TODD R. DAVIES

DETERMINATION, UNIFORMITY, AND RELEVANCE: NORMATIVE CRITERIA FOR GENERALIZATION AND REASONING BY ANALOGY

INTRODUCTION: THE IMPORTANCE OF PRIOR KNOWLEDGE IN REASONING AND LEARNING FROM INSTANCES

If an agent is to apply knowledge from its past experience to a present episode, it must know what properties of the past situation can justifiably be projected onto the present on the basis of the known similarity between the situations. The problem of specifying when to generalize or reason by analogy, and when not to, therefore looms large for the designer of a learning system. One would like to be able to program into the system a set of criteria for rule formation from which the system can correctly generalize from data as they are received. Otherwise, all of the necessary rules the agent or system uses must be programmed in ahead of time, so that they are either explicitly represented in the knowledge base or derivable from it.

Much of the research in machine learning, from the early days when the robot Shakey was learning macro-operators for action (Nilsson, 1984) to more recent work on chunking (Rosenbloom and Newell, 1986) and explanation-based generalization (Mitchell et al., 1986), has involved getting systems to learn and represent explicitly rules and relations between concepts that could have been derived from the start. In Shakey's case, for example, the planning algorithm and knowledge about operators in STRIPS were jointly sufficient for deriving a plan to achieve a given goal. To say that Shakey "learned" a specific sequence of actions for achieving the goal means only that the plan was not derived until the goal first arose. Likewise, in explanation-based generalization (EBG), explaining why the training example is an instance of a concept requires knowing beforehand that the instance embodies a set of conditions sufficient for the concept to apply, and chunking, despite its power to simplify knowledge at the appropriate level, does not in the logician's terms add knowledge to the system.

The desire to automate the acquisition of rules, without programming them into the system either implicitly or explicitly, has led to a good

D. H. Helman (ed.), Analogical Reasoning, 227–250.

deal of the rest of the work in symbolic learning. Without attempting a real summary of this work, it can be said that much of it has involved defining heuristics for inferring general rules and for drawing conclusions by analogy. For example, Patrick Winston's program for learning and reasoning by analogy (Winston, 1980) attempted to measure how similar a source and target case were by counting equivalent corresponding attributes in a frame, and then projected an attribute from the source to the target if the count was large enough. In a similar vein, a popular criterion for enumerative induction of a general rule from instances is the number of times the rule has been observed to hold. Both types of inference, although they are undoubtedly part of the story for how people reason inductively and are good heuristic methods for a naive system,[1] are nonetheless frought with logical (and practical) peril. In reasoning by analogy, for example, a large number of similarities between two children does not justify the conclusion that one child is named "Skippy" just because the other one is. First names are not properties that can be projected with any plausibility based on the similarity in the childrens' appearance, although shirt size, if the right similarities are involved, can be. In enumerative induction, likewise, the formation of a general rule from a number of instances of co-occurrence may or may not be justified, as Nelson Goodman's well-known unprojectible predicate "grue" makes very clear (Goodman, 1983). So in generalizing and reasoning by analogy we must bring a good deal of prior knowledge to the situation to tell us whether the conclusions we might draw are justified. Tom Mitchell has called the effects of this prior knowledge in guiding inference the inductive "bias" (Mitchell, 1980).

A LOGICAL FORMULATION OF THE PROBLEM OF ANALOGY

Reasoning by analogy may be defined as the process of inferring that a *conclusion* property Q holds of a particular situation or object T (the *target*) from the fact that T shares a property or set of properties P with another situation/object S (the *source*) which has property Q. The set of common properties P is the *similarity* between S and T, and the conclusion property Q is *projected* from S onto T. The process may be summarized schematically as follows:

$$\frac{\begin{array}{l} P(S) \wedge Q(S) \\ P(T) \end{array}}{Q(T).}$$

The form of argument defined above is nondeductive, in that its conclusion does not follow syntactically just from its premises. Instances of this argument form vary greatly in cogency. As an example, Bob's car and Sue's car share the property of being 1982 Mustang GLX V6 hatchbacks, but we could not infer that Bob's car is painted red just because Sue's car is painted red. The fact that Sue's car is worth about $3500 is, however, a good indication that Bob's car is worth about $3500. In the former example, the inference is not compelling; in the latter it is very probable, but the premises are true in both examples. Clearly the plausibility of the conclusion depends on information that is not provided in the premises. So the justification aspect of the logical problem of analogy, which has been much studied in the field of philosophy (see, e.g. Carnap, 1963; Hesse, 1966; Leblanc, 1969; Wilson, 1964), may be defined as follows.

> THE JUSTIFICATION PROBLEM:
> Find a criterion which, if satisfied by any particular analogical inference, sufficiently establishes the truth of the projected conclusion for the target case.

Specifically, this may be taken to be the task of specifying background knowledge that, when added to the premises of the analogy, makes the conclusion follow soundly.

It might be noticed that the analogy process defined above can be broken down into a two-step argument as follows: (1) From the first premise $P(S) \wedge Q(S)$, conclude the *generalization* $\forall x\, P(x) \Rightarrow Q(x)$, and (2) instantiate the generalization to T and apply modus ponens to get the conclusion $Q(T)$. In this process, only the first step is nondeductive, so it looks as if the problem of justifying the analogy has been reduced to the problem of justifying a single-instance inductive generalization. This will in fact be the assumption henceforth — that the criteria for reasoning by analogy can be identified with those for the induction of a rule from one example. This amounts to the assumption that a set of similarities judged sufficient for projecting conclusions from the source to the target would remain sufficient for such a

projection to any target case with the same set of similarities to the source. There are clearly differences in plausibility among different single-instance generalizations that should be revealed by correct criteria. For example, if inspection of a red robin reveals that its legs are longer than its beak, a projection of this conclusion onto unseen red robins is plausible, but projecting that the scratch on the first bird's beak will be observed on a second red robin is implausible. However, the criteria that allow us to distinguish between good and bad generalizations from one instance cannot do so on the basis of many of the considerations one would use for enumerative induction, when the number of cases is greater than one. The criteria for enumerative induction include (1) whether or not the conclusion property taken as a predicate is "entrenched" (unlike 'grue', for instance) (Goodman, 1983), (2) how many instances have confirmed the generalization, (3) whether or not there are any known counterexamples to the rule that is to be inferred, and (4) how much variety there is in the confirming instances on dimensions other than those represented in the rule's antecedent (Thagard and Nisbett, 1982). When we have information about only a single instance of a property pertinent to its association with another, then none of the above criteria will provide us with a way to tell whether the generalization is a good one. Criteria for generalizing from a single instance, or for reasoning by analogy, must therefore be simpler than those required for general enumerative induction. Identifying those more specialized criteria thus seems like a good place to start in elucidating precise rules for induction.

