CONSTRAINTS AND PREFERENCES IN INDUCTIVE LEARNING: AN EXPERIMENTAL STUDY COMPARING HUMAN AND MACHINE PERFORMANCE

by

D. L. Medin
W. D. Wattenmaker
R. S. Michalski

ISG 86-1, UIUCDCS-F-86-952, Department of Computer Science, University of Illinois, Urbana, February 1986.

Reports

of the

Intelligent

Systems

Group

File No. UIUCDCS-F-86-952

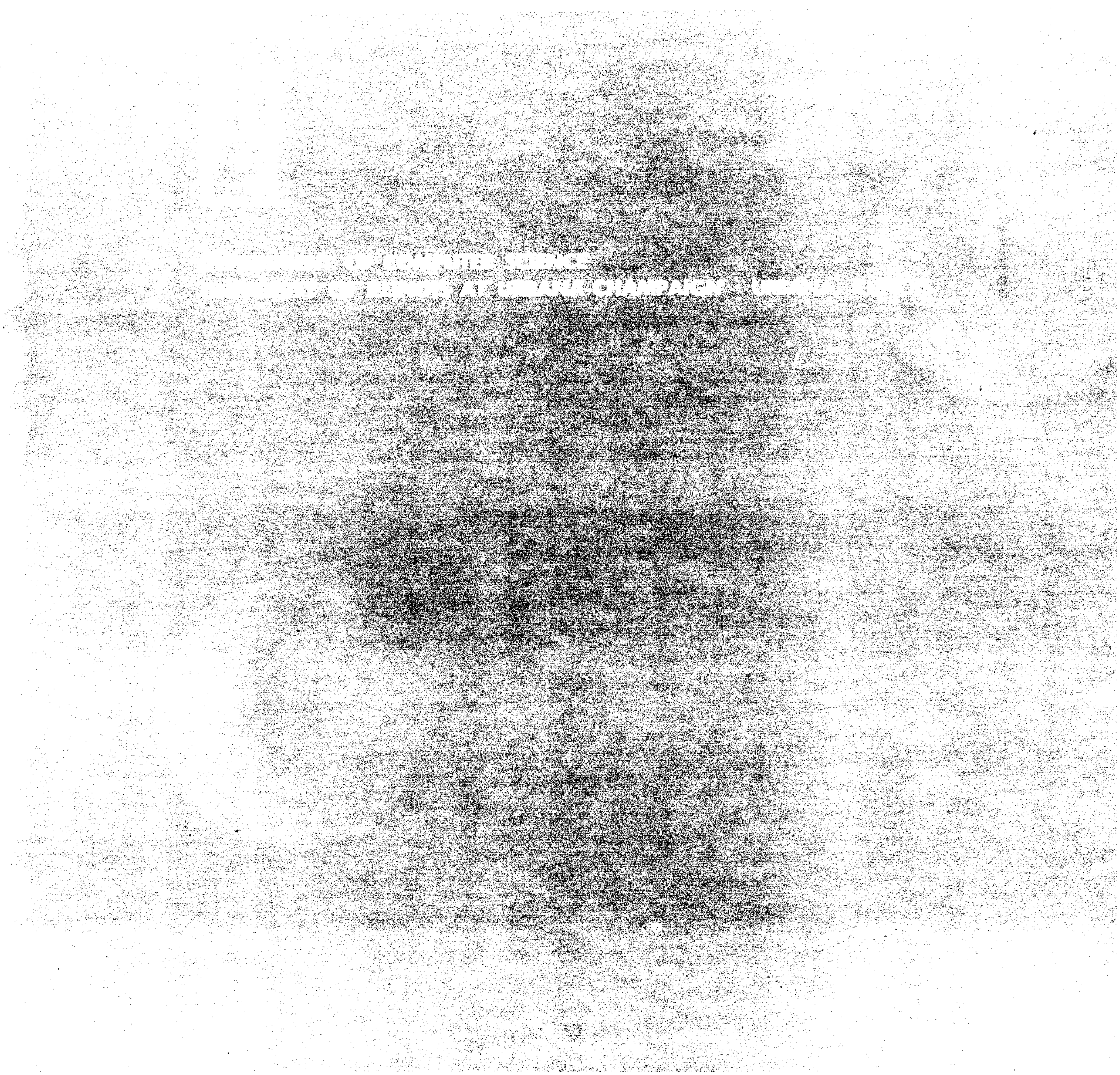
CONSTRAINTS AND PREFERENCES IN INDUCTIVE LEARNING
An Experimental Study Comparing
Human and Machine Performance

by

Douglas L. Medin
William D. Wattenmaker
Ryszard S. Michalski

January 1986

ISG 86-1



File No. UIUCDCS-F-86-952

CONSTRAINTS AND PREFERENCES IN INDUCTIVE LEARNING
An Experimental Study of Human and Machine Performance

Douglas L. Medin, William D. Wattenmaker
Psychology Department
University of Illinois

and

Ryszard S. Michalski
AI Laboratory
Department of Computer Science
University of Illinois

ISG 86-1

March 1986

Acknowledgements

We thank Renee Baillargeon, Mark Gluck, Elissa Newport, Larry Rendell, Brian Ross and Robert Stepp, for constructive comments on earlier drafts of this paper.

This research was supported in part by National Library of Medicine Grant LM04375 and NSF Grant BNS 84–19756 awarded to the first author, and in part by NSF Grant DCR–84–06801, ONR Grant N00014–82–K–0186, and DARPA ONR N0014–K–85–0878 awarded to the third author. Some of this research was completed while the third author was a visitor at the MIT Artificial Intelligence Laboratory (where he was supported in part by the Advance Research Projects Agency, Office of Naval Research, Contract N00014–80–C–0505).

Abstract

This paper examines constraints and preferences employed by people in learning decision rules from examples, and in constructing classifications of observed objects. Results from four different experiments with human subjects are analyzed and compared with those obtained from two AI inductive learning programs, INDUCE and CLUSTER, developed at the University of Illinois. These results indicate that human rule inductions and classifications tend toward greater specificity than would be expected if conceptual simplicity were the key preference. Such a bias may be explained by the fact that greater specificity, and thus greater the number of conjunctively linked assertions in the hypothesis, both maximizes the number of inferences that may be drawn from it, and protects the hypothesis against overgeneralization.

A significant correspondence has been observed between results from people and from the inductive programs investigated. One difference was that people tended to emphasize category validity (probability of some property, given a category) as contrasted with cue validity (probability of a category, given a property), to a greater extent than did the INDUCE program. The results of conjunctive conceptual clustering performed by CLUSTER seem to correspond well to the results obtained from human subjects.

The study seems to have a doubly positive effect. From the cognitive science viewpoint, the experiments with people and the analysis of performance of inductive learning programs present new clues for building a better cooperative model of human inductive processes. From the artificial intelligence viewpoint, the research reveals interesting new ideas for improving the current inductive learning programs.

CONSTRAINTS AND PREFERENCES IN INDUCTIVE LEARNING:

An Experimental Study of Human and Machine Performance

I. Introduction

Probably the most impressive fact about inductive learning is not that it occurs naturally in intelligent systems, but rather that it does not get out of hand. Any limited set of experiences will be consistent with an unlimited set of possible inductive generalizations. To give but one example, the next item in the sequence 1, 2, 4 might justifiably be 5 (increasing integers not divisible by 3) 8 (as in the equation $a_{n+1}= 2a_n$, $a_1= 1$), 14 (as in $a_{n+1}=a_{n_2}-a_n + 2, a_1= 1$), A (as in 1, 2, 4, A, B, D), or really anything. Therefore, a major issue concerns which of these possible inductive generalizations of given facts are preferred by people. This issue has become particularly salient with the advent of computer programs capable of inductive learning (e.g., see Michalski, Carbonell, & Mitchell, 1983, for a recent review). Aside from the general issue of how to form useful inductive generalizations, an important research topic for studies of human-computer interaction is the extent to which humans and computer programs form compatible inductive generalizations.

This paper is concerned with two types of inductive learning: rule induction from examples, and the construction of category partitionings ("clustering"). The search for contraints and preferences associated with these processes raises two related questions. First, given that experiences may be partitioned in a virtually limitless variety of ways (the set of potential partitionings of n objects increases factorially with n), why do we have the categories we have and not others? Second, how do we select among the large set of potential rules that can describe any particular classification or partitioning? Presumably only some of the possible partitionings and some of the possible rules are natural for human beings.

Why look for constraints? Our explicit assumption is that some rule inductions associated with partitions of entities are natural and others are awkward or unnatural. One possibility is that naturalness is strongly context-dependent; i.e., it varies with the specific contents of the entities under consideration. On that view, it simply is not possible to develop formal, universal constraints and preferences on rule induction, or they might have to be stated at a level too general to be useful. A more optimistic attitude is that fairly universal constraints or biases on rule inductions exist, and that they might provide important general principles for the question of how intelligent systems structure their experiences.

A more specific reason for seeking constraints and preferences on inductive generalizations concerns the compatibility between human and computer inductive learning. Inductive learning programs in artificial intelligence (AI) can be thought of as "expert systems" that can suggest new meaningful groupings of observations or generate descriptions of given classes of observations. If these new groupings or descriptions are to be useful, they must be understood. Therefore, it is essential that the groupings be described in a way that is compatible with human biases or descriptive preferences (for an example involving practical results from automated induction of descriptions of soybean diseases, see Michalski & Chilausky, 1980; Mozetic and Michalski, 1986). Conversely, constraints derived from human data provide candidate principles for AI programs.

Although the present studies are exploratory, they are motivated in part by principles derived from both AI and cognitive psychology. The next section describes some of these principles.

II. Candidate Constraints and Preferences

Cognitive psychologists have generated a large body of data on classification learning from examples and on the difficulty of learning different types of rules. In rule learning experiments, the experimenter creates a stimulus partitioning that conforms to some prespecified rule, and the data of interest concern the speed with which subjects converge on that rule. There has not been a concomitant interest in the situation where a partitioning admits of many possible rules, and the major issue is what forms and types of rules are developed from experience. Nonetheless, if there is a close link between ease of learning and naturalness, then one may be able to use results on learning difficulty to generate candidate biases in rule induction. Several factors that seem to influence the inductive learning process are considered below:

1. *Preference for simple rules.* It is true almost by definition that simple rules are easier to learn than complex ones. In fact the notion of simplicity or parsimony is so well engrained in the scientific community that one might wonder if any other preferences are needed. Simplicity, however, is a very elusive concept.

Informally speaking, rule simplicity is the inverse of conceptual complexity, where complexity reflects the time expended and resource costs, i.e., "mental effort," needed to employ the rule in decision making. One problem with this definition is that, for the same task, mental effort may differ with practice, background knowledge, and other contextual factors. If simplicity is defined only in terms of mental effort and cannot be specified in advance, then it becomes a dependent rather than an independent variable. For simplicity to provide a meaningful constraint on inductive learning it must be operationally defined.

In one attempt to be specific about simplicity, Neisser and Weene (1962) posited some basic logical operations (i.e., conjunction, disjunction, negation), and defined simplicity in terms of the number of operations needed to describe a partitioning. They also found that ease of learning was directly related to simplicity so defined. To the extent that one can specify which operations are basic, one can test the idea that simplicity provides a useful constraint on rule induction (see also Pinker, 1979). Because simplicity can change with the language of descriptions employed, it is important to evaluate simplicity within a theoretical framework that specifies basic operations and elementary concepts.

