I., Hampton, J., & Lea, S. (1973). New problem in concept formation. *Nature, 243,* 101–102.

Dominowski, R. (1973). Requiring hypotheses and the identification of unidimensional, conjunctive and disjunctive concepts. *Journal of Experimental Psychology, 100,* 387–394.

Ericsson, K., & Simon, H. (1984) *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Flann, N., & Dietterich, T. (1989). A study of explanation-based methods for inductive learning. *Machine Learning, 4,* 187–261.

Haygood, R., & Bourne, L. (1965). Attribute- and rule-learning aspects of conceptual behavior. *Psychological Review, 72,* 175–195.

Hovland, C., & Weiss, W. (1953). Transmission of information concerning concepts through positive and negative instances. *Journal of Experimental Psychology, 45,* 175–182.

Indow, T., Dewa, S., & Tadokoro, M. (1974). Strategies in attaining conjunctive concepts: Experiment and simulation. *Japanese Psychological Research, 16,* 132–142.

Indow, T., & Suzuki, S. (1972). Strategies in concept identification: Stochastic model and computer simulation I. *Japanese Psychological Research, 14,* 168–175.

Kelley, H. (1971). Causal schemata and the attribution process. In E. Jones, D. Kanouse, H. Kelley, N. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151–174). Morristown, NJ: General Learning Press.

Kraut, R., & Lewis, S. (1982). Person perception and self-awareness: Knowledge of influences on one's own judgments. *Journal of Personality and Social Psychology, 42,* 448–460.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence, 33,* 1–64.

Langley, P., Gennari, J., & Iba, W. (1987). Hill climbing theories of learning. In Pat Langley (Ed.), *Proceedings of the Fourth International Machine Learning Workshop* (pp. 312–323). Irvine, CA: Morgan Kaufmann.

Lebowitz, M. (1986). Integrated learning: Controlling explanation. *Cognitive Science, 10,* 145–176.

Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology, 71,* 331–338.

Levine, M. (1967). The size of the hypothesis set during discrimination learning. *Psychology Review, 74,* 428–430.

Lewicki, P. (1986). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 135–146.

Meyer, D., & Schvaneveldt, R. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Michalski, R. (1983). A theory and methodology of inductive learning. In R. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 83–134). San Mateo, CA: Morgan Kaufmann.

Mitchell, T. (1982). Generalization as search. *Artificial Intelligence, 18,* 203–226.

Mitchell, T., Kedar-Cabelli, S., & Keller, R. (1986). Explanation-based learning: A unifying view. *Machine Learning, 1,* 47–80.

Murphy, G., & Medin, D. (1985). The role of theories in conceptual coherence. *Psychology Review, 92,* 289–316.

Murphy, G., & Wisniewski, E. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Advances in cognitive science* (Vol. 2). New York: Wiley.

Nakamura, G. V. (1985). Knowledge-based classification of ill-defined categories. *Memory & Cognition, 13,* 377–384.

Nisbett, R., & Ross, L. (1978). *Human inference: Strategies and shortcomings of social judgments.* Engelwood Cliffs, NJ: Prentice-Hall.

Nisbett, R., & Wilson, T. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Palmer, S. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519–526.

Pazzani, M. (1990). *Creating a memory of causal relationships: An integration of empirical and explanation-based learning methods.* Hillsdale, NJ: Erlbaum.

Pazzani, M. (in press). A computational theory of learning causal relationships. *Cognitive Science.*

Pazzani, M., Dyer, M., & Flowers, M. (1986). The role of prior causal theories in generalization. In *Proceedings of the National Conference on Artificial Intelligence* (pp. 545–550). Philadelphia, PA: Morgan Kaufmann.

Pazzani, M., & Silverstein, G. (1990). Learning from examples: The effect of different conceptual roles. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 221–228). Cambridge, MA: Erlbaum.

Rajamoney, S., & DeJong, G. (1987). The classification, detection and handling of imperfect theory problems. In John McDermott (Ed.), *Proceedings of the Tenth International Joint Conference on Artificial Intelligence* (pp. 205–207). Milan, Italy: Morgan Kaufmann.

Reber, A. (1976). Implicit learning of synthetic languages: The role of instruction set. *Journal of Experimental Psychology: Human Memory and Learning, 2,* 88–94.

Reber, A., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition, 6,* 189–221.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Schank, R., Collins, G., & Hunter, L. (1986). Transcending inductive category formation in learning. *Behavioral and Brain Sciences, 9,* 639–686.

Shepard, R., Hovland, C., & Jenkins, H. (1961). Learning and memorization of classifications. *Psychological Monographs, 75,* 1–42.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18,* 158–194.

Wisniewski, E. (1989). Learning from examples: The effect of different conceptual roles. In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 980–986). Ann Arbor, MI: Erlbaum.

Wright, J. C., & Murphy, G. L. (1984). The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General, 113,* 301–322.

Wright, P., & Rip, P. (1981). Retrospective results on the causes of decisions. *Journal of Personality and Social Psychology, 40,* 600–614.

Zelano, P., & Shultz, T. (1989). Concepts of potency and resistance in causal prediction. *Child Development, 60,* 1307–1315.

Appendix A

The Influence Theory Used to Model Subjects' Knowledge of Inflating Balloons

easier	strong actor	inflate
easier	more elastic	inflate
implies	act stretch	more elastic
implies	old actor	strong actor
implies	age adult	old actor

(Appendixes continue on next page)

Appendix B

Productions in PostHoc

Initializing hypothesis

I1: *If* there is an influence that is present in the example, *Then* initialize the hypothesis to a single conjunction representing the features of that influence.

I2: *Otherwise,* initialize the hypothesis to a conjunction of all features of the initial example.

Errors of omission

O1: *If* the hypothesis is consistent with the influence theory *and* there are features that indicate an additional influence, *then* create a disjunction of the current hypothesis and a conjunction of the features of the example indicative of the influence.

O2: *If* the hypothesis is a single conjunction *and* a feature of the conjunction is not in the example *and* the conjunction consists of more than one feature, *then* drop the feature from the conjunction.

O3: *Otherwise,* create a new disjunction of the current hypothesis and a random feature from the example and simplify the disjunction.

Errors of commission

C1: *If the hypothesis is consistent with the background theory and* for each true conjunction there are features not present in the current example that would be necessary for an influence, *then* modify the conjunction by adding the additional features that are indicative of the influence.

C2: *Otherwise,* specialize each true conjunction of the hypothesis by adding the inverse of a feature of the example that is not in the conjunction and simplify.

Received September 18, 1989
Revision received August 13, 1990
Accepted October 30, 1990 ■