One approach to the analogy problem has been to regard the conclusion as plausible in proportion to the *amount* of similarity that exists between the target and the source (see Mill, 1900). Heuristic variants of this have been popular in research on analogy in artificial intelligence (AI) (see, e.g. Carbonell, 1983; Winston, 1980). Insofar as these "similarity-based" methods and theories of analogy rely upon a measure over the two cases that is independent of the conclusion to be projected, it is easy to see that they fail to account for the differences in plausibility among many analogical arguments. For example, in the problem of inferring properties of an unseen red robin from those of one already studied, the amount of similarity is fixed, namely that both things are red robins, but we are much happier to infer that the bodily proportions will be the same in both cases than to infer that the unseen robin will also have a scratched beak. It is worth emphasizing that this

is true no matter how well constructed the similarity metric is. Partly in response to this problem, researchers studying analogy have recently adverted to *relevance* as an important condition on the relation between the similarity and the conclusion (Kedar-Cabelli, 1985; Shaw and Ashley, 1983). However, to be a useful criterion, the condition of the similarity P being relevant to the conclusion Q needs to be weaker than the inheritance rule $\forall x\, P(x) \Rightarrow Q(x)$, for then the conclusion in plausible analogies would always follow just by application of the rule to the target. Inspection of the source would then be redundant. So a solution to the logical problem of analogy must, in addition to providing a justification for the conclusion, also ensure that the information provided by the source instance is used in the inference. We therefore have the following.

> THE NONREDUNDANCY PROBLEM:
> The background knowledge that justifies an analogy or single-instance generalization should be insufficient to imply the conclusion given information only about the target. The source instance should provide new information about the conclusion.

This condition rules out trivial solutions to the justification problem. In particular, although the additional premise $\forall x\, P(x) \Rightarrow Q(x)$ is sufficient for the validity of the inference, it does not solve the nonredundancy problem and is therefore inadequate as a general solution to the logical problem of analogy. To return to the example of Bob's and Sue's cars, the nonredundancy requirement stipulates that it should not be possible, merely from knowing that Bob's car is a 1982 Mustang GLX V6 hatchback, and having some rules for calculating current value, to conclude that the value of Bob's car is about $3500 — for then it would be unnecessary to invoke the information that Sue's car is worth that amount. The role of the source analogue (or instance) would in that case be just to point to a conclusion which could then be verified independently by applying general knowledge directly to Bob's car. The nonredundancy requirement assumes, by contrast, that the information provided by the source instance is not implicit in other knowledge. This requirement is important if reasoning from instances is to provide us with any conclusions that could not be inferred otherwise. As was noted above, the rules formed in EBG-like systems are justified, but the instance information is redundant, whereas in systems that use heu-

ristics based on similarity to reason analogically, the conclusion is not inferrable from prior knowledge but is also not justified after an examination of the source.

There has been a good deal of fruitful work on different methods for learning by analogy (e.g., Burstein, 1983; Carbonell, 1983, 1986; Greiner, 1985; Kedar-Cabelli, 1985; Winston, 1980) in which the logical problem is of secondary importance to the empirical usefulness of the methods for particular domains. Similarity measures, for instance, can prove to be a successful guide to analogizing when precise relevance information is unavailable, and the value of learning by chunking, EBG, and related methods should not be underestimated either. The wealth of engineering problems to which these methods and theories have been applied, as well as the psychological data they appear to explain, all attest to their importance for AI. In part, the current project can be seen as an attempt to fill the gap between similarity-based and explanation-based learning, by providing a way to infer conclusions whose justifications go beyond mere similarity but do not rely on the generalization being implicit in prior knowledge. In that respect, there will be suggestions of methods for doing analogical reasoning. The other, perhaps more important, goal of this research has been to provide an underlying *normative* justification for the plausibility of analogy from a logical and probabilistic perspective, and in so doing to provide a general form for the background knowledge that is sufficient for drawing reliable, nonredundant analogical inferences, regardless of the method used. The approach is intended to complement, rather than to compete with, other approaches. In particular is not intended to provide a *descriptive* account of how people reason by analogy or generalize from cases, in contrast to much of the work in cognitive psychology to date (e.g., Gentner, 1983; Gick and Holyoak, 1983). Descriptive theories may also involve techniques that are not logically or statistically sound. The hope is that, by elucidating what conclusions are justified, it will become easier to analyze descriptive and heuristic techniques to see why they work and when they fail.

DETERMINATION RULES FOR GENERALIZATION AND ANALOGICAL INFERENCE

Intuitively, it seems that a criterion that simultaneously solves both

the justification problem and the nonredundancy problem should be possible to give. As an example, consider again the two car owners, Bob and Sue, who both own 1982 Mustang GLX V6 hatchbacks in good condition. Bob talks to Sue and finds out that Sue has been offered $3500 on a trade-in for her car. Bob therefore reasons that he too could get about $3500 if he were to trade in his car. Now if we think about Bob's state of knowledge before he talked to Sue, we can imagine that Bob did *not* know and could not calculate how much his car was worth. So Sue's information was not redundant to Bob. At the same time, there seemed to be a prior expectation on Bob's part that, since Sue's car was also a 1982 Mustang GLX V6 hatchback in good condition, he could be relatively sure that *whatever* Sue had had offered to her, that would be about the value of his (Bob's) car as well, and indeed of *any* 1982 Mustang GLX V6 hatchback in good condition. What Bob knew prior to examining the instance (Sue's car) was some very general but powerful knowledge in a form of a *determination relation*, which turns out to be a solution to the justification and nonredundancy problems in reasoning by analogy. Specifically, Bob knew that the make, model, design, engine-type, condition and year of a car determine its trade-in value. With knowledge of a single determination rule such as this one, Bob does not have to memorize (or even consult) the Blue Book, or learn a complicated set of rules for calculating car values. A single example will tell him the value for all cars of a particular make, model, engine, condition, and year.