2. *Preference for conjunctive rather than disjunctive rules.* Rosch and her associates have persuasively argued that real-world categories are formed to exploit clusters of correlated attributes (Mervis & Rosch, 1981; Rosch, 1975, 1978). For example, animals with feathers are very likely to have wings and beaks, whereas animals with fur are very unlikely to have wings and beaks. In other words, correlated attributes carry information that permits one to go from knowledge of some attributes to predictions about others. An organism sensitive to these correlated or co-occurring attributes would find conjunctive concepts or rules more natural than disjunctive concepts or rules. Another important advantage is that conjunctive class descriptions allow one to determine properties of an object from knowledge of its class membership.

There is a fair amount of experimental evidence that conjunctive rules are easier to learn than disjunctive rules (Haygood & Bourne, 1965). Bourne (1974) has proposed that the relative difficulty of conjunctive and disjunctive rules arises from pre-experimental biases or preferences that favor conjunctive concepts, but results of experimental tests of this idea have either contradicted it (e.g., Dominowski & Wetherick, 1976) or suggested that biases may not be consistent over stimulus types (Reznick & Richman, 1976). Therefore, although the preponderence of evidence suggests that conjunctive rules are easier to learn than disjunctive rules, the support for this claim is far from universal.

3. *Sensitivity to cue validity.* Cue validity has long played a part in theories of perceptual categorization (e.g., Beach, 1964). The validity of a given cue or property is defined as the probability that an entity is a member of a given category. For the special case where cue validity is equal to unity, a cue is said to be *sufficient* (though it may not be necessary) for determining category membership. The basic idea is that organisms are sensitive to properties or cues which allow them to make correct categorizations. Elio and Anderson (1981) noted that people seemed especially sensitive to sufficient features in classification learning. As applied to rule induction in categorization, features entering into inductions should tend to be those that discriminate between categories. For example, having hollow bones has greater cue validity than being of a particular size in differentiating birds and mammals.

4. *Sensitivity to category validity.* Category validity is defined as the logical converse of cue validity, namely, as the probability that an entity has some property or cue given that it belongs to a category (Tversky, 1977). For the special case where category validity is equal to unity, a cue or feature may be said to be necessary (though it may not be sufficient) for category membership. To see that category validity is not the same as cue validity, one may note that category validity does not take into account whether a feature or cue is possessed by members of alternative categories. For example, having two legs has no cue validity with respect to differentiating *birds* from *people*. Category validity is similar to the correlated attribute principle in that it focuses on inferences that can be made from knowledge of

category membership. As applied to rule induction, one might speculate that features entering into inductions will tend to be those that are widespread within a category.

5. *Preference for positive over negative properties.* There is a substantial body of evidence suggesting that people have difficulty in processing negative information (e.g., Wason & Johnson-Laird, 1972). In the Neisser and Weene (1962) framework, negative assertions always involve extra operations that increase the task complexity. An exception to the suggestion that negatives always involve more complexity is internal disjunction. For example, the description "East or West or North" might be summarized more efficiently as "not South." In any event, one might expect people to prefer descriptions (rules) which minimize or do not involve negative features or properties. Recent studies show that this holds specifically in cases when subjects use a verbal problem representation. Subjects using a mental imagery strategy apparently are not affected by the negation. (Hunt, 1983).

These five candidate preferences do not add up to a theory of induction. Rather, they reveal an unsettled state of affairs. The various factors may trade-off against or complement each other. An immediate question concerns how one ought to express contraints associated with rule induction. Specifically, one may think of constraints as directly determining the *result* or outcomes of induction or they may act indirectly by being embodied in the *process* of rule induction.

The majority of psychological research has been directed at constraints stated in terms of products or outcomes. Keil (1981) offers some cogent arguments and evidence for the view that one should look for domain-specific constraints developed in terms of structures (or products) rather than processes. Keil takes the somewhat uneven picture on the relative difficulty of different types of rules as supporting the futility of looking for domain-general constraints.

Our position is compatible with Keil's, although in some respects it is the logical converse. We agree with Keil in that if one is committed to developing constraints in terms of particular structures or outputs, then such constraints will very likely be doman-specific. The focus of our present work, however, is the claim that if one is looking for domain-general constraints, then they should be embodied in the process of forming models of performance. To some extent, the distinction between process and output is artifical in that the two must necessarily be intimately linked. Pragmatically, however, there is a clear difference. The on focus products reflects the faith that output constraints will form a coherent picture. This may arise because there is a many-to-one mapping between alternative underlying processes and outputs, or because domains limit the set of plausible processing mechanisms. In contrast, the focus on processing principles carries with it the conviction that coherence more readily emerges from process constraints. For example, it may be possible to account for the mixed picture on the relative difficulty of conjunctive versus disjunctive rules in terms of a single underlying processing model. The danger associated with this commitment to processing principles is that one will formulate models which are too narrow and task-specific. Although any small set of experiments is likely to be susceptible to this latter criticism, we believe that our studies do illustrate the value of looking for processing constraints.

Our rationale for seeking such constraints is that this seems to be the natural way to evaluate relationships among the above candidate preferences that we have just discussed. Unfortunately, there is no extant psychological model that provides an account of how people construct categories or provide inductive generalizations or rules for preclassified categories.

Research on inductive learning in AI, however, has proposed answers to the questions we have been considering in the form of working computer programs. One major purpose of the present paper is to examine the extent to which the preferences embodied in these AI programs also act as preferences for human rule induction. That is, we will treat these programs as a first approximation to a psychological theory of rule induction. As will be seen, there are numerous parallels between candidate preferences derived from cognitive psychology and biases incorporated into these programs. The second major purpose in our comparison of human and machine rule induction is to see if processing principles from human rule induction provides any clues for the enhancement of methods embodied in machine inductive learning.

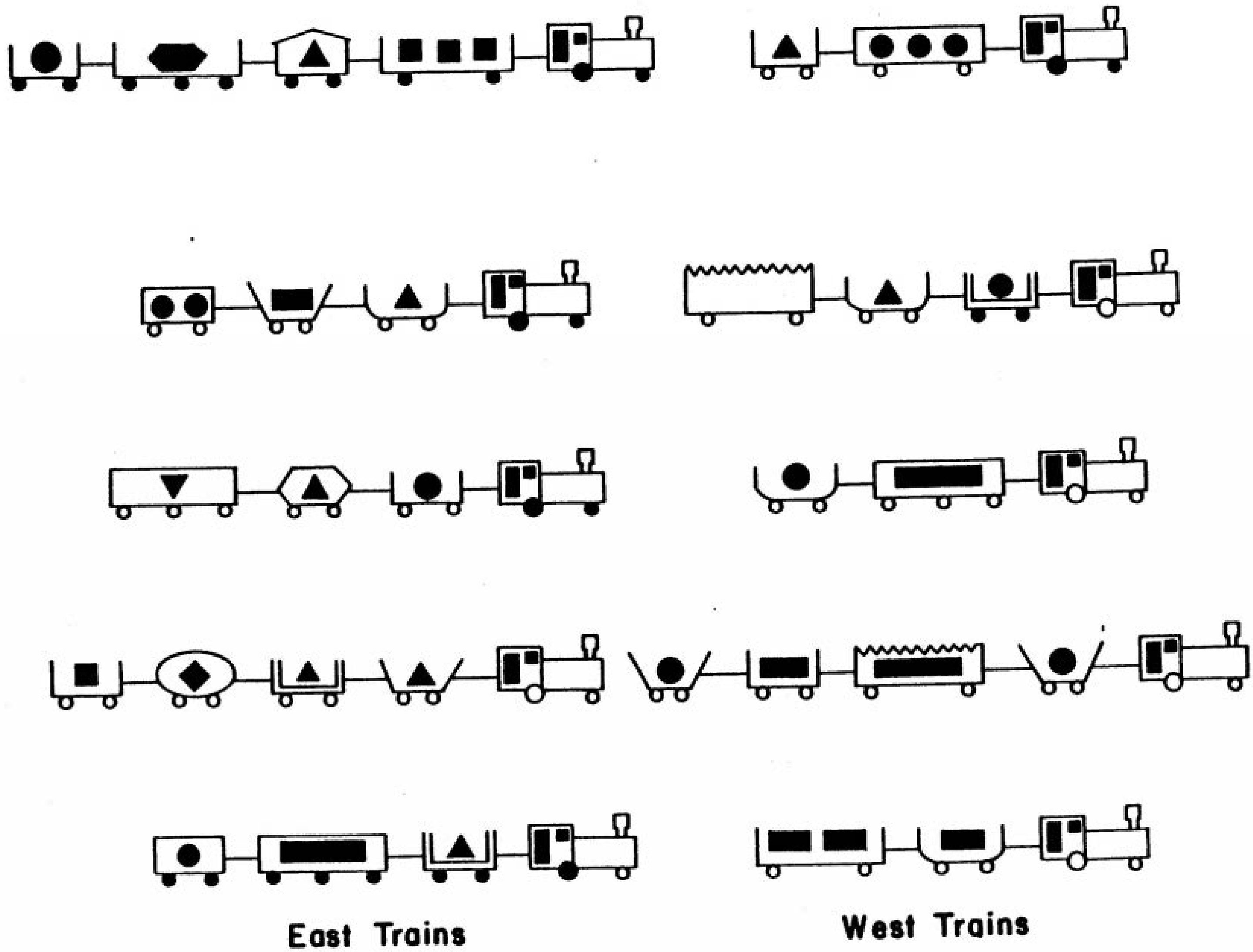
Although there are numerous inductive learning systems (see e.g., Dietterich and Michalski, 1981, 1983 for a detailed review) we will primarily concentrate on two particular programs, INDUCE AND CLUSTER. There are three main reasons for our choice. One is that stimulus materials which we employed require structured descriptions, and many AI systems do not have descriptive languages that are this powerful. The second reason is that INDUCE and CLUSTER were specifically designed with the criterion of human comprehensibility in mind (the rules should make sense or seem natural to people; see Michalski, 1980, 1983 a,b), and, therefore, they seem like particularly good candidates for psychological models. The third reason is more pragmatic and not specific to CLUSTER and INDUCE. It is not feasible to provide a description of the algorithms associated with each inductive learning program because to illustrate our approach requires far more detail than otherwise might be provided. In the general discussion we will provide a more detailed summary of the adequacy of other AI induction programs as psychological models. First, however, we turn our attention to INDUCE and CLUSTER.

III. Constraints in the INDUCE and CLUSTER programs

A. INDUCE

Since Michalski's program, INDUCE, is designed to provide inductive generalizations that are biased in such a way as to be comprehensible by people, it can be thought of as a psychological model embodying conjectures about what is natural for people. This program performs a heuristic search through a space of possible symbolic descriptions which in turn are generated by the application of various inference rules to the initial observational statements. The following paragraphs describe INDUCE in a general way and the reader is referred to Michalski (1980, 1983 a,b) for a more detailed, technical presentation of INDUCE.

To see how INDUCE works, it is helpful to have a specific example in mind. Figure 1 shows the set of trains that was used in the first experiment. The trains differ in numerous properties such as wheel color, car shape, and load shape. The five trains on the left are said to be Eastbound, and the five trains on the right are said to be Westbound. The task for INDUCE, as well as our experimental participants, is to come up with a simple rule that could be used to determine whether a train is Eastbound or Westbound. It should be obvious that there is a large set of candidate rules, ranging from a disjunction of descriptions of individual examples to the most general possible assertion about each set of trains. The issue of interest is whether or not the forms of rules people develop are similar to those constructed by INDUCE.



Eastbound and Westbound Trains presented in subjects in Experiment 1

Figure 1

Description of Rules. The initial input to INDUCE consists of a set of observational statements characterizing each example. For instance, each car of the train may be described as being long or short, as having a particular shape, and so on. These elementary descriptors, attributes, functions or predicates may be nominal (e.g., sex), linear (e.g., length), or hierarchically structured (e.g., shape, with values such as triangle, square, polygon, etc.).

