CONSISTENCY AND PLAUSIBLE INFERENCE

J. R. Quinlan

October 1982

P-6831

I. INTRODUCTION

Research in expert systems is concerned with how to represent and reproduce the problem-solving skills that experts exhibit in their respective domains. One of the most basic of these skills is the ability to put two and two together--to draw reasoned conclusions that supplement direct observations. This poses a difficulty because our models of reasoning are derived from the deduction mechanisms of logic and, almost without exception, investigators have noted that expert reasoning beyond a superficial level cannot be understood in terms of such precise schema. Logic deals with an idealized world in which facts are known with certainty and rules of inference allow other facts to be deduced with equal certainty. Experts, on the other hand, are usually required to form judgments based on evidence. Such evidence may be subject to uncertainties arising from errors of measurement or difficulty of interpretation. The argument that justifies a conclusion in terms of the evidence may also be more tenuous than a syllogism. Again, experts can operate in environments containing inconsistent or contradictory "facts," but such environments are useless in the logical sense because a set of propositions that includes implicitly both the affirmation and the denial of a proposition can be used to prove anything whatsoever.

The study of how to overcome difficulties such as inconsistency and lack of definitiveness and still reach reasonable, supportable conclusions is called <u>plausible</u> or <u>uncertain inference</u>. Systems developed for this task typically operate in a zeroth-order world of

propositions and relations among them, where the "zeroth-order" means that propositions are unquantified or atomic. Propositions may be interpreted as facts ("the car won't start"), hypotheses ("the trouble is in the ignition system"), findings ("the distributor cap is defective"), or any concept in the domain that is relevant to the expert's problem-solving behavior. Relations connecting subsets of the propositions are usually expressed as logical definitions ("A is the conjunction of B, C, and D") or inferential links ("If A is true, then so is B"). The feature that distinguishes uncertain inference from the familiar propositional calculus is the qualified nature of knowledge about both the relations and propositions. Propositions have associated with them some (usually continuous) measure of their validity instead of being either true or false. Inferential relations also have a validity measure that weakens the connection between antecedent and consequent; the relation "If A, then B" may support a less-than-categorical affirmation of B even when A is known with certainty.

A useful way of viewing this formalism is as an <u>inference net</u> [Hayes-Roth, Waterman, and Lenat, 1978; Duda, Gaschnig, and Hart, 1979; Gaschnig, 1981]. The propositions are represented as nodes and the relations among propositions become the links of the network. Whenever the validity measure of a node is changed, such as by the arrival of new evidence, this information propagates along the links to related nodes and may cause changes to their validity measures in turn. The secondary changes propagate in the same way so that, when the net stabilizes again, the evidence responsible for the initial change may be reflected in altered validity measures of many propositions or, to put it another

way, all the inferences that can be made from that evidence have been made.

The general inference net framework does not address the important questions of how validity is to be represented and how the propagation above is to be carried out. Some approaches measure the validity of a proposition as its posterior probability or likelihood given all the evidence to hand, and use Bayes' Theorem together with various assumptions to compute this likelihood. Others use probability intervals rather than values as a measure of validity, relying on more general schemes of updating such as the Dempster-Shafer theory of evidence [Barnett, 1981; Garvey, Lowrance, and Fischler, 1981]. It is not uncommon for the form of the inference net to be restricted; for example, [Pearl, 1982] requires that it be a tree. Many systems treat the links representing relations as directional, so that the relation "If A, then B" allows updating of B's validity when A is known to be true but does not allow A's validity to be altered if B is found to be false. A review and critique of the more common approaches can be found in [Quinlan, 1982]. The account of two current systems that appears later in this paper should convey some feel for the techniques used.

Quite a few expert systems embodying mechanisms for uncertain inference have achieved notable successes, as exemplified by two of the pioneering efforts. MYCIN [Shortliffe and Buchanan, 1975], an early program that diagnosed bacterial infections and prescribed appropriate antibiotic therapy, was rated highly by a panel of experts [Buchanan, 1982] and its general-purpose successor EMYCIN [van Melle, 1979, 1980] has formed the basis of many expert systems. A recent article

[Campbell, Hollister, Duda and Hart, 1982] reporting a verified strike by Prospector, SRI's geological consultant program, attracted widespread news coverage. Despite these and other achievements, there appear to be applications requiring an uncertain inference capability that are not handled well by any current system. The characteristics of these applications are discussed in later sections, but the gist of the difficulty and the proposed solution can be obtained from the following example.