In the above example, Bob's knowledge, that the make, model, design, engine, condition, and year determine the value of a car, expresses a determination relation between functions, and is therefore equivalent to what would be called a "functional dependency" in database theory (Ullman, 1983). The logical definition for function G being functionally dependent on another function F is the following (Vardi, 1982):

$$(*) \quad \forall x, y\, F(x) = F(y) \Rightarrow G(x) = G(y).$$

In this case, we say that a function (or set of functions) F *functionally determines* the value of function(s) G because the value assignment for F is associated with a unique value assignment for G. We may know this to be true without knowing exactly which value for G goes with a particular value for F. If the example of Bob's and Sue's cars (Car_B and Car_S respectively) from above is written in functional terms, as follows:

$$\begin{array}{ll} Make(Car_S) = Ford & Make(Car_B) = Ford \\ Model(Car_S) = Mustang & Model(Car_B) = Mustang \\ Design(Car_S) = GLX & Design(Car_B) = GLX \\ Engine(Car_S) = V6 & Engine(Car_B) = V6 \\ Condition(Car_S) = Good & Condition(Car_B) = Good \\ Year(Car_S) = 1982 & Year(Car_B) = 1982 \\ Value(Car_S) = \$3500 & \\ \hline Value(Car_B) = \$3500 & \end{array}$$

then knowing that the make, model, design, engine,condition, and year determine value thus makes the conclusion valid.

Another form of determination rule expresses the relation of one predicate *deciding* the truth value of another, which can be written as:

$$(**) \quad (\forall x\, P(x) \Rightarrow Q(x)) \vee (\forall x\, P(x) \Rightarrow \neg Q(x)).$$

This says that either all P's are Q's, or none of them are. Having this assumption in a background theory is sufficient to guarantee the truth of the conclusion $Q(T)$ from $P(S) \wedge P(T) \wedge Q(S)$, while at the same time requiring an inspection of the source case S to rule out one of the disjuncts. It is therefore a solution to both the justification problem and the nonredundancy problem. We often have knowledge of the form "P decides whether Q applies". Such rules express our belief in the rule-like relation between two properties, prior to knowledge of the direction of the relation. For example, we might assume that either all of the cars leaving San Francisco on the Golden Gate Bridge have to pay a toll, or none of them do.

Other, more complicated formulas expressing determination relations can be represented. It is interesting to note that determination cannot be formulated as a connective, i.e. a relation between propositions or closed formulas. Instead it should be thought of as a relation between predicate *schemata*, or open formulas. In the semantics of determination presented in the next section, even the truth value of a predicate or schema is allowed to be a variable. Determination is then defined as a relation between a *determinant* schema and its *resultant* schema, and the free variables that occur only in the determinant are viewed as the *predictors* of the free variables that occur only in the resultant (the *response* variables). It is worth noting that there may be more than one determinant for any given resultant. For example, one's zip code and capital city are each individually sufficient to determine one's state. In our generalized logical definition of determination (see the section on "Representation and Semantics"), the forms (*) and (**) are subsumed as special cases of a single relation "*P determines Q*", written as $P \succ Q$.

Assertions of the form "P determines Q" are actually quite common in ordinary language. When we say "The IRS decides whether you get a tax refund," or "What school you attend determines what courses are available," we are expressing an invariant relation that reflects a causal theory. At the same time, we are expressing weaker information than is contained in the statement that P formally implies[2] Q. If P implies Q then P determines Q, but the reverse is not true, so the inheritance relation falls out as a special case of determination. That knowledge of a determination rule or of "relevance" underlies preferred analogical inferences seems transparent when one has considered the shortcomings of alternative criteria like how similar the two cases are, or whether the similarity together with our background knowledge logically imply the conclusion. It is therefore surprising that even among very astute philosophers working on the logical justifications of analogy and induction, so much emphasis has until recently been placed on probabilistic analyses based on numbers of properties (Carnap, 1963), or on accounts that conclude that the analogue is redundant in any sound analogical argument (e.g., Copi, 1972). Paul Thagard and Richard Nisbett (Thagard and Nisbett, 1982) speculate that the difficulty in specifying the principles that describe and justify inductive practice has resulted from an expectation on the part of philosophers that inductive principles would be like deductive ones in being capable of being formulated in terms of the syntactic structure of the premises and conclusions of inductive inferences. When, in 1953—54 Nelson Goodman (Goodman, 1983) made his forceful argument for the importance of background knowledge in generalization, the Carnapian program of inductive logic began to look less attractive. Goodman was perhaps the first to take seriously the role and form of semantically-grounded background criteria (called by him "overhypotheses") for inductive inferences. The possibility of valid analogical reasoning was recognized by Julian Weitzenfeld (Weitzenfeld, 1984), and Thagard and Nisbett (Thagard and Nisbett, 1982) made the strong case for semantic (as opposed to syntactic, similarity- or numerically-based) criteria for generalization. In the process both they and Weitzenfeld anticipated the argument made herein concerning determination rules. The history of AI approaches to analogy and induction has largely recapitulated the stages that were exhibited in philosophy. But the precision required for

making computational use of determination, and for applying related statistical ideas, gives rise to questions about the scope and meaning of the concepts that seem to demand a slightly more formal analysis than has appeared in the philosophical literature. In the next section, a general form is given for representing determination rules in first order logic. The probabilistic analogue of determination, herein called "uniformity", is then defined in the following section, and finally the two notions — logical and statistical — are used in providing definitions of the relation of "relevance" for both the logical and the probabilistic cases.