The descriptors used in the input data are not necessarily the final descriptors used in inductive assertions. In the process of formulating inductive generalizations INDUCE applies various inductive inference rules to develop general descriptions of the initial observational statements. These inductive rules can be classified as either *selective* or *constructive.* Selective inference rules directly incorporate descriptors used in initial concept descriptions. Examples of selective rules include *turning constants into variables,* (e.g., replacing *red* by *any color*), dropping conditions (assuming that some property is irrelevant), and closing intervals (e.g., if entities have values of either four or six on some dimension, then this operation would transform the description to "value between 4 and 6"), creating *internal disjunction* (e.g., "value 4 or 5 or 6"), and *climbing generalization* tree for hierarchically structured variables (e.g., transforming "Chicago or Dekalb or Peoria" into the city in "Illinois"). Negative descriptors are not normally employed except in two situations. The first situation is when a negative descriptor yields a more succinct expression. For example, given a choice between "triangle or rectangle or pentagon or ellipse or circle" and "not square," the latter description would be used. The second situation occurs when using the generalization rule called *extension against.* If example A is positive, and example B is negative, then the rule creates the negation of any property in B that is not shared by A. Such a negation is the most general assertion describing A and excluding B (Michalski, 1983), called *extension against.*

Constructive generalization rules involve creating new descriptors not present in the original observational statements. For example, there is a *counting rule* that creates descriptors measuring the number of occurrences of the same term or attribute value in a description (e.g., *two red circles).* Another rule, *generating chain properties,* creates descriptors characterizing ordered relationships, such as "first," "middle" or "last" in a series. Other constructive generalization rules exploit descriptor interdependence, such as might be present when attributes are correlated. For particular domains, the user may suggest additional constructive generalization rules.

General Algorithm. INDUCE begins with a set of descriptions of entities, then selects a target concept (say, the Eastbound trains) and proceeds as follows.

1. An example called the "seed", is randomly selected from the target set, i.e., the set of examples representing the target concept.

2. The seed is then described in alternative and most general ways, so that they may apply to as many other examples of the target set as possible, but not to examples of the contrasting category. In the process of generating these descriptions, both selective and constructive generalization rules are applied.

3. Descriptions on the candidate list are tested for consistency and completeness. A description is *consistent* if it does not apply to any members of the contrast set (i.e., it has no counter-example). This is equivalent to cue validity being equal to unity, if the entire description is treated as a single cue. A description is *complete* if it applies to all members of the target category. This is equivalent to a description having a category validity of unity. Descriptions that are consistent and complete represent alternative

solutions and are saved.

4. Alternative descriptions are evaluated in terms of a preference criterion which can be varied (and will be described below). A list of the k most preferred descriptions is developed (the candidate hypothesis list).

5. If no admissable solution appears from the initial candidate list, then attention focuses on descriptions that are consistent but not complete. In this event, the best description (e.g., the one covering most of the events) is saved, the target set is reduced to those examples not covered by the saved description, and the process reverts to step 1.

6. The disjunction of all generated descriptions is a consistent and complete concept description. Since the process starts with a single example of a concept and any single example can always be characterized by a conjunctive expression, the solutions have a bias toward conjunctions. It may not always be the case that a single conjunctive statement will be complete and consistent; therefore, the general form of solutions will be a disjunction of conjunctive descriptions.

The Preference Criterion. The quality of any particular consistent and complete description *(candidate* description) may be evaluated according to multiple criteria. Two major criteria that INDUCE considers are simplicity, and the "fit" between examples and a description. The measure of simplicity may involve costs of measuring attributes, the memory requirements, and the number of descriptors and operators used in the generated inductive assertion. Simplicity encourages short, general, and easily computed descriptions. The notion of fit is designed to avoid overly general rules, and to a large extent is in opposition to simplicity. Fit refers to how well an inductive assertion matches the examples of the target set. It reflects the amount of uncertainty that any given object description satisfying the inductive assertion corresponds to an actual example. For example, if the target set contained a small red triangle and a large red triangle, and the contrast set consisted of a blue circle and a green triangle, then the solution "red and triangle" would have a better fit (and be less general) than the assertion "red."

The descriptions are ranked using a preference criterion based on a "lexicographical evaluation functional." It is an ordered list of "criterion-tolerance" pairs supplied by the user where each criterion on the list represents one elementary attribute characterizing candidate assertions. The program offers a collection of elementary criteria from which the user selects the most relevant for the current problem. The list represents a set of successive hurdles for candidate assertions to meet until the best one is selected or all hurdles have been met. Associated with each criterion is some tolerance value that defines the threshold for meeting the criterion. The preference function is similar to Tversky's (1972) elimination by aspects model for choice behavior. The importance of any one criterion is determined by its *rank* in the list (the higher, the more important) and by its *tolerance* (the greater the tolerance, the less stringent the criterion). Since the user supplies the criterion-tolerance pairs, the preference criterion is fairly flexible. For example, by appropriate ranking of simplicity and fit one can produce either *characteristic descriptions,* which focus on properties common to a class, or *discriminant descriptions,* which focus on properties necessary to differentiate between classes.

The parallels between the constraints drawn from cognitive psychology and those associated with INDUCE are quite close. INDUCE has a notion of simplicity, embodies a bias favoring conjunctive descriptions, and generally avoids negative features. The issue of cue validity versus category validity corresponds to the difference between discriminant and characteristic descriptions. INDUCE embodies certain constraints and parameterizes others in terms of its preference criterion function. Although one can readily imagine that relative importance or weighting would vary considerably as a function of task factors, it is an empirical question whether or not consistent patterns or preferences in ranking criteria associated with human inductive generalizations are observed.

B. CLUSTER

Our other main interest here is in constraints and preferences associated with the construction of classifications, as opposed to creating rules characterizing preclassified sets of entities, as done by INDUCE. In this case people are given a set of stimuli, and asked to sort them into categories in a way that seems sensible. Michalski and his associates (e.g., Michalski & Stepp, 1983; Stepp & Michalski, 1986) have developed a program for constructing category classifications (sorting) called CLUSTER. When CLUSTER is presented with a set of descriptions of objects and asked to put them into k classes, it constructs k clusters of objects, and describes each cluster by a single conjunctive statement. Like INDUCE, CLUSTER has a preference function incorporating the criteria of similarity and fit, and current versions of CLUSTER have both selective and constructive generalization rules. The psychological implications of CLUSTER are analogous to those of INDUCE: if people follow the same constraints as CLUSTER they will prefer category partitionings that can be described conjunctively, and these descriptions will not necessarily be determined exclusively by simplicity.

The CLUSTER program works by converting the problem of clustering into a sequence of rule construction tasks. To illustrate this program very briefly, assume that the number of desired clusters is k. The program randomly selects k objects (seeds) from the population, and treats them as hypothetical representatives of the k classes. The program then develops k mutually disjoint general conjunctive descriptions, one for each seed, which together cover the whole population, and score best on the clustering *quality evaluation criterion*. Next, sets of objects covered by each description are determined, and from each set a new seed is selected. In one implementation the seeds selected alternate between those which are central in obtained clusters, and those occupying an extreme position in clusters. This produces more rapid convergence to quasi-optimal clusters. The algorithm then reverts to the first step. This process repeats until a stable clustering is determined.

IV. Overview of Experiments

The preceding analysis of constraints and preferences in inductive learning suggests a research strategy. First of all, evidence is needed bearing on the validity and importance of these candidate constraints on rule induction in classification. If more than one factor emerges as important, then followup studies can be targeted at the relative performance of each factor. A related question will be how general any constraints prove to be across tasks.

In the present studies the program INDUCE is used both to operationally define simplicity and as a model of human performance. To the extent that INDUCE captures people's inferences, it will receive support as a psychological model, and will provide a framework for evaluating the relative importance of different factors influencing the naturalness of inductions. If the processes associated with people's development of inductive generalizations show systematic differences from INDUCE, then these differences can be used either to modify INDUCE (if the differences involve factors that may provide useful constraints on induction or increase comprehensibility), or to develop psychological models of people's inductive generalizations (in the event that the differences are that people depart from what is useful or optimal).

Experiment 1

The first experiment was exploratory and employed a combination of classification construction (sorting) and rule induction tasks. The stimulus materials consisted of the 10 trains shown in Figure 1. Participants were initially asked to arrange the trains into any number of groups (classes) in a way that made sense to them. They were then asked to describe the basis for their classifications. Next, participants were asked to perform two additional classification construction tasks. The first had the constraint that there should be exactly two categories of equal size (of five members each). The second was identical to the initial unconstrained task except that participants were told that they could employ an "else" category for trains that did not fit any of their preferred groupings. Finally, participants were told that the 5 trains on the left side of Figure 1 were Eastbound, that the trains on the right side were Westbound, and that their task was to come up with a rule that could be used to decide if a new train was East– or Westbound.

There were several objectives in this initial study. We were interested in the relationship between classification constructions and descriptions of them, particularly for the initial task when participants did not know they would be asked to provide justifications for their groupings. In particular, the experiment provided a data base of descriptions that could be used to evaluate the adequacy of the generalization rules associated with CLUSTER and INDUCE. To sharpen this comparison, half of the participants were told which features were relevant (the same ones as used in the initial input to INDUCE) and half were not. If human rule inductions correspond to those of INDUCE, then processing constraints associated with INDUCE will be supported.

Method

Subjects. The subjects were 64 undergraduates (male and female) attending the University of Illinois, who were paid for their participation in the experimental session which lasted about one hour. The participants were randomly assigned to either the Standard group or the Informed group.

Stimuli. The stimulus materials consisted of the drawings of 10 trains shown in Figure 1. The trains were mounted on 7.6 cm by 12.7 cm index cards. As may be seen in Figure 1, the trains could differ in the number and shape of cars, in their tops and loads, and in the number and color of their wheels.

Procedure. The experimental procedure consisted of three classification construction tasks followed by a rule induction task. Participants were tested individually. Details of these procedures were as follows

1. *Free Classification.* For the initial task participants were asked to carefully look over the trains and then to put them into groups in a way that made sense to them. After this free classification was completed, each participant was asked for the basis of his or her groupings.

2. *Constrained Category Construction.* For the next task, participants were asked to put the trains into two equal-sized groups in a way that made sense. Then participants were asked again to justify their partitionings.

3. *Free Classification with "Else" Category.* The last partitioning task was identical to the first, except that participants were told that they could have a "junk" category for trains that did not fit in with other groups.

4. *Rule-Induction.* For the rule induction task, participants were presented with the two groups of trains corresponding to the left and right half of Figure 1 and told that one group was Eastbound and the other Westbound. They were told that their task was to come up with a rule that could be used to decide if a train was East– or Westbound. Participants performing these tasks were divided into the Standard and the Informed

group.

Participants in the Standard condition were not presented with any description of the trains. Participants in the Informed condition were told at the start of the experiment that the following set of attributes was relevant: shape of cars, number of cars, length of cars, number of loads, shape of loads, type of car top (open or closed), number of wheels, and color of wheels (white or black).

Results

Because of a procedural error the data from one of the participants in the Informed condition could not be used. The results will be presented separately for each of the sortings and the rule induction test.

Free Classification. Most of the participants constructed groups of trains on the basis of a single property although a significant minority used a conjunction of properties. No one described their sorting as involving a disjunction of properties. A breakdown of partitioning strategies is shown in Table 1.[1]

Of the partitionings based on a single property, *number of cars* was the predominant basis for classification, accounting for about three-fourths of the unidimensional groupings. The relation between bases for classification and number of categories was straightforward. For example, people sorting on the basis of number of cars created three groups corresponding to 2, 3, or 4 cars in a train. Sorting by wheel color typically involved three groups: wheels all black, wheels all white, and wheels mixed in color.