Consider the task of a fictional detective investigating a case in which (as usual) there are many apparent contradictions in the evidence that he unearths. How is he to proceed? Current approaches to plausible inference would have him weigh evidence for and against each hypothesis, considering the hypothesis confirmed to the extent that the balance of evidence supports it. But any mystery buff knows that this approach differs from the one Poirot would adopt, and might even lead to the anomalous situation in which the balance of evidence individually supports propositions A and B, but where A and B cannot both have occurred. This paper suggests an alternative method of forming conclusions that our detective would find more familiar. Instead of making deductions from contradictory information, we divide the evidence into two classes, items to be believed and items to be disregarded, so that all the evidence in the former category is consistent and "makes sense." Where there are many possible divisions we use some model to weigh the validity, not of individual propositions, but of the <u>division itself</u>. For example, a division that would require our detective to disregard significantly more data than another might be judged to be

less valuable. Regardless of how the divisions are evaluated, drawing conclusions only from one or more of these consistent subsets has the effect of giving precedence to global considerations over the more local assessment of individual hypotheses.

The paper is organized as follows: Section II sets out a seemingly simple uncertain inference problem. Section III sketches Prospector as an example of a directed Bayesian architecture, shows that the problem must be redrafted to fit the Prospector formalism, and discusses the difficulties of interpreting the findings for this case. Section IV describes INFERNO, a non-directed non-Bayesian architecture sensitive to the consistency of information, and shows that it is also less than satisfactory for this task. Section V introduces a new system called Ponderosa that performs uncertain inference by evidence division rather than by propagation of validity. The final section summarizes the paper and speculates on the possibility of merging two approaches.

II. DESCRIPTION OF THE TRIAL APPLICATION

The setting for this application of uncertain inference is an attempt to model the interactions among five econometric indicators. We are given several assertions concerning both general relationships among the indicators and predictions about what will happen in the near future. The goal is to draw meaningful inferences from these assertions so as to arrive at a composite picture of what will happen to all the indicators.

Table T1 contains the ten assertions that define the model.[1] Numbers in brackets following assertions are validity measures in the range 0 to 1; where there are two such numbers following an assertion they correspond to the "if" and "only if" cases respectively. Since we have not defined what we mean by "validity", the precise interpretation of these numbers is open. It is intended that a proposition or relation with validity 1 be equivalent to a categorical assertion and that one with validity 0 be totally vacuous. We will accept any of the different meanings of a middle-ground validity that are used in current systems.

The model is typical of real-world applications in form if not in content. The validities of the assertions or beliefs range from very weak (as in A1) to near-categorical (as in A8 and A9). Assertions like A4 that relate directly to an indicator of interest are relatively uncertain, but it is often possible to make a stronger statement about a less interesting proposition as illustrated by A3.

[1] This model was derived from an exercise in a Mathematical Logic text; any resemblance to any theory of Economics, past or present, is purely coincidental.

Table T1

ASSERTIONS DEFINING THE MODEL

A1	Stocks will fall [.55]
A2	Either taxes will not be raised or both stocks will fall and interest rates will fall [.85]
A3	Either taxes will be raised or interest rates will not fall [.9]
A4	Interest rates won't fall [.75]
A5	Either taxes will be raised or there will be a high deficit [.85]
A6	Bonds will rise or interest rates will fall if, and only if, stocks fall or taxes are not raised [.6,.85]
A7	Stocks will fall if, and only if, bonds rise and taxes are raised [.7,.8]
A8	If interest rates fall, either stocks will not fall or bonds won't rise [.95]
A9	Interest rates will not fall if there is a high deficit [.95]
A10	If there is a high deficit, stocks will fall [.8]

The application maps directly into the zeroth-order formalism described in Sec. I. There are five basic propositions corresponding to the indicators of primary concern,

- stocks will fall
- interest rates will fall
- taxes will be raised
- bonds will rise
- there will be a high deficit

We have also a small number of derived propositions stated as logical combinations of these basic propositions, such as "bonds will rise or interest rates will fall." Each such derived proposition is defined by

one or more logical relations, e.g., that the above is the disjunction of "bonds will rise" and "interest rates will fall." Assertions A1 through A5 each provides evidence in the form of a validity for one of the basic or derived propositions. Each of the last five assertions becomes either one or two inferential relations. All in all there are 16 propositions, 7 inferential relations and 12 logical relations.