THE REPRESENTATION AND SEMANTICS OF DETERMINATION

To define the general logical form for determination in predicate logic, we need a representation that covers (1) determination of the truth value or polarity of an expression, as in example cases of the form "$P(x)$ *decides* whether or not $Q(x)$" (formula (**) from previous section), (2) functional determination rules like (*) above, and (3) other cases in which one expression in first order logic determines another. Rules of the first form require us to extend the notion of a first order predicate schema in the following way. Because the truth value of a first order formula cannot be a defined function within the language, let us introduce the concept of a *polar variable* which can be placed at the beginning of an expression to denote that its truth value is not being specified by the expression. For example, the notation " $iP(x)$" can be read "whether or not $P(x)$", and it can appear on either side of the determination relation sign "$>$" in a determination rule, as in

$$P_1(x) \wedge i_1 P_2(x) > i_2 Q(x).$$

This would be read, "$P_1(x)$ and whether or not $P_2(x)$ together jointly determine whether or not $Q(x)$", where i_1 and i_2 are polar variables.

As was mentioned above, the determination relation cannot be formulated as a connective, i.e. a relation between propositions or closed formulas. Instead, it should be thought of as a relation between predicate *schemata*, or open formulas with polar variables. For a first order language **L**, the set of predicate schemata for the language may be characterized as follows. If S is a sentence (closed formula or wff) of **L**, then the following operations may be applied, *in order*, to S to generate a predicate schema:

(1) Polar variables may be placed in front of any wffs that are contained as strings in S,
(2) Any object variables in S may be unbound (made free) by removing quantification for part of S, and
(3) Any object constants in S may be replaced by object variables.

All of and only the expressions generated by these rules are schemata of **L**.

To motivate the definition of determination, let us turn to some example pairs of schemata for which the determination relation holds. As an example of the use of polar variables, consider the rule that, being a student athlete, one's school, year, sport, and whether one is female determine who one's coach is and whether or not one has to do sit-ups. This can be represented as follows:

EXAMPLE 1:

$$\begin{aligned}&(Athlete(x) \wedge Student(x) \wedge School(x) = s\\ &\quad \wedge Year(x) = y \wedge Sport(x) = z \wedge i_1\, Female(x))\\ &> (Coach(x) = c \wedge i_2 Sit-ups(x)).\end{aligned}$$

As a second example, to illustrate that the component schemata may contain quantified variables, consider the rule that, not having any deductions, having all your income from a corporate employer, and one's income determine one's tax rate:

EXAMPLE 2:

$$\begin{aligned}&(Taxpayer(x) \wedge Citizen(x, US) \wedge\\ &\quad (\neg \exists d\, Deductions(x, d)) \wedge (\forall i\, Income(i, x) \Rightarrow\\ &\quad Corporate(i)) \wedge Personal\ Income(x) = p)\\ &> (Tax\ Rale(x) = r).\end{aligned}$$

In each of the above examples, the free variables in the component schemata may be divided, relative to the determination rule, into a *case* set $\underline{x}$ of those that appear free in both the *determinant* (left-hand side) and the *resultant* (right-hand side), a *predictor* set $\underline{y}$ of those that appear only in the determinant schema, and a *response* set $\underline{z}$ of those that appear only in the resultant. These sets are uniquely defined for each determination rule. In particular, for example 1 they are $\underline{x} = \{x\}$, $\underline{y} = \{s, y, z, i_1\}$, and $\underline{z} = \{c, i_2\}$; and for example 2 they are $\underline{x} = \{x\}$, $\underline{y} = \{p\}$, $\underline{z} = \{r\}$. In general, for a predicate schema Σ with free variables $\underline{x}$ and $\underline{y}$, and a predicate schema X with free variables $\underline{x}$

(shared with Σ) and $\underline{z}$ (unshared), whether the determination relation holds is defined as follows:

$$\Sigma[\underline{x}, \underline{y}] > X[\underline{x}, \underline{z}]$$
iff
$$\forall \underline{y}, \underline{z}(\exists \underline{x}\ \Sigma[\underline{x}, \underline{y}] \wedge X[\underline{x}, \underline{z}]) \Rightarrow (\forall \underline{x}\ \Sigma[\underline{x}, \underline{y}] \Rightarrow X[\underline{x}, \underline{z}]).$$

For interpreting the right-hand side of this formula, quantified polar variables range over the unary Boolean operators (negation and affirmation) as their domain of constants, and the standard Tarskian semantics is applied in evaluating truth in the usual way (see Genesereth and Nilsson, 1987). This definition covers the full range of determination rules expressible in first order logic, and is therefore more expressive than the set of rules restricted to dependencies between frame slots, given a fixed vocabulary of constants. Nonetheless, one way to view a predicate schema is as a frame, with slots corresponding to the free variables.

USING DETERMINATION RULES IN DEDUCTIVE SYSTEMS

Determination rules can provide the knowledge necessary for an agent or system to reason by analogy from case to case. This is desirable when the system builds up a memory of specific cases over time. If the case descriptions are thought of as conjunctions of well-formed formulas in predicate logic, for instance, then questions about the target case in such a system can be answered as follows:

(1) Identify a resultant schema corresponding to the question being asked. The free variables in the schema are the ones to be bound (the response variables $\underline{z}$).
(2) Find a determination rule for the resultant schema, such that the determinant schema is instantiated in the target case.
(3) Find a source case, in which the bindings for the predictor variables $\underline{y}$ in the determinant schema are identical to the bindings in the target case for the same variables.
(4) If the resultant schema is instantiated in the source case, then bind the shared free variables $\underline{x}$ of the resultant schema to their values in the target case's instantiation of the determinant schema, and bind the response variables to their values in the source case's instantiation of the resultant schema. The well-formed formula thus produced is a sound conclusion for the target case.