---

[1] A more detailed description of these data is available upon request.

There were few, if any differences between the Standard and the Informed condition. The possible exception is that in partitions based on a conjunction of properties six persons in the Standard condition but only one in the Informed condition used some combination of car position (first, middle, last) with another property. Car position was not given as a relevant dimension to participants in the Informed condition. The CLUSTER program was not given car position as a descriptor but it could produce it as a descriptor using the Generating Chain Properties Rule (Michalski, 1983). Finally, none of the descriptions involved negative properties or attributes.

Constrained Category Construction. When participants were asked to sort the trains into equal-sized groups, they continued to employ single properties or conjunctions of properties. A breakdown of reported sorting strategies is shown in Table 2.

There were no obvious differences between the Standard and Informed conditions. Color of engine wheels was the most common basis for sorting in both groups. The presence or absence of a particular shape (e.g., rectangles) was the next most popular strategy among people using a single property. One participant in each condition sorted on the basis of whether or not the loads on a train were all different. This strategy would be captured by constructive generalization rules in CLUSTER (e.g., Stepp & Michalski, 1984). We were unable to understand the partitionings of two participants in the Standard condition because the partitions and their descriptions did not seem consistent. One of these descriptions mentioned number of cars and the other mentioned simple versus complex trains without elaborating on the basis of this complexity.

Slightly more than a fourth of the participants used a combination of properties. For example, the partition might be defined in terms of whether or not there was a circle load in the last car. An equal number of people in both conditions (4) used conjunctions involving car position. Finally one participant in the Standard condition used a disjunctive description. Negative properties were not mentioned except where an entire category was defined by exclusion from the alternative category. No participant sorted the trains in a manner corresponding to Eastbound and Westbound categories in Figure 1.

Free Classification with "Else" Category. Almost every participant used a different classification principle when they were allowed to employ a miscellaneous category from the one they used on the initial free classification. In addition, every participant put at least one train into the else category. Both of these results probably arise from implicit task demands rather than some intrinsic property associated with being able to use a junk category. One major change which does not appear to be a function of implicit expectations is that the predominant basis for sorting shifted from being based on a single property to a combination of properties. A breakdown of these data is shown in Table 3.

The increased use of conjunctions of properties was associated with an increased variety of property combinations. For example, participants used combinations of load shapes, combinations of car shapes, and load shapes in same versus adjacent cars. Descriptions involving combinations could be readily incorporated into INDUCE. Car position was used by more participants in the Standard condition (8) than in the Informed condition (2). No descriptions involved negative properties and no one employed a disjunctive description.

East-West Rule. The rule inductions provided the most straightforward test of INDUCE. The results are summarized in Table 4.

The task proved to be quite difficult. Two people in each condition discovered a simple classifier based on the number of different loads. About a third of the participants employed conjunctions of properties. A large majority of conjunctive rules made use of negative properties. About half of the participants used a disjunctive description, the most popular of which was the simple rule that Westbound trains have two cars or a jagged top. Many of the disjunctive descriptions, however, were fairly elaborate and involved conjunctions of properties as part of the disjunctive rule. When negative properties were part of these complex disjunctions, they usually (but not always) were associated with the conjunctive part of them. Finally, a few participants were unable to come up with rules and either gave partial rules or detailed descriptions of particular trains. Three participants in the standard condition and four in the informed condition gave rules that did not perfectly partition the trains. A complete list of rules is given in the Appendix (Table A1).

Overall, the results are generally consistent with INDUCE. The one exception is that people tended to begin with rules that had counterexamples (e.g., three or more cars) and then eliminate the counterexamples by using negative properties (as in the rule, East: three or more cars and not jagged top). As will be seen, this pattern is consistent with the fairly straightforward processing model for rule induction to be considered next. This processing model is very similar in spirit to INDUCE. It also reflects, on a small scale, Kuhn's notion of a paradigm shift (Kuhn, 1962). That is, when observations do not fit the current theory (description) a very common strategy is to attempt to patch up the current theory and only when these modifications become too unwieldy is the theory abandoned.[2]

[2]Rick Lathrop from MIT AI Laboratory suggested this relationship to Kuhn's ideas.

Theoretical Analysis

A Process Model for Rule Induction. It is convenient to characterize performance in terms of consistency and completeness. Recall that consistency refers to descriptions that have no counterexamples but may not cover all known members of a category, whereas completeness refers to descriptions that cover all members of a category but may have counterexamples (apply to members of alternative categories). Current versions of INDUCE look for consistent and complete descriptions ("candidate hypotheses") but give more weight to consistency than completeness. The data from human subjects are best accounted for by the idea that completeness may be more important than consistency in the initial phases of rule formulation. Therefore, not only is it the case that consistent rules are modified to make them complete but also complete rules are modified to make them consistent.

One way to formalize these ideas about consistency and completeness is in terms of the following process model: People focus on one category and begin by looking for a descriptor that spans the positive set and does not apply to any counterexample. If one is found, then a simple rule can be generated. If no single descriptor works, because there are counterexamples, then one of two strategies may be applied. If there are numerous counterexamples, then people may look for combinations of properties (e.g., "X and Y") that span the set but do not generate counterexamples. If there are only a few counterexamples, then people may attempt to eliminate them by negating properties of the counterexamples not present in the positive set. For example, a person may notice that all Eastbound trains have a triangle load but that two Westbound trains also do. This description is complete but not consistent. They might then look for combinations of properties that apply to the East but not the West trains. For example they might consider the rule "triangle load in nonlast car," but that rule would still have a counterexample. Next a person might consider properties true of these two Westbound trains that are not shared by the East trains. For example, they might notice that the two West train counterexamples have a long car with two white wheels and then generate the rule "Eastbound trains have a triangle load and not long cars with two white wheels."

The other main possibility is that a descriptor has no counterexamples but fails to span the positive set. In that event people form a disjunction using the initial descriptor and then confine attention to the reduced positive set and the contrast set. For example, they might notice that only Westbound trains have two cars, and then focus on differences between the remaining two Westbound trains and the Eastbound trains. They might notice that the remaining West trains both have jagged tops and generate the rule "Westbound trains have two cars or a jagged top." This part of the process model is functionally equivalent to INDUCE and the above rule is one of those that INDUCE actually discovers.

This account seems quite consistent with the present results. The descriptor, number of different loads, was apparently not very salient (it would involve a constructive rule for INDUCE) and few participants found the simple rule based on it. As judged by the initial free sorting, *number of cars* was quite salient and many people found the simple disjunction, number of cars and jagged top. According to this process model, negative descriptors (e.g., not jagged top) should be part of conjunctions and not part of disjunctions. This held for 17 of the 20 cases where negative descriptors were used. The three exceptions seem to be cases where the reference (positive) set and the contrast (counterexample) set shifted at some point during the rule search. Two exceptions were of the form "not triangle or triangle and..." and the third was "not dark engine wheels or dark engine wheels and...." In this model the relative number of disjunctive and conjunctive rules would depend on the exact structure of the trains and the salience of the associated descriptors. In general, however, because people are assumed to initially focus on properties that members of the positive set have in common, conjunctive rules are likely to result.

Relation to INDUCE. In general the people's rules were quite similar to those produced by INDUCE. Both INDUCE and many participants appeared to discover consistent but not complete descriptors and then confine attention to the reduced positive set and the contrast set. This would produce disjunctive rules where one or more parts of the disjunction might consist of a conjunction of descriptors (again, see Table A1 in the Appendix for a detailed listing of rules). The descriptors in the rules were either consistent with the original descriptions given to INDUCE or could be readily produced by constructive generalization rules.

The largest difference between solutions given by people and by INDUCE is that a fair number of people appeared to find descriptors that were complete but not consistent and then remove the inconsistencies by negating properties that were unique to these counterexamples (for example, the rule that East trains "have a triangle load and not three circular loads and not a jagged top"). The current implementations of INDUCE focus on a list of consistent (but not necessarily complete) descriptions but do not allot similar attention to complete (but not consistent) descriptions.

Discussion

The category constructions, their descriptions, and the rule inductions were consistent with at least some of the biases outlined in the introduction. The partitionings were predominately either on the basis of a single property or on a conjunction of properties. This is consistent with the principles of simplicity, category validity, and a preference for conjunctions over disjunctions, all of which are embodied in CLUSTER. The descriptions of these partitionings did not involve negative properties. The informed group did not confine itself to the original list of properties but their new descriptors were consistent with the constructive generalization rules associated with CLUSTER and INDUCE.

The rule induction data revealed both disjunctions and negative properties. The negative properties almost always were part of conjunctive descriptions and fit quite well with both INDUCE and a less formal process model that assumes that when people find a descriptor that spans a set but is consistent with some members in the contrast set, they attempt to eliminate these counterexamples by developing a rule based on negating their properties. Disjunctive descriptions arise when a salient descriptor has no counterexamples but fails to span the positive set. The only significant difference between INDUCE and the informal process model is that the former places more emphasis in its typical parameter settings on consistency than completeness in the process of developing solutions. By setting the parameters associated with the preference criterion differently, more importance can be attached to completeness, but current versions of INDUCE do not have an intermediate stage where complete but not consistent solutions are saved. Although in principle INDUCE could be modified along these lines, for present purposes it will be most convenient to describe our results both in terms of INDUCE and the less formal process model. It should be kept in mind, however, that the two models are highly similar and embody almost identical biases in rule induction.