Despite the simplicity of this model, it may not be immediately apparent that the information in the assertions is inconsistent. A1 and A7, for example, jointly support the inference that taxes will be raised, while assertions A2 and A4 together suggest that taxes will not be raised. In the logical sense, therefore, this collection of assertions is of no value because anything at all can be inferred from it via the tautology A -> (~A -> B). However, it seems that most plausible reasoning tasks involve inconsistent information so that the example is not an unfair one.

III. PROSPECTOR

Prospector [Duda, Hart, and Nilsson, 1976; Duda, Hart, Nilsson, Reboh, Slocum, and Sutherland, 1977; Gaschnig, 1980, 1981] is a general-purpose architecture for uncertain inference that has been used with several geological models and whose basic approach has been taken up by other systems such as AL/X [Reiter, 1980, 1981; Paterson, 1981]. It is therefore representative of a well-developed school of thought about uncertain inference.

OVERVIEW OF PROSPECTOR

Prospector and other Bayesian systems model the validity of a proposition by its posterior probability given the evidence at hand. Let H be some proposition about which inferences are to be drawn and E another proposition. Bayes' theorem gives the posterior probability (or likelihood) of H given E as

$$P(H|E) = P(E|H) \times P(H) \;/\; P(E)$$

where P(E) and P(H) are prior probabilities, and correspondingly

$$P(\sim H|E) = P(E|\sim H) \times P(\sim H) \;/\; P(E)$$

Assuming that the latter is non-zero, we can divide the first equation by the second to obtain

$$O(H|E) = O(H) \times [P(E|H) \;/\; P(E|\sim H)]$$

which may be stated as, the posterior odds of H is its prior odds

multiplied by a factor (called λ) that characterizes the sufficiency of E as a predictor of H. A similar analysis can be performed replacing E by ~E in the above, and the corresponding factor λ' characterizes the necessity of E if H is to hold.

Unfortunately, this formalism is insufficient by itself to determine what should happen to the odds of H when several propositions E_1, E_2, ... are relevant to it, or when the E's are known with less than certainty. The approach taken in Prospector is to make two additional assumptions:[2]

(Conditional independence) The probability $P(E_i|H,E_j)$ of E_i given H and E_j is equal to $P(E_i|H)$, and similarly for ~H.

(Interpolation) The effective multiplying factor to use when E_i is known with less than certainty is obtained from a piecewise linear interpolation:

-- If the observed probability of E_i is greater than its prior probability, interpolate the posterior probability of H between $P(H)$ and $P(H|E_i)$

-- Otherwise, interpolate between $P(H)$ and $P(H|\sim E_i)$

In either case the effective multiplying factor is the interpolated posterior odds of H divided by its prior odds.

---

[2] Pednault, Zucker, and Muresan (1981) and Szolovits and Pauker (1980) have commented on the inappropriateness of these assumptions, particularly the first. Konolige [Konolige 1982; Duda, Hart, Konolige, and Reboh, 1979] has developed an appealing scheme in which these assumptions are replaced by a single unifying assumption: that the posterior distribution chosen should contain minimal information while still conforming to user-specified constraints on marginal and conditional probabilities.

Under these assumptions, the posterior odds of proposition H is simply the product of its prior odds and the effective multiplying factors obtained as above for each E_i.

Inferential links from one proposition to another can thus be implemented by choosing appropriate values for the factors λ and λ'. Prospector allows logical relations among propositions as follows:

- If A is the negation of B, the odds of A is the reciprocal of the odds of B.
- If A is the disjunction of B_1, B_2, ... , the odds of A is the maximum of the odds of any B_i.
- If A is the conjunction of B_1, B_2, ... , the odds of A is the minimum of the odds of any B_i.

Prospector's control structure comes from its intended application as a consultation-style system. Each relation can cause the odds of only one proposition to be altered directly; inferential relations "If E, then H" as before affect only H, and logical relations as above affect only A. Accordingly, the links representing relations are thought of as directed into the affected proposition. Prospector requires that there be no cycles in the corresponding inference net and allows observed probabilities to be given only for "evidence" propositions that have no links directed into them.