Such a system might start out with a knowledge base consisting only of determination rules that tell it what information it needs to know in order to project conclusions by analogy, and as it acquires a larger and larger database of cases, the system can draw more and more conclusions based on its previous experience. The determination rule also provides a matching constraint in searching for a source case. Rather than seeking to maximize the similarity between the source and the target, a system using determination rules looks for a case that matches the target on predictor bindings for a determinant schema, which may or may not involve a long list of features that the two cases must have in common.

A second use of determination rules is in the learning of generalizations. A single such rule, for example that one's species determines whether one can fly or not, can generate a potentially infinite number of more specific rules about which species can fly and which cannot, just from collecting case data on individual organisms that includes in each description the species and whether that individual can fly. So the suggestion for machine learning systems that grows out of this work is that systems be programmed with knowledge about determination rules, from which they can form more specific rules of the form $\forall x\, P(x, Y) \Rightarrow Q(x, Z)$. Determination rules are a very common form of knowledge, perhaps even more so than knowledge about strict implication relationships. We know that whether you can carry a thing is determined by its size and weight, that a student athlete's coach is determined by his or her school, year, sport, and sex. In short, for many, possibly most, outcomes about which we are in doubt, we can name a set of functions or variables that jointly determine it, *even though we often cannot predict the outcome from just these values.*

Some recent AI systems can be seen to embody the use of knowledge about determination relationships (e.g., see Baker and Burstein, 1987; Carbonell, 1986; Rissland and Ashley, 1986). For example, Edwina Rissland and Kevin Ashley's program for reasoning from hypothetical cases in law represents cases along dimensions which are, in a loose sense, determinants of the verdicts. Likewise, research in the psychology and theory of induction and analogy (see, e.g. Nisbett et al.,

1983) has postulated the existence of knowledge about the "homogeneity" of populations along different dimensions. In all of this work, the reality that full, indefeasible determination rules cannot be specified for complicated outcomes, and that many of the determination rules we can think of have exceptions to them, has prompted a view toward weaker relations of a partial or statistical nature (Russell, 1986), and to determination rules that have the character of defaults (Russell and Grosof, 1987). The extension of the determination relation to the statistical case is discussed in the next section on uniformity.

A third use of determination rules is the representation of knowledge in a more compact and general form than is possible with inheritance rules. A single determination rule of the form $P(x, y) \succ Q(x, z)$ can replace any number of rules of the form $\forall x\, P(x, Y) \Rightarrow Q(x, Z)$ with different constants Y and Z. Instead of saying, for instance, "Donkeys *can't* fly," "Hummingbirds *can* fly," "Giraffes *can't* fly," and so forth, we can say "One's species *determines* whether or not one can fly," and allow cases to build up over time to construct the more specific rules. This should ease the knowledge acquisition task by making it more hierarchical.

UNIFORMITY: THE STATISTICAL ANALOGUE OF DETERMINATION

The problem of finding a determining set of variables for predicting the value of another variable is similar to the problem faced by the applied statistician in search of a predictive model. Multiple regression, analysis of variance, and analysis of covariance techniques all involve the attempt to fit an equational model for the effects of a given set of independent (predictor) variables on a dependent (response) variable or vector (see Johnson and Wichern, 1982; Montgomery and Peck, 1982). In each case some statistic can be defined which summarizes that proportion of the variance in the response that is explained by the model (e.g. multiple R^2, ω^2). In regression, this statistic is the square of the correlation between the observed and model-predicted values of the response variables, and is, in fact, often referred to as the "coefficient of determination" (Johnson and Wickern, 1982). When the value of such a statistic is 1, the predictor variables clearly amount to a determinant for the response variable. They are, in such cases, exhaustively relevant to determining its value in the same sense in which a particular schema

determines a resultant in the logical case. But when the proportion of the variance explained by the model is less than 1, it is often difficult to say whether the imperfection of the model is that there are more variables that need to be added to determine the response, or that the equational form chosen (linear, logistic, etc.) is simply the wrong one. In low dimensions (one or two predictors), a residual plot may reveal structure not captured in the model, but at higher dimensions this is not really possible, and the appearance of randomness in the residual plot is no guarantee in any case. So, importantly, the coefficient of determination and its analogues measure *not* the predictiveness of the independent *variables* for the dependents, but rather the predictiveness of the model. This seems to be an inherent problem with quantitative variables.

If one considers only categorical data, then it is possible to assess the predictiveness of one set of variables for determining another. However there are multiple possibilities for such a so-called "association measure". In the statistics literature one finds three types of proposals for such a measure, that is, a measure of the dependence between variables in a k-way contingency table of count data. Firstly, there are what have been termed "symmetric measures" (see Haberman, 1982; Hays and Winkler, 1970) that quantify the degree of dependence between two variables, such as Pearson's index of mean square contingency (Hays and Winkler, 1970). Secondly, there are "predictiveness" measures, such as Goodman and Kruskal's λ (Goodman and Kruskal, 1979), which quantify the proportional reduction in the probability of error, in estimating the value of one variable (or function) of an individual, that is afforded by knowing the value of another. And thirdly, there are information theoretic measures (e.g. Theil, 1970) that quantify the average reduction in uncertainty in one variable given another, and can be intepreted similarly to the predictive measures (Hays and Winkler, 1970). In searching for a statistic that will play the rule in probabilistic inference that is played by determination in logic, none of these three types of association measure appear to be what we are looking for. The symmetric measures can be ruled out immediately, since determination is not a symmetric relation. The predictive and information theoretic measures quantify how determined a variable is by another *relative* to prior knowledge about the value of the dependent variable. While this is a useful thing to know, it corresponds more closely to what in this paper is termed "relevance" (see next section), or the value of the information provided by a variable relative to what we already know.

Logical determination has the property that a schema can contain some superfluous information and still be a determinant for a given outcome; that is, information added to our knowledge when something is determined does not change the fact that it is determined, and this seems to be a useful property for the statistical analogue of determination to have.