Experiment 2

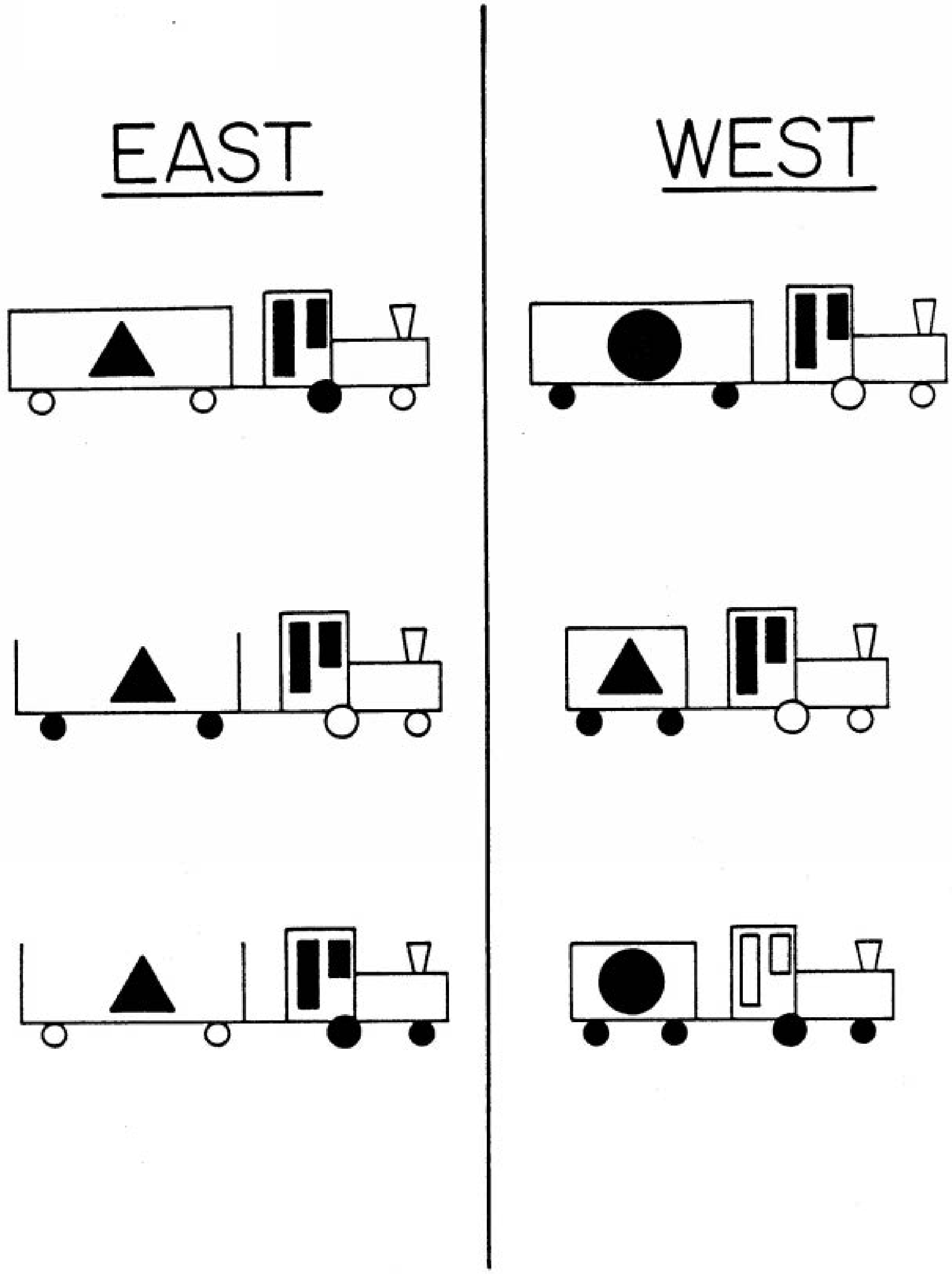
Although the first experiment was useful by being complicated enough to give INDUCE and CLUSTER a serious test, the study did not provide any strong contrasts among alternative constraint principles. The second experiment was concerned only with rule induction. It was designed to pit conjunction and category validity against disjunction and cue validity. The stimuli were simplified trains shown in Figure 2. The experimental task was to come up with a basis for determining whether a train was Eastbound or Westbound. As in the first experiment, there are many possible inductive generalizations consistent with Figure 2 and the main question is which of these people prefer. We were particularly interested in the relative preponderance of conjunctive and disjunctive rules. For example, Eastbound trains can be described either by the rule "long car *and* triangle load in car" or by the rule "open car *or* white wheels on car." The conjunctive rule combines two properties each having high category validity and lower cue validity and the disjunctive rule combines two properties each high in cue validity and lower in category validity. The two types of rules are equally simple but the program INDUCE predicts that conjunctive solutions will be more frequent than disjunctive solutions. Although the various train properties are not counterbalanced, it would be hard to explain rule preferences in terms of the salience of stimulus dimensions. For example, if car length and load type were salient it might produce a bias for conjunctive rules involving Eastbound trains but it also ought to produce a corresponding bias for disjunctive rules (West = short car or circle load) involving

Westbound trains.

It should be noted that this prediction of a bias toward rule constituents that have high category validity is not a property of all inductive learning algorithms. For example, one might imagine a process model which initially computes the cue validity of each descriptor, orders descriptors first by cue validity and secondly by category validity, and then develops rules by going down the list of descriptors until a rule is created which is consistent and complete. Whenever no single descriptor was both consistent and complete, disjunctive rules would be produced. A related algorithm would determine the information value (rather than the cue validity) of candidate test properties and develop a discrimination net with the most informative test occupying each node in the network. This is the procedure embodied in the ID3 technique of Quinlan (1975,1979). In the present task the consistent–but–not–complete and complete–but–not–consistent descriptors are mirror images of each other, so there is no reason to expect a preference for one type of rule over the other, according to Quinlan's framework.

Method

Subjects. The subjects were 66 male and female undergraduates attending the University of Illinois who participated in this experiment in partial fulfillment of course requirements in introductory psychology. Participants were run in groups of 10 to 15 and the experiment lasted about 10 minutes.



Eastbound and Westbound Trains presented to subjects in Experiment 2

Figure 2

Stimuli. The stimulus materials consisted of six trains placed on a single sheet of paper as shown in Figure 2. The trains differed from each other in color of car wheels, car loads, car length, car top, engine door and window color, and engine wheel color. A given property either was true of all members of one class but had counterexamples in the contrast category, or was true of only some members of one class but had no counterexamples. These two types of properties can be thought of as maximizing category validity and cue validity, respectively.

Procedures. Participants were given the sheet of six trains and their East–West designation and asked to examine them. They were told to come up with a basis for classification that could be used to predict whether a new train would be Eastbound or Westbound and that, at a minimum, the basis for classification should properly classify the six trains on the sheet.

Results

There was some ambiguity as to whether a basis for classifying both sets of trains was needed or whether one set could be defined by exclusion. Out of 66 participants, three gave a criterion characterizing only one of the sets. The remaining 63 people provided some basis for classifying each set, but there is reason to believe that the primary focus was on Eastbound trains. In scoring descriptions or rules for whether or not they could be used to successfully classify the trains, one finds that for 17 participants the East classification principle was adequate but the West one incomplete, whereas for only 3 participants the West principle was adequate and the East incomplete. For 7 participants both the East and West classification principles were incomplete. Also, the instructions did not specifically ask for statement of a decision rule and a significant number of people, 27, only provided a list of descriptors that might be useful in classifying trains.

Overall, people showed a very strong preference for conjunctive rules. Since it was possible for people to give different forms of classification rules for the East and West sets, the details supporting this generalization are a little complicated. Altogether, 34 people gave a conjunctive rule for East trains, and of these, 20 also gave a conjunctive rule for West trains, 5 gave a disjunctive rule for West trains, and 7 simply gave a description of West trains but no rule that could be used to classify the trains. One person gave a disjunctive rule for both East and West trains and three people gave a disjunctive rule for one set (2 West, 1 East) and did not provide a basis for classifying the alternative set. Two people gave both a conjunctive and disjunctive rule for East Trains and a disjunctive rule for West trains. As a whole, then, the rule statements showed a strong bias for conjunctive over disjunctive rules.

A further breakdown of classification principles offered by participants is shown in Table 5. Conjunctive rules predominate over disjunctive rules. More than the minimum necessary to classify the trains was contained in 21 of the 57 conjunctive rules. For example, a typical East rule was "long cars and triangle load and black rectangles on engine." This implies that the people were not focusing exclusively on discriminant descriptions. Only two rule statements mentioned negative features and both of those cases appear to embody the Extension Against principle (e.g., West: circle or short; East: not circle and not short).

The descriptions also seem consistent with a conjunctive bias or at least a preference for category validity over cue validity. A very large majority of the descriptions mentioned properties that members of a set had in common (maximizing individual property category validity) compared with those possessed by some members of a set that were not present in the contrast set (maximizing individual property cue validity).

These results cannot be explained simply in terms of component salience. Although there is some evidence that people tended to use rules based on car length and load type, the dimensions employed in rules varied with whether the trains were East or West. Overall, 93 percent of the East rules mentioned car length or load type but only 27 percent of the West rules mentioned car length or load type. It appears that the form of the rule, conjunctive versus disjunctive, influenced performance much more than the salience of component properties.

Discussion

The main results of this experiment are in terms of both descriptions and rule statements and they form a coherent picture. Although both rules were equally complex in terms of number of descriptors and operators, people showed a strong bias for conjunctive rather than disjunctive rules. The common properties entering into conjunctions maximize component category validity (probability of the property given the category) in contrast with the discriminative properties of disjunctive rules which maximize component cue validity (probability of the category given the property). For the protocols giving descriptions rather than rules there was a corresponding bias for common over discriminating properties.

This bias for conjunctive rules and common properties is consistent with INDUCE and the less formal process model outlined earlier. This bias arises from the assumption that the first stage in rule induction involves generating a description of properties that members of a set have in common and then refining it to exclude counterexamples. For the trains in Figure 2 the conjunction of two descriptors (e.g., dark wheels, closed top) has no counterexamples and a simple conjunctive rule can be discovered. The task of finding common properties should have been and apparently was easier than in the first experiment because fewer, less complex trains were employed. The results are inconsistent with the idea that properties are ordered by cue validity alone or information value alone and then developed into rules (by, for example, generating a discrimination net). Ordering by cue validity predicts a bias for disjunctive rules and ordering by information value predicts no bias.

Although a bias for conjunctive rules and for component category validity are closely associated, they are not indistinguishable. For instance, if a single descriptor is complete but has a counter-example and the counter-example has a distinct, salient property, then one ought to see opportunistic conjunctions based on negating that property. We gave an additional 22 subjects the rule induction task involving the trains in Figure 2 but we added a smokestack to either the West train that had a triangle load (for half the subjects) or to the West train that had a long car. This change led to 11 simple or redundant conjunctive rules and, more importantly, 6 opportunistic conjunctive rules of the form "triangle load and not smokestack" or "long car and not smokestack" which may represent an opportunistic disjunction. The data from these additional subjects suggest that the tendency to conjoin properties within a category does not eliminate opportunistic conjunctions and disjunctions.

Aside from supporting one of the key assumptions of INDUCE, the results of this experiment show that rule induction is guided by more than simplicity or parsimony. Many of the rules contained more than the minimum number of necessary descriptors, which, in the framework of INDUCE, suggests that "fit" also influences inductive generalizations. These observations are also in accord with a bias toward characteristic descriptions over discriminant descriptions. This preference for fit to data (or, in other words, avoiding excessive generalizations) comes at the cost of simplicity but it has the benefit that the descriptions list the inferences about properties that can be reliably drawn from knowledge of category membership.

Experiment 3

The third experiment used the same trains as the second and also was concerned with rule induction. The difference was that the examples were not presented all at once, but sequentially one by one. The examples were trains and participants had to learn to classify each of the 6 trains as East- or Westbound. At the end of learning, participants were asked for their basis of classification (i.e., the rules they had learned). The main question concerns how the rules will change under this sequential presentation procedure, which places more demands on memory than the simultaneous presentation used in Experiment 2.

In terms of our descriptive model for inductive generalization the learning procedure might make it more difficult to discover properties that are complete or consistent. If a person finds a property that is complete but not consistent (e.g., long for East trains has one counterexample), they might treat the counterexample as an exception and eliminate by describing it in detail. This might lead to a rule like "Trains with long cars are Eastbound except if they have a circle load." Another possibility is that a descriptor might be found which is consistent but not complete (e.g., the descriptor, "Westbound trains are short"). In that event, attention should focus on the remaining West train and one might see a rule like "Westbound trains are *short* or *long with a circular load*."[3] Note that such a rule is different from the rule "Westbound trains are *short or have a circular load* " because it specifically combines circular load with long car. The descriptive model, then, is consistent with

[3]The underlining of various sections of rule statement is designed to facilitate proper parsing, since statements like "short or long with a circular load" are ambiguous.

disjunctive rules but for the trains in Figure 2 at least one part of the disjunction should contain a conjunctive description.

Method

Subjects. The subjects were 20 male and female undergraduates attending Emory University who participated in this experiment in partial fulfillment of course requirements in introductory psychology.

Stimuli. The stimulus materials consisted of the six trains shown in Figure 2 which were individually mounted on index cards. The stimuli were otherwise identical to those used in Experiment 2. For half the subjects the trains on the left side of Figure 2 were in the East category and the train on the right side were in the West category, and for the other half of the subjects this assignment was reversed.

Procedures. Each participant was tested individually. They were told that they would see trains differing in a number of properties and that their task was to learn to correctly classify the trains as Eastbound or Westbound. The individual cards were presented in a random order, subject to the constraints that a given train was never presented twice in a row and a given category never appeared more than four times in a row.

The experimenter first ran through the set of six trains twice and gave the correct category assignment as each card was presented. Thereafter the cards continued to be presented one at a time and the subject said whether they thought the train was in the East or West category and then was told whether they were correct or incorrect. There was a brief pause between every two runs and training continued until a participant was correct for each train in a block of two such runs. When the training criterion was met, the experimenter asked the subject to explain their criterion for classifying the trains as East or West. In addition to this general question, participants were specifically asked if they focused on one of the two categories.

Results

Every participant met the learning criterion and the overall average number of errors to criterion (calling an Eastbound train Westbound or vice versa) was 7.50. The solutions were generally in accord with our general processing model and INDUCE. All but 3 of the 20 participants focused on one of the two categories. The solution types are summarized in Table 6. With two exceptions, the solutions were conjunctive, involved a consistent descriptor disjunctively combined with conjunctive description to make the rule complete, or involved a complete descriptor combined with a conjunctive description of the counter-example to make the rule complete. One person simply memorized the trains and another described a configural property involving openess and brightness. Of the 23 solutions stated, 14 included redundant features. An example of this is the rule that West trains are *short* or *long with a circular load* when "short or circular load" was equally accurate.

Discussion

The main effect of switching to a learning paradigm appeared to be to make it more difficult to discover sets of consistent and complete descriptors. The predominant strategy was to select a single descriptor and narrow it by conjunctively describing the counter-example (in the case of a complete but not consistent descriptor) or to extend it by describing the additional train (in the case of a consistent but not complete description). Furthermore a majority of solutions contained redundant components and there were not cases in which the most simple disjunctive solution was given. Frequently, these redundant components were associated with descriptions that applied to a single train, either to include it or to exclude it. It is not clear whether this form of redundancy differs in any fundamental way from the type of redundancy noted earlier. This pattern of results is consistent with INDUCE and the general processing model outlined earlier.

Experiment 4

Although the results of the second and third experiments were clearcut, they are based on a single set of stimulus materials. This experiment used verbal descriptions of two categories of hypothetical people in the rule induction task. The abstract structure is again such that comparison can be made between conjunctive rules derived from properties that are complete but not consistent and disjunctive rules derived from properties that are consistent but not complete. One reason for anticipating a different pattern of results with verbal materials is that combinations of properties might be much less salient. A second factor varied was whether or not the two properties that could be conjoined into a disjunctive or conjunctive rule were adjacent in the descriptions. Again, nonadjacent descriptions may favor consistency and disjunctive rules because it may be difficult to integrate information that is spatially separated.

Method

Subjects. The subjects were 54 male and female undergraduates attending the University of Illinois who participated in the study in partial fulfillment of course requirements in introductory psychology. Participants were run in groups of 3 to 4 and the experiment lasted about 10 minutes. The subjects were either assigned to a condition where relevant dimensions were adjacent (Adjacent Group, n=30) or nonadjacent (Nonadjacent

Group, n=24).

Stimuli. The stimulus materials consisted of descriptions of two groups of six people placed on a single sheet of paper partitioned by group. Each description consisted of a value on each of six dimensions: Marital Status (Single or Married), Education (M.A. or B.A.), Sports (Golf or Tennis), Music (Rock or Jazz), Employment (Self-employed or Corporation) and Hobby (Painting, Photography, or Ceramics). For four of the six dimensions a given value was true of all members of one class but had two counterexamples in the contrast category, or was true of some (four) members of one class but had no counterexamples. The former properties have maximal category validity and the latter have maximal cue validity.

It was possible to combine two complete but not consistent descriptors to form a valid conjunctive rule or to combine two consistent but not complete descriptors into a disjunctive rule. The relevant dimensions involved in either type of conjoining were either adjacent (first and second, third and fourth, or fifth and sixth) or nonadjacent (first and fourth, second and fourth, third and fifth, second and fifth). An example from the Nonadjacent condition is shown in Table 7. The two possible rules of central interest for the left category in Table 7 are "M.A. and Rock" versus "Golf or Self-employed" and for the right category are "Tennis and Corporation" versus "B.A. or Jazz." Although each participant saw the same abstract structure, several different randomizations of positions and properties were employed to realize this abstract structure.

Procedure. Participants were given the sheet of twelve descriptions and their left-right grouping and asked to read them over carefully. They were told to come up with a basis for classifying the two groups that could be used to describe the groups and to determine the correct category membership for any new descriptions. For the Adjacent Group the pair of consistent descriptors or the pair of complete descriptors was always adjacent and for the Nonadjacent Group there was at least one intervening descriptor between the two members of a potential pair (see Table 7).

Results

The results were generally the same as for the second experiment — there was a strong preference for conjunctive rules based on complete but non consistent descriptors over disjunctive rules derived from consistent but not complete descriptors. In the Adjacent Group 21 people gave a conjunctive rule and only 2 a disjunctive rule. Of the remaining six people, three simply listed relevant properties, one gave a very complex (and incorrect) rule and two integrated the dimensions into a composite personality statement (e.g., dependent versus independent people). All together, there were 34 conjunctive rules given and only 5 disjunctive rules. For 8 of the 34 conjunctive rules additional properties were mentioned, again suggesting that rules are not strongly constrained by simplicity. On three occasions only a single property was mentioned for a rule and in each case this was a complete but not consistent property.

The rule induction task proved to be more difficult for the Nonadjacent Group but the main pattern of results was the same. Eleven of the people gave incomplete rules which can be further classified as consisting of a necessary feature (two people), a sufficient feature (one person), and both a necessary and sufficient feature (five people). Three people integrated the dimensions into a composite personality statement and the last person gave no rule. At the level of rules all 20 were conjunctive and 5 of these included an additional property.

Discussion

The switch from simple trains to verbal descriptions of people did not change the preference for conjunctive rules based on complete properties over disjunctive rules based on consistent properties. Furthermore, although the Nonadjacent condition dramatically reduced the proportion of people coming up with a successful rule (from 80 percent to 46 percent), it dod not diminish this preference for conjunctive over disjunctive rules (it went from 88 percent to 100 percent).

This evidence that category validity plays an important role in rule induction apparently has at least modest generality. We found no evidence that components are ordered by information value alone or cue validity alone and then developed into rules.

Table 1.

Bases for classification on the initial clustering in Experiment 1. The number in the table refers to the number of participants using a particular clustering strategy. The numbers in parenthesis under partitioning bases give the modal category sizes associated with each strategy. Thus (3)(3)(4) refers to partitioning into three categories of respective sizes of 3, 3, and 4 trains.

Basis for Partitioning		Standard Method (# of people)	Informed Method (# of people)
Single Property			
Number of Cars (3) (3) (4)		16	19
Color of Wheels (1) (4) (5)		3	5
Car Shape (8) (2)		3	0
Load Shape (1) (4) (5)		1	0
Number of Loads		0	2
	Total Single	23	26
Conjunction of Properties			
Number of cars and engine wheel color (1)(1)(2)(2)(2)(2)		1	3
Number of cars and car shape (3)(3)(2)(2)		1	1
Number of loads per car (4)(6)		1	0
Car position and type of load (4)(6)		5	0
Car position and car shape (2)(8)		1	1
	Total Conjunction	9	5
	Total Participants	32	31

Table 2.

Basis for classification on the constrained clustering task where two equal-sized groupings were created in Experiment 1. The numbers refer to the number of participants using a particular strategy.

Basis for Partitioning	Condition	Standard Method (# of people)	Informed Method (# of people)
Single Property			
Color of Engine Wheels		13	8
Wheel color		1	4
Load Shape			
Shape vs Not		4	2
All Different vs Not		1	1
Number of Loads		1	3
Car Shape			
Rounded vs Straight		1	1
Other			
Number of Cars (loosely)		1	0
Simple vs Complex Trains		1	0
	Total Single	23	19
Conjunction of Properties			
Load Shape and Car position		4	4
Car Shape and Load Shape		4	2
Car Length and Car Top		0	2
Number of Cars and Car Top		0	1
Number of Cars & Engine Wheel Color		0	2
Load Shape & Wheel Color		0	1
	Total Conjunction	8	12
Disjunction of Properties			
Number of Cars or Wheel Color		1	0
	Total People	32	31

Table 3.
Basis for classification on the free sorting with an "else" category in Experiment 1. The numbers in the Table refer to the number of participants using a particular strategy.

Basis for Partitioning		Standard Method (# of people)	Informed Method (# of people)
Single Property			
Number of Cars		2	0
Wheel Color		1	10
Number of Loads		1	1
Load Shape		1	1
Car Shape		0	1
Number of particular Load Shapes		3	1
Number of Wheels		2	0
	Total Single	10	14
Conjunction of Properties			
Car Position and Shape		6	1
Car Top and Wheel Color		1	1
Car Position and Load		2	1
Car Shape and Loads		3	3
Car Shape and Number of Wheels		0	1
Car Length and Shape		1	1
Number of Cars and Shape		1	1
Number of Cars and Loads		0	1
Number of Cars and Wheel Color		0	2
Combinations of Load Shapes		3	4
Load Shape in Same vs Adjacent Cars		3	0
Number of Cars and Shapes Order		0	1
Combinations of Car Shapes		2	0
	Total Conjunctive	22	17
	Total People	32	31

Table 4.

Breakdown of solutions to rule induction task in Experiment 1. The numbers in the Table refer to the number of participants giving a particular type of rule.

Solution Type		Standard Method	Informed Method
Simple Property			
Number of Different Loads (East: 3 or more different loads)		2	2
Conjunction of Properties			
Positive Features only (e.g. East: triangle load and 3 or more loaded cars)		2	2
With Negative Features (e.g. East: 3 or more cars and triangle load and not jagged car top)		8	7
	Conjunction Total	10	9
Disjunction of Properties			
Simple (e.g. West: two cars or jagged top)		12	5
Disjunction of Conjunction Positive properties only (e.g. 2 cars or long cars and 2 white wheels)		1	9
Negative properties included (e.g. East: at least 1 black wheel on engine and not 3 circular loads or (diamond shape load and not black wheels)		4	1
	Disjunction Total	17	15
Mixed Types (e.g. East; conjunctive, West; disjunctive)		1	2
Other (e.g. partial rules, descriptions of the various trains		2	3
	Total People	32	31

Table 5.

Bases for classification provided by participants in Experiment 2. The numbers refer to the number of participants for a given classification basis. The numbers in parentheses are the number of descriptions that would not successfully classify the trains.

	East	West
Rule		
Conjunctive		
Simple	17 (2)	19 (4)
Redundant	17	4
Disjunctive	0	8 (1)
Both	2	0
Description		
Common properties	20 (8)	24 (17)
Distinctive Properties	4	3 (1)
Common and Distinctive Properties		7 (1)
None	2	1

Table 6.

Bases for classification provided by participants in Experiment 3. The numbers refer to number of solutions for a given type and since 3 of the 20 participants said they had paid equal attention to both categories the total number of solutions is 23. The underlinings for the rule statements are intended to help parse the rule components.

Basis for Classification	Number of Solutions
Conjunctive rule	
Simple (e.g. East: Long and triangle load)	6
Redundant (e.g. East: Two wheels and triangle load and not short)	1
Consistent descriptor plus conjunction (e.g. East: open top or closed top and clear rear wheels)	11
Complete descriptor plus conjunction to eliminate counterexample (e.g. East: Long cars and not long with dark wheels and a circular load)	3
Memorized Individual Trains	1
Configural ("East trains looked more open and bright")	1

Table 7.

An example of the classification materials used in Experiment 4. Each cluster of descriptors corresponds to an individual. In this example the dimensions relevant to a simple disjunctive or conjunctive rule are nonadjacent (1st and 4th or 2nd and 5th).