So a review of existing statistical measures apparently reveals no suitable candidates for what will hereinafter be called the *uniformity* of one variable or function given the value of another, or the statistical version of the determination relation. Initially we might be led simply to identify the uniformity of a function G given another function F with the conditional probability:

$$Pr\{G(x)=G(y)\,|\,F(x)=F(y)\}$$

for randomly select pairs x and y in our population. Similarly, the uniformity of G given a particular value (property or category) P might defined as:

$$Pr\{G(x)=G(y)\,|\,P(x) \wedge P(y)\},$$

and permutations of values and variables in the arguments to the uniformity function could be defined along similar lines. This possibility is adverted to by Thagard and Nisbett (Thagard and Nisbett, 1982), though they are not concerned with exploring the possibility seriously. If the uniformity statistic is to underlie our confidence in a particular value of G being shared by additional instances that share a particular value of F, where this latter value is newly observed in our experience, then it seems that we will be better off, in calculating the uniformity of G given F, if we conditionalize on randomly chosen *values of F,* and then measure the probability of a match in values for G, rather than asking what is the probability of a match on G given a match on F for a randomly chosen pair of elements in our past experience, or in a population.

An example should illustrate this distinction and its importance. If we are on a desert island and run across a bird of a species unfamiliar to us (say, "shreebles," to use Thagard and Nisbett's term) and we further observe that this bird is green, we want the uniformity statistic to tell us, based on our past experience or knowledge of birds, how likely it is that the next shreeble we see will also be green. Let us say, for illustration, that we have experience with ten other species of birds, and that among these species nine of them are highly uniform with respect to color, but the other is highly varying. Moreover, let us assume that we have had far greater numerical exposure to this tenth, highly variable species, than to the others, or that this species (call them "variabirds") is a lot more numerous generally. Then if we were to define uniformity as was first suggested, sampling at random from our population of birds, we would attain a much lower value for uniformity than if we average over species instead, for in the latter case we would have high uniformities for all but one of our known species and therefore the high relative population of variabirds would not skew our estimate. Intuitively the latter measure, based on averaging over species rather than individuals in the conditional, provides a better estimate for the probability that the next shreeble we see will be green. The important point to realize is that there are multiple possibilities for such a statistic, and we should choose the one that is most appropriate for what we want to know. For instance, if the problem is to find the probability of a match on color given a match on species for randomly selected pairs of birds, then the former measure would clearly be better. Another factor that plays in the calculation when we average over species is the relative confidence we have in the quality of each sample, i.e. the sample size for each value of F. We would want to weigh more heavily (by some procedure that is still to be specified) those values for which we have a good sample. Thus the uniformity statistic for estimating the probability of a match given a new value of F would be the weighted average,

$$U(G\,|\,F)=\frac{1}{p}\sum_{i=1}^{p} w_i \Pr\{G(x)=G(y)\,|\,F(x)=F(y)=P_i\},$$

where p is the number of values P_i of F for which we have observed instances and also know their values for G. In the absence of information about the relative quality of the samples for different values of F, all of the weights w_i would equal 1.

How might we make use of such a statistic in learning and reasoning? Its value is that, under the assumption that the uniformity of the function given another can be inferred by sampling, we can examine a relatively small sample of a population, tabulate data on the subsets of values appearing in the sample for the functions in question, and compute an estimate of the extent to which the value of one function is

determined by the other. This will in turn tell us what confidence we can have in a generalization or inference by analogy based on a value for a predictor function (variable) co-occurring with a value for a response function, when either or both have not been observed before. The experience of most people in meeting speakers of foreign languages provides a good example. In the beginning, we might think, based on our early data, that one's nationality determines one's native language. But then we come across exceptions — Switzerland, India, Canada. We still think that native language is highly *uniform* given nationality, however, because its conditional uniformity is high. So in coming across someone from a country with which we are not familiar, we can assume that the probability is reasonably high that whatever language he or she speaks is likely to be the language that a randomly selected other person from that country speaks.[3]

RELEVANCE: LOGICAL AND STATISTICAL DEFINITIONS FOR THE VALUE OF INFORMATION

The concepts of determination and uniformity defined above can be used to help answer another common question in learning and problem solving. Specifically, the question is, how should an agent decide whether to pay attention to a given variable? A first answer might be that one ought to attend to variables that determine or suggest high uniformity for a given outcome of interest. The problem is that both determination and uniformity fail to tell us whether a given variable is *necessary* for determining the outcome. For instance, the color of Smirdley's shirt determines how many steps the Status of Liberty has, as determination has been defined, because the number of steps presumably does not change over time. As another example, one's zip code and how nice one's neighbors are determine what state one lives in, because zip code determines state. This property for determination and uniformity is useful because it ensures that superfluous facts will not get in the way of a sound inference. But when one's concern is what information needs to be sought or taken into account in determining an outcome, the limits of resource and time dictate that one should pay attention only to those variables that are *relevant* to determining it.

The logical relation of relevance between two functions F and G may be loosely defined as follows: F is relevant to determining G if and only if F is a necessary part of some determinant of G. In particular, let us say that

> F is *relevant to determining* G iff there is some set of functions $\mathbf{D}$ such that (1) $F \in \mathbf{D}$, (2) $\mathbf{D} > G$, and (3) $\mathbf{D} - \{F\}$ does not determine G.[4]

We can now ask, for a given determinant of a function, which part of it is truly relevant to the determination, and which part gives us no additional information. Whether or not a given function has value[5] to us in a given situation can thus be answered from information about whether it is relevant to a particular goal. Relevance as here defined is a special case of the more general notion because we have used only functional determination in defining it. Nonetheless, this restricted version captures the important properties of relevance. Devika Subramanian and Michael Genesereth (1987) have recently done work demonstrating that knowledge about the *ir*relevance of, in their examples, a particular proposition, to the solution of a logical problem, is useful in reformulating the problem to a more workable version in which only the aspects of the problem description that are necessary to solve it are represented. In a similar vein, Michael Georgeff has shown that knowledge about independence among subprocesses can eliminate the frame problem in modeling an unfolding process for planning (Georgeff, 1987). Irrelevance and determination are dual concepts, and it is interesting that knowledge in both forms is important in reasoning.