Category A	Category B
Married	Married
M.A.	B.A.
Golf	Tennis
Rock	Jazz
Self–employed	Corporation
Ceramics	Ceramics
Married	Single
M.A.	B.A.
Tennis	Golf
Rock	Rock
Self–employed	Corporation
Painting	Painting
Married	Single
B.A.	B.A.
Golf	Tennis
Rock	Rock
Self–employed	Corporation
Photography	Photography
Married	Single
M.A.	B.A.
Golf	Tennis
Rock	Jazz
Corporation	Corporation
Ceramics	Ceramics
Married	Single
M.A.	B.A.
Tennis	Golf
Rock	Jazz
Corporation	Corporation
Painting	Painting
Married	Married
B.A.	B.A.
Golf	Tennis
Rock	Jazz
Self–employed	Corporation
Photography	Photography

VI. General Discussion

The set of experiments in the present paper form a coherent pattern. The first study found that people's category constructions and rule inductions were quite similar to those associated with two corresponding AI inductive learning and generalizaton programs, CLUSTER and INDUCE. People appear to set out to find descriptors that will span the target category without applying to examples from contrasting categories. If an assertion is consistent (covers no counter-examples) but not complete (does not span the target category), it is retained, attention shifts to the members of the target category not covered by the original assertion, and new assertions are sought that are consistent and complete for the reduced set (i.e., they form a disjunction). This is precisely how INDUCE works. The second major possibility is that an assertion will be complete but not consistent. In this event, people focus on the counter-examples and attempt to eliminate them by specializing their description, which can be done by negating properties that are true of the counter-examples but not for the positive examples. In support of this interpretation, negations (e.g., not triangular) appear almost exclusively with conjunctive rules. One may mention that by appropriate parameter settings INDUCE may be able to capture this emphasis on completeness, though in its current implementation INDUCE places more stress on consistency in the process of developing solutions. Finally, the similarity of performance of participants given the list of descriptors and those who were not, provides support for the selective and constructive generalization rules associated with INDUCE. People did not confine themselves to the original descriptive language and neither do CLUSTER and INDUCE.

The second study showed that people are far more likely to develop conjunctive rules with complete but not consistent descriptors than disjunctive rules with consistent but not complete descriptors. In addition, many rules derived by subjects contained redundant components. This observation is consistent with the idea that degree of "fit" to data and not just simplicity influences people's inductive generalizations. The third study used a learning procedure, and again component completeness (category validity) appeared to be more important than component consistency (cue validity). No participant gave a simple disjunctive rule. Instead, rules took one of three forms: (1) simple conjunctive, (2) disjunctive based on a consistent but not complete description supplemented by a description of the remaining example (e.g., "short or long with a circular load"), and (3) conjunctive based on a complete but not consistent descriptor supplemented with a description of the remaining counter-example. Again, a majority of the rule statements included more than the minimum necessary descriptions. The fourth study indicated that category validity continues to be important when the stimulus materials consist of verbal descriptions of people.

Relation to AI Models

We have concentrated on the program INDUCE for reasons given in the introduction. As a psychological process model INDUCE fares rather well. Although it manifests a bias for conjunctive solutions it does allow for disjunctive solutions of the form we have been referring to as "opportunistic disjunctions." It's main shortcoming as a psychological model is that it does ot contain an algorithm for "opportunistic conjunctions" where complete but not consistent rules are modified by negating properties of counter-examples. Although both types of opportunistic rules lack the elegance of a simple conjunctive description they do offer certain advantages. First of all, most concepts probably do not have singly necessary and jointly sufficient properties (see Medin & Smith, 1984, for a recent review) and, therefore, would allow for simple conjunctive rules. A second, related reason for allowing for opportunistic rules in AI programs is that it would provide better immunity to noisy or partially inconsistent data. The first part of opportunistic rules would not be affected by a few inconsistencies or counter–examples.

Other AI programs fare less well as psychological models. In part, this is to be expected in that they were not intended to be models for human rule induction. The reasons why these alternative induction procedures do not mirror the human data are varied. First of all, some programs do not provide for disjunction or constructive generalization rules (e.g., Mitchell, 1977). Although other programs employ constructive generalization rules (e.g. Winston, 1975; Hayes–Roth & McDermott, 1978) they contain no mechanisms for representing disjunctions. Most of the programs that do allow for disjunctions (e.g., Quinlan, 1975, 1979) assume that a discrimination net ordered by information value is developed to construct rules. These programs could not predict the strong preference for conjunctive rules and component category validity over disjunctive rules and component cue validity that was particularly salient in the second and third experiments. Finally, to our knowledge no AI program makes provision for the opportunistic conjunctions that were fairly prevalent in our human rule induction data.

Generality

The generality of the present results is certainly open to question. So far we have sampled from a small set of stimulus materials, procedures, and category structures. Yet to be determined is the extent to which we are studying fairly general processing constraints as opposed to constraints associated with our particular tasks and stimulus materials. Even if we are sanguine with respect to general constraints, we know little about the range of flexibility available to people in comparison with that of INDUCE and CLUSTER. By changing the preference criteria, we can easily cause INDUCE to give simplicity priority over fit. It would be surprising if people demonstrated this range of flexibility in processing, if only because of time and memory limitations. On the other hand, people may manifest a deeper form of flexibility such that their performance falls outside that available to INDUCE and CLUSTER even when all the parameters in the programs are free to vary.

As one approach to the issue of human flexibility, we have conducted followup work using a rule induction task and employing the trains from the first experiment. The main independent variable was that instead of labeling the trains as East– or Westbound, different labels and cover stories were presented. For example, a participant might be told that the categories were trains run by smugglers versus legal trains, or trains constructed by creative versus uncreative children, or trains that travel in mountainous versus flat terrains. Our preliminary data suggest that these different labels influence rule inductions in systematic ways but these systematic changes are compatible with INDUCE and the general process model. As one example of a change, the mountainous versus flat terrain labels make it much more likely that a participant will come up with the rule that the trains in one category have three or more different loads. In addition, certain salient properties that are readily linked to labels may lead participants to rules suggesting a greater bias toward consistency. For example, when the smuggler category included the train carrying a diamond–shaped load, a participant might give a rule of the form "diamond shaped load load. Finally, for these more meaningful categories, we have some evidence that participants are more likely to tolerate rules which either are incomplete or have counter–examples.

Although one could probably demonstrate that a semantically-rich but syntactically-awkward rule will be preferred to a semantically-impoverished but syntactically-simple rule, such a demonstration is unlikely to constitute a powerful constraint on the generality of the present results. In most domains of interest semantic considerations may narrow down the set of properties which might enter into inductive generalizations but still leave an innumerable set of possible inductions. Among this set, syntactic considerations may play a powerful role. Of course, syntax and semantics may not be orthogonal. In novel domains, syntactic constraints may guide the search for semantically meaningful properties — a complete but not consistent descriptor is a good candidate for a necessary property and a consistent but not complete descriptor may turn out to be a sufficient property.

The Importance of Category Validity

Probably the most striking result was the emergence of category validity as a significant factor in rule inductions. The preference for conjunctive over disjunctive rules in the second and third studies may be seen as deriving from an opportunistic combining of complete but not consistent descriptors. Again, we hasten to add that stating constraints in terms of products or outputs derives from the underlying processing model we have outlined combined with the particular category structures employed. With different processing demands and alternative category structures the same processing model that continues to give an important role to category validity may give rise to a preponderance of disjunctive rather than conjunctive rules (e.g., Experiment 1).

There is still the question of whether these results on category validity have any significant generality. We think there are two strong reasons for thinking that they do. One is that our tasks are heavily biased toward discriminating rather than characterizing the categories and, therefore, heavily biased toward cue validity. Still, category validity emerged as a very significant factor and if that is true in the present circumstance, it ought to be even more true in the more general case where characterizing and understanding categories are more important. The second support for generality derives from some related research in diagnostic classification.

One domain that may be particularly relevant to the present studies is the diagnostic classification associated with medical problem solving. Some recent research in this area can be interpreted as supporting the importance of category validity. One fairly elaborate study by Fox (1980) employed a task where an initial symptom was presented and the person performing in the task could either make a diagnosis or perform tests for additional symptoms. Both the symptoms and diseases were realistic and the participants were third, fourth, and fifth year medical school students. All symptoms were associated with more than one disease and the probability of a symptom given a disease could and did vary from disease to disease. The medical students received extensive training on this task until their performance was asymptotic. Fox (1980) analyzed the sequential tests for symptoms in terms of a production system model and he did not directly consider the role of cue and category validity. There was one case, however, where the presenting sysmptom narrowed down the set of possible diseases to two and where some of the additional symptoms had approximately same informative value but varied in category validity. Specifically, one sympton was associated with one disease half the time (probability of symptom/disease = .50) and never appeared with the other disease, whereas another symptom was associated with the first disease three–fourths of the time and appeared with the second one–fourth of the time. Because the diseases did not appear equally often the first symptom had a slightly greater information value but the second had a higher category validity. The results showed that the symptom with the higher category validity was tested for far more frequently than the other (33 out of 41 occasions). This suggests the influence of category validity is not confined by meaningless stimuli, short tasks, and naive subjects.

A related study with first-year house officers (Wolf, Gruppen, & Billi, 1985) also suggests the cue validity is not the sole factor determining diagnostic classification. Wolf et al. used a highly simplified task but one that tends to underline their results. The medical personnel were presented with cards labeled with two diseases (A and B) and two symptoms and given information about the prevalence of one of the symptoms in one of the disease categories. Participants were allowed to select one of the other three sources of information. To determine cue validity, one would need to test for the prevalence of the given symptom in the alternative disease category. Only a minority of the house officers (24 percent) consistently selected this optimal diagnostic information. Most of the nonoptimal choices were testing for the alternative symptom in the initial disease category. In general, if physicians organize their medical knowledge in terms of diseases and the likelihood that different symptoms are associated with them, then category validity may play a more important role in induction and diagnostic reasoning. Eddy's (1982) recent review of probabilistic reasoning in clinical medicine showing that people often act as if cue validity is the same as category validity is consistent with this suggestion.

Relative emphases on cue versus category validity have different implications for which procedural variations should optimize learning. Consider a classification learning task involving two categories where in the initial phases of learning the examples from alternative categories are either randomly intermixed or blocked by category (i.e., all the examples of one category appear before the examples of the other category). To determine cue validity, one needs to have a contrast category so mixing examples should facilitate learning. On the other hand, acquiring information about category validities ought to be facilitated when examples are blocked by categories. Although there is not a great deal of evidence bearing on the relative effectiveness of these two training procedures the data which do exist show that learning is considerably more efficient under blocked rather than mixed presentation for both rule-based (Whitman & Garner 1963) and fuzzy categories (Murphy, 1984).

The present findings, along with results from the studies just reviewed, undermine the idea that people classify and form inductive generalizations by computing cue validity or information value and then developing something like a discrimination net model. On the other hand, cue validity is not totally ignored. For example, although the rules given for the trains in Experiment 2 were often redundant, they did not include properties that were true of all members of both categories (i.e., those with zero cue validity). In addition, one might readily imagine that rule redundancy could readily be decreased (or increased) by different instructions or task demands. The results do suggest, however, that category validity plays a more significant role than implied by previous accounts of rule induction. Given that this pattern of results apparently holds for medical diagnosis and classification learning, where the emphasis is on discrimination, it ought to be even more powerful for natural object categories where the emphasis is often on the inferences which can be derived from knowledge of category membership.

Implication for Constraints

The models we have been discussing suggest some fairly general biases or constraints on rule inductions. If we take as our starting point the vague notion that the only constraint needed is that people prefer simple rules to complex rules, then we can claim considerable progress. First of all, simplicity is not the whole story. Whether we define simplicity in terms of number of operators or complexity of descriptors, our experiments demonstrate inductive generalizations are influenced by factors other than simplicity. People show strong preferences among equally simple rules and their rules very frequently contain more than the minimal content needed to discriminate between the categories. And it is not the case that this lack of parsimony arises from people's failures to discover simple rules. In a large number of cases people stated rules that could be made more simple by dropping conditions. These and other observations support the idea that people's inductions are also influenced by the concept of *fit* or degree of specificity. The concept of fit implies that rule inductions may tend toward greater specificity than the most simple and general discriminating rules. One could think of this emphasis on fit as protecting the system from drawing generalizations that are too broad and difficult to recover from. Also, the fit biases descriptions toward including the maximum number of correlated descriptors in one conjunctive statement. This bias toward correlated attributes allows for convenient representation of inferences which may be drawn from category membership and may set the stage for causal linkages among descriptors.

Our process models also embody other constraints. According to these models, one cannot specify independent of particular structures whether conjunctive or disjunctive rules are more likely to predominate. It is the case, however, that processes such as initially searching for completeness and then modifying descriptions to insure consistency will provide powerful biases in rule inductions and allow one to make predictions about the relative preponderance of disjunctive and conjunctive rules for any particular structure. That is, *the* *constraints are embodied in the process model for performance and not in some abstract statement of the general difficulty of different types of rules.*

The notion that constraints are embodied in process models suggests a future direction of research. The difference in rule statement between the second and third studies versus the fourth study shows that demands on memory associated with learning procedures provides an additional source of constraints. A more detailed model for human rule induction that included a limited working memory would provide a framework for exploring additional constraints on human rule induction.

APPENDIX

Table A.1

List of solutions to the East–West rule problem in Experiment 1. An asterisk indicates that the rule will not perfectly partition the trains.

Solution Type	Standard Method (# people)	Informed Method (# people)
Simple Property		
1) East: 3 or more different loads	2	2
Conjunctions of Properties		
1) East: triangle load and 3 or more loaded cars	1	0
2) East: short car and closed top	1	1
3) West: less than 3 load types and last car with 2 wheels	0	1
4) East: 3 or more cars and triangle load and not jagged top	1	2
5) East: 3 or more cars and not jagged top	2	0
6) East: 3 or more cars and triangle load and not triangle load in non first rounded car	1	0
7) East: triangle load and not 3 circle loads and not jagged top	2	1
8) East: 3 or more cars in first car open and not jagged top and not two identical cars	0	1
9) East: 3 or more cars and all loaded and no two cars identical	0	1
10) West: Long car and white wheel and not carrying triangle load	0	1
11) West: Long car and white wheels and (no loads or two small rectangles three circles or a large rectangle as loads)	0	1
12) East: Triangle load and not in last car and nonengine wheels all same color	1	0
13) West: Two or fewer black wheels and not hexagon or oval car and not short rectangular car with closed top	1	0
Disjunction of Properties		
1) West: 2 cars or jagged top	10	3
2) West: 2 card or (long cars and		

2 white wheels)	1	1
3) West: not triangle load or (triangle and 2 white wheels)	1*	0
4) West: two cars or 2 identical cars or only 1st car wheels black	1	0
5) West: (Clear engine wheels and not an oval shaped car) or three circular loads	1	0
6) West: triangle in 2nd car or second with closed top	1*	0
7) East: (At least one black engine wheel and not 3 circular load) or diamond shaped load	1	1
8) West: No dark engine wheels or (not white engine wheels and circular loads in first car)	1	0
9) West: Not triangle load or (triangle load and last two cars of same type open)	0	1*
10) East: (triangle load in first car and first car open top) or (triangle load in second car and second car closed top)	0	1
11) East: (four cars and one with closed top) or (3 cars and two with closed top)	0	1*
12) East: (Long cars and black wheels) or (triangle load and closed tops) or (triangle load in first car)	0	1
13) East: (3 cars with all white car wheels and black engine wheels) or (4 cars with all black wheels or diamond load) or (3 cars and inverted triangle load)	0	1
14) East: (at least one engine wheel colored and first car open) or (all wheels white and second car with triangle load)	0	1
15) East: (at least one engine wheel colored and three or four cars) or (engine wheels white and four cars and one oval shaped car)	0	1
16) East: (triangle load and open top in first car) or triangle load and closed top in second car)	0	1
17) West: (Engines with all white wheels) or (black engine wheels and open		

last car with triangle load)	0	1*
18) West: (middle car closed with clear wheels) or two cars adjacent to middle have wheels of different color)	0	1
Mixed Types		
1) East: four or more cars and load and not jagged top	0	2
2) East: four cars and open top West: two cars or (jagged top and long rectangular car and short rectangular car)	1	0
Other		
1) Single property plus very detailed descriptions of various trains	1*	2
2) Detailed descriptions of either East or West trains	1	1*

References

Beach, L. R. (1964). "Cue probabilism and inference behavior." *Psychological Monographs, 78* (5) (Whole No. 582).

Bourne, Jr., L. E. (1974). "An inference model of conceptual rule learning." In R. Solso (Ed.)., *Theories in cognitive psychology,* pp. 231–256, Washington DC: L. Erlbaum Associates.

Dietterich, T., & Michalski, R. (1981). "Inductive learning of structural descriptions." *Artificial Intelligence, 16.* (3) pp. 257–294.

Dietterich, T., & Michalski, R. (1983). "A comparative review of selected methods for learning from examples." In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine Learning.* pp. 41–81. Palo Alto, CA: Tioga Publishing.

Dominowski, R. L., & Wetherick, N. E. (1976). "Inference processes in conceptual rule learning." *Journal of Experimental Psychology: Human Learning and Memory,* 2, pp. 1–10.

Hunt, E. (1983). "On the nature of intelligence." *Science,* 219, pp. 141–146.

Imai, S. (1966). "Classification of sets of stimuli with different stimulus characteristics and numerical properties. *Perception & Psychophysics,* 1, pp. 48–54.

Kuhn, T. S. (1962). "The historical structure of scientific discovery." *Science,* 136, pp. 760–764.

Medin, D. L. (1983a). "Linear separability and concept naturalness." in *Proceedings of the 2nd International Conference on Machine Learning.* pp. 213–217. Champaign, IL: University of Illinois Press.

Medin, D. L. (1983b). "Structural principles in categorization." in B. Shepp & T. Tighe (Eds.), *Interaction: Perception Development and Cognition.* pp. 203–230. Hillsdale, NJ: Erlbaum.

Medin, D. L., & Smith, E. E. (1984). "Concepts and concept formation." *Annual Review of Psychology,* 35, pp. 113–138.

Mervis, C. B., & Rosch, E. (1981). "Categorization of natural objects." *Annual Review of Psychology,* 32, pp. 89–115.

Michalski, R. B. (1980). "Pattern recognition as rule–guided inductive." *IEEE Transactions on Pattern Analysis and Macine Intelligence.* Vol. PAMI–2, No. 4, pp. 349–361.

Michalski, R. S. (1983a). "A theory and methodology of inductive learning." *Artificial Intelligence,* 20, pp. 111–161.

Michalski, R. S., (1983b). "A theory and methodology of inductive learning." in R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach,* Vol 1, pp. 83–134. Palo Alto, CA: Tioga Publishing.

Michalski, R. S., & Chilausky, R. L. (1980). "Learning by being told and learning from examples: An experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis." *Policy Analysis and Information Systems*, 4, pp. 125–160.

Michalski, R. S., & Stepp, R. E. (1983). "Learning from observation Conceptual clustering." in R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach*, Vol 1, pp. 331–363. Palo Alto, CA: Tioga Publishing.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1983). *Machine Learning: An Artificial Intelligence Approach*, Vol 1, Palo Alto, CA: Tioga Publishing.

Murphy, G. L., & Medin, D. L. (1984). "The role of theories in conceptual coherence." *Psychological Review*, in press.

Neisser, U., & Weene, P. (1962). "Hierarchies in concept attainment." *Journal of Experimental Psychology*, 64, pp. 640–645.

Pinker, S. (1979). "Formal models of language learning." *Cognition*, 7, pp. 217–283.

Reznick, J. S., & Richman, C. L. (1976). "Effects of class complexity, class frequency, and pre-experimental bias on rule learning." *Journal of Experimental Psychology: Human Learning and Memory*, 2, pp. 774–782.

Rosch, E. (1975). "Universals and cultural specifics in human categorization." in R. Brislin, S. Bochner, & W. Lonner (Eds.), *Cross- cultural perspectives on learning*. New York: Halsted Press.

Rosch, E., & Mervis, C. B. (1975). "Family resemblances: Studies in the internal structure of categories." *Cognitive Psychology*, 7, pp. 573–605.

Stepp, R. S., & Michalski, R. S. (1984). "Learning hierarchical descriptions: Conceptual clustering of structured objects." in R. S. Michalski, J. G. Carbonell, & T. MN. Mitchell (Eds.). *Machine Learning: An Artificial Intelligence Approach*, Vol 1, Palo Alto, CA, Tioga Publishing.

Tversky, A. (1972). "Elimination by aspects: A theory of choice." *Psychological Review*, 79, pp. 281–299.

Tversky, A. (1977). "Features of similarity." *Psychological Review*, 84, pp. 327–352.

Wason, P. C., & Johnson-Laird, P. N. (1972). "Psychology of reasoning Structure and content." Cambridge, MA: Harvard University Press.

BIBLIOGRAPHIC DATA SHEET	1. Report No. UIUCDCS-F-86-952	2. ISG 86-1	3. Recipient's Accession No.

4. Title and Subtitle

CONSTRAINTS & PREFERENCES IN INDUCTIVE LEARNING
An Experimental Study Comparing Human and Machine Performance

5. Report Date
January 1986

6.

7. Author(s)

Douglas L. Medin, William D. Wattenmaker & Ryszard S. Michalski

8. Performing Organization Rept. No.

9. Performing Organization Name and Address

Departments of Psychology and
Computer Science
University of Illinois
Urbana-Champaign, Illinois

10. Project/Task/Work Unit No.

11. Contract/Grant No.

12. Sponsoring Organization Name and Address

National Institute of Mental Health, Rockville, MD
National Science Foundation, Washington, DC
Office of Naval Research, Arlington, VA
Defense Advanced Research Project Agency, Arlington, VA

13. Type of Report & Period Covered

14.

15. Supplementary Notes

16. Abstracts

Four experiments examined constraints associated with category construction or sorting and the learning decision rules from category partitionings. The candidates constraints were derived from rule learning experiments in cognitive psychology and compared to those used by two inductive learning programs, INDUCE and CLUSTER, from artificial intelligence (AI) research. The results showed that people's category constructions and rule inductions were generally similar to those obtained from programs CLUSTER and INDUCE. One difference was that people tended to emphasize category vlidity (probability that an entity is a member of a category given that it has some property) to a greater extent than did the AI programs. More generally, the results show that human inductive generalizations tend toward greater specificity than would be expected if conceptual simplicity were the key constraints on inductions. This bias toward specificity may be due to the fact that this criterion both maximizes inferences that may be drawn from category membership and protects rule induction systems from developing overgeneralizations.

17. Key Words and Document Analysis. 17a. Descriptors

Inductive Learning
Concept Acquisition
Category Construction
Classification
Clustering
Machine Learning
Cognitive Modeling of Learning

17b. Identifiers/Open-Ended Terms

17c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 67
	20. Security Class (This Page) UNCLASSIFIED	22. Price

FORM NTIS-35 (10-70) USCOMM-DC 40329-P71