Irrelevance in the statistical case can, on reflection, be seen to be related to the concept of probabilistic independence. In probability theory, an event A is said to be independent of an event B iff the conditional probability of A given B is the same as the marginal probability of A. The relation is symmetric. The statistical concept of irrelevance is a symmetric relation as defined in this paper. The definition is the following:

> F is (statistically) *irrelevant to determining* G iff
> $U\{G(x) = G(y) \mid F(x) = F(y)\} = Pr\{G(x) = G(y)\}$.

That is, F is irrelevant to G if it provides no information about the value of G. For cases when irrelevance does not hold, one way to define the *relevance* of F to G is as follows:

$$R(F, G) = |U\{G(x) = G(y) \mid F(x) = F(y)\} - Pr\{G(x) = G(y)\}|.$$

That is, relevance is the absolute value of the change in one's information about the value of G afforded by specifying the value of F. Clearly,

if the value of G is known with probability 1 prior to inspection of F then F cannot provide any information and is irrelevant. If the prior is between 0 and 1, however, the value of F may be highly relevant to determining the value of G. It should be noted that relevance has been defined in terms of uniformity in the statistical case, just as it was defined in terms of determination in the logical case. The statistic of relevance is more similar to the predictive association measures mentioned in the last section for categorical data than is the uniformity statistic. As such it may be taken as another proposal for such a measure. Relevance in the statistical case gives us a continuous measure of the value of knowing a particular function, or set of functions, or of knowing that a property holds of an individual, for purposes of determining another variable of interest. Knowledge about the relevance of variables can be highly useful in reasoning. In particular, coming up with a set of relevant functions, variables, or values for determining an outcome with high conditional uniformity should be the goal of an agent when the value of the outcome must be assessed indirectly.

CONCLUSION

The theory presented here is intended to provide normative justifications for conclusions projected by analogy from one case to another, and for generalization from a case to a rule. The lesson is not that techniques for reasoning by analogy must involve sentential representations of these criteria in order to draw reasonable conclusions. Rather it is that the soundness of such conclusions, in either a logical or a probabilistic sense, can be identified with the extent to which the corresponding criteria (determination and uniformity) actually hold for the features being related. As such it attempts to answer what has to be true of the world in order for generalizations and analogical projections to be reliable, irrespective of the techniques used for deriving them. That the use of determination rules without substantial heuristic control knowledge may be intractable for systems with large case libraries does not therefore mean that determination or uniformity criteria are of no use in designing such systems. Rather, these criteria provide a standard against which practical techniques can be judged on normative grounds. At the same time, knowledge about what information is relevant for drawing a conclusion, either by satisfying the logical relation of relevance or by being significantly relevant in the probabilistic sense, can be used to prune the factors that are examined in attempting to generalize or reason by analogy.

As was mentioned earlier, logic does not prescribe what techniques will be most useful for building systems that reason by analogy and generalize successfully from instances, but it does tell us what problem such techniques should solve in a tractable way. As such, it gives us what David Marr (1982) called a "computational theory" of case-based reasoning, that can be applied irrespective of whether the (in Marr's terms) "algorithmic" or "implementational" theory involves theorem proving over sentences (Davies and Russell, 1987) or not. A full understanding of how analogical inference and generalization can be performed by computers as well as it is performed by human beings will surely require further investigations into how we measure similarity, how situations and rules are encoded and retrieved, and what heuristics can be used in projecting conclusions when a valid argument cannot be made. But it seems that logic can tell us quite a lot about analogy, by giving us a standard for evaluating the truth of its conclusions, a general form for its justification, and a language for distinguishing it from other forms of inference. Moreover, analysis of the logical problem makes clear that an agent can bring background knowledge to bear on the episodes of its existence, and soundly infer from them regularities that could not have been inferred before.

ACKNOWLEDGMENTS

Much of this paper is based on my senior thesis, submitted to Stanford University in 1985 and issued as (Davies, 1985). I owe a great deal to my advisor for the project, John Perry, whose work with John Barwise on a theory of situations provided exactly the right framework for analysis of these issues (Barwise and Perry, 1983). In addition, I have profited greatly from discussions with Stuart Russell, Amos Tversky, Devika Subramanian, Benjamin Grosof, David Helman, Leslie Kaelbling, Kurt Konolige, Doug Edwards, Jerry Hobbs, Russ Greiner, David Israel, Michael Georgeff, Stan Rosenschein, Paul Rosenbloom, Anne Gardner, Evan Heit, Yvan Leclerc, Aaron Bobick, and J. O. Urmson.

The research reported here was made possible in part by a grant from the System Development Foundation to the Center for the Study

of Language and Information, and in part by the Office of Naval Research under Contract Nos. N00014-85-C-0013 and N00014-85-C-0251. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research or the United States Government.

Artificial Intelligence Center,
SRI International and Department of Psychology,
Stanford University,
USA.

NOTES

[1] See the essay by Stuart Russell elsewhere in this volume.

[2] The term 'formal implication' is due to Bertrand Russell and refers to the relation between predicates P and Q in the inheritance rule $\forall x P(x) \Rightarrow Q(x)$.

[3] I am indebted to Stuart Russell for this example, and for the suggestion of the term 'uniformity'.

[4] This definition can easily be augmented to cover the relevance of *sets* of functions, and values, to others.

[5] 'Value' as used here refers only to usefulness for purposes of inference.

REFERENCES

Baker, M. and Burstein, M. H. (1987), 'Implementing a model of human plausible reasoning', in *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*, Los Altos, CA: Morgan Kaufmann, pp. 185—188.

Barwise, J. and Perry, J. (1983), *Situations and Attitudes*, Cambridge, MA: MIT Press.

Burstein, M. H. (1983), 'A model of incremental analogical reasoning and debugging', in *Proceedings of the National Conference on Artificial Intelligence (AAAI-83)*, Los Altos, CA: Morgan Kaufmann, pp. 45—48.

Carbonell, J. G. (1983), 'Derivational analogy and its role in problem solving', in *Proceedings of the National Conference on Artificial Intelligence (AAAI-83)*, Los Altos, CA: Morgan Kaufmann, pp. 64—69.

Carbonell, J. G. (1986), 'Derivational analogy: A theory of reconstructive problem solving and expertise acquisition', in Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (eds.), *Machine Learning: An Artificial Intelligence Approach, Volume II*. Los Altos, CA: Morgan Kaufmann, pp. 371—392.

Carnap, R. (1963), *Logical Foundations of Probability*, Chicago: University of Chicago press.

Copi, I. M. (1972), *Introduction to Logic*, New York: The Macmillan Company.

Davies, T. (1985), *Analogy*, Informal Note No. IN-CSLI-85-4, Center for the Study of Language and Information, Stanford, CA.

Davies, T. R. and Russell, S. J. (1987), 'A logical approach to reasoning by analogy', in *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*, Los Altos, CA: Morgan Kaufmann, pp. 264—270. Also issued as Technical Note 385, Artificial Intelligence Center, SRI International, Menlo Park, CA, July 1987.

Genesereth, M. R. and Nilsson, N. J. (1987), *Logical Foundations of Artificial Intelligence*, Los Altos, CA: Morgan Kaufmann.

Gentner, D. (1983), 'Structure mapping: A theoretical framework for analogy', *Cognitive Science* **7**: 155—170.

Georgeff, M. P. (1987), *Many Agents Are Better Than One*, Technical Note 417, Artificial Intelligence Center, SRI International, Menlo Park, CA.

Gick, M. L. and Holyoak, K. J. (1983), 'Schema induction and analogical transfer', *Cognitive Psychology* **15**: 1—38.

Goodman, L. A. and Kruskal, W. H. (1979), *Measures of Association for Cross Classifications*, New York: Springer-Verlag.

Goodman, N. (1983), *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press.

Greiner, R. (1985), *Learning by Understanding Analogies*, Technical Report STAN-CS-85-1071, Stanford University, Stanford, CA.

Haberman, S. J. (1982), 'Association, measures of', in Kotz, S. and Johnson, N. L. (eds.), *Encyclopedia of Statistical Science, Volume I*, New York: John Wiley and Sons, pp. 130—137.

Hays, W. L. and Winkler, R. L. (1970), *Statistics, Volume II: Probability, Inference, and Decision*, San Francisco: Holt, Rinehart and Winston.

Hesse, M. B. (1966), *Models and Analogies in Science*, Notre Dame: University of Notre Dame Press.

Holland, J., Holyoak, K., Nisbett, R. and Thagard, P. (1986), *Induction: Processes of Inference, Learning, and Discovery*, Cambridge, MA: MIT Press.

Johnson, R. A. and Wichern, D. A. (1982), *Applied Multivariate Statistical Analysis*, Englewood Cliffs, NJ: Prentice-Hall.

Kedar-Cabelli, S. (1985), 'Purpose-directed analogy', in *The Seventh Annual Conference of the Cognitive Science Society*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 150—159.

Leblanc, H. (1969), 'A rationale for analogical inference', *Philosophical Studies* **20**: 29—31.

Marr, D. (1982), *Vision*, New York: W. H. Freeman and Company.

Mill, J. S. (1900), *A System of Logic*, New York: Harper & Brothers Publishers.

Mitchell, T. M. (1980), *The Need for Biases in Learning Generalizations*, Technical Report CBM-TR-117, Rutgers University, New Brunswick, NJ.

Mitchell, T. M., Keller, R. M., and Kedar-Cabelli, S. T. (1986), 'Explanation-based generalization: A unifying view', *Machine Learning* **1**: 47—80.

Montgomery, D. C. and Peck, E. A. (1982), *Introduction to Linear Regression Analysis*, New York: John Wiley & Sons.

Nilsson, N. (1984), *Shakey the Robot*, Technical Note 323, Intelligence Center, SRI International, Menlo Park, CA.

Nisbett, R. E., Krantz, D. H. Jepson, D., and Kunda, Z. (1983), 'The use of statistical heuristics in everyday inductive reasoning', *Psychological Review* **90**: 339—363.

Rissland, E. L. and Ashley, K. D. (1986), 'Hypotheticals as heuristic device', in *Proceedings of the National Conference on Artificial Intelligence (AAAI-86)*, Los Altos, CA: Morgan Kaufmann, pp. 289—297.

Rosenbloom, P. S. and Newell, A. (1986), 'The chunking of goal hierarchies: A generalized model of practice', in Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (eds.), *Machine Learning: An Artificial Intelligence Approach, Volume II*. Los Altos, CA: Morgan Kaufmann, pp. 247—288.

Russell, S. J. (1986), *Analogical and Inductive Inference*, PhD Thesis, Stanford University, Stanford CA.

Russell, S. J. and Grosof, B. N. (1987), 'A declarative approach to bias in inductive concept learning', in *Proceedings of the National Conference on Artificial Intelligence (AAAI-87)*, Los Altos, CA: Morgan Kaufmann, pp. 505—510.

Shaw, W. H. and Ashley, L. R. (1983), 'Analogy and inference', *Dialogue: Canadian Journal of Philosophy* **22**: 415—432.

Subramanian, D. and Genesereth, M. R. (1987), 'The relevance of irrelevance', in *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*, Los Altos, CA: Morgan Kaufmann, pp. 416—422.

Thagard, P. and Nisbett, R. E. (1982), 'Variability and confirmation', *Philosophical Studies* **42**, 379—394.

Theil, H. (1970), 'On the estimation of relationships involving qualitative variables', *American Journal of Sociology* **76**: 103—154.

Ullman, J. D. (1983), *Principles of Database Systems*, Rockville, MD: Computer Science Press.

Vardi, M. Y. (1982), *The Implication and Finite Implication Problems for Typed Template Dependencies*, Technical Report STAN-CS-82-912, Stanford University, Stanford, CA.

Weitzenfeld, J. S. (1984), 'Valid reasoning by analogy', *Philosophy of Science* **51**: 137—149.

Wilson, P. R. (1964), 'On the argument by analogy', *Philosophy of Science* **31**: 34—39.

Winston, P. H. (1980) 'Learning and reasoning by analogy', *Communications of the Association for Computing Machinery* **23**: 689—703.