APPLYING PROSPECTOR TO THE MODEL

Several difficulties arise when we attempt to use the Prospector architecture for the task described in Sec. II. The more serious of these are consequences of Prospector's tacit assumption that

propositions can be arranged in a hierarchy with inference chains flowing smoothly from raw evidence through to conclusions.

Consider, for example, the proposition "stocks will fall." Assertion A1 establishes this proposition as being true with some validity, implying that this proposition is evidence. Several other assertions, however, establish conditions under which the proposition can be inferred to be true, thus establishing inferential links to the proposition and so preventing it from being evidence. Again, assertions A2, A3, and A5 establish that certain logical combinations of propositions are valid, and Prospector contains no mechanism that would allow evidence to bear directly on such propositions. Similar problems arise from A6, A7, and A8, where logical combinations are on the receiving end of the inferential links.

The steps taken to reformulate the example are as follows: (1) The two propositions "stocks will fall" and "interest rates will fall" that appear both as evidence and as potential conclusions are represented each by two nodes in the net. The first node is a conventional evidence node with a very strong inferential link to the second copy that is also the recipient of other inferential links. (2) Assertions such as A2 of the form "A or B" are represented functionally as the pair of inference relations "If A is false, then B" and "If B is false, then A." (3) Complex assertions are broken down into more primitive relations that have a single proposition as the inference. For example, A6, of the form "A or B if, and only if, C or D," becomes the set of relations:

If (A or B) and ~C, then D

If (A or B) and ~D, then C

If ~A and ~B, then ~C

If ~A and ~B, then ~D

(4) Finally, all prior probabilities are taken by default as 0.5 since the example does not specify other values, and the strengths of the multipliers λ and λ' are determined so that, if the relation "If A, then B" has validity V, the posterior probability of proposition B given A is also V.

Table T2 lists the Prospector-style inferential relations that this reformulation produces, together with their multiplying factors λ and

Table T2

INFERENTIAL LINKS IN PROSPECTOR FORM

To	From	λ	λ'	Source
stocks-	same (evidence)	10,000	0.0001	
	taxes+	5.67	1	A2
	(bonds+ v interest-) & taxes+	5.67	1	A6
	~bonds+ & ~interest-	0.67	1	A6
	interest- & bonds+	0.05	1	A8
	high deficit	4.00	1	A10
interest-	same (evidence)	10,000	0.0001	
	high deficit	0.05	1	A9
	taxes+	5.67	0.11	A2,A3
taxes+	stocks-	4.0	0.18	A2,A7
	interest-	9.0	0.18	A2,A3
	high deficit	1	5.67	A5
	(bonds+ v interest-) & ~stocks-	0.18	1	A6
	~bonds+ & ~interest-	1.5	1	A6
	~stocks- & bonds+	0.43	1	A7
bonds+	stocks-	4.0	1	A7
	interest- & stocks-	0.05	1	A8
	~stocks- & taxes+	0.43	1	A7
high deficit	taxes+	1	5.67	A5

λ'. The final column of the Table shows the assertion(s) responsible for each link. At this point a serious flaw becomes evident--the network of propositions and relations contains cycles and so violates another Prospector requirement. The cycles are the result of strong interconnections among the five indicators, however, and there seems to be no way of eliminating them. Rather than abandon the enterprise forthwith, we will generalize the Prospector algorithm to allow computation of posterior probabilities by relaxation, terminating when changes are small so that the cycles will not cause infinite loops.

A Prospector-like system embodying this change was used to obtain the results shown in Table T3. (Since a relaxation algorithm was used, there is no guarantee that this is the only set of posterior probabilities that is stable; reordering of the computational steps could give rise to a different solution.) Although they appear to be the kind of results that are commonly obtained from plausible inference systems, they are deficient in at least two respects. First, they do not highlight the fact, noted in Sec. II, that the set of assertions from which the model was derived is internally inconsistent. The

Table T3

RESULTS FROM A PROSPECTOR-STYLE SYSTEM

Proposition	Posterior Probability	Categorical Interpretation
stocks will fall	.64	T
interest rates will fall	.08	F
taxes will be raised	.27	F
bonds will rise	.59	T
there will be a high deficit	.66	T

assumptions that Prospector makes will never produce an overconstrained system, so any collection of evidence and relations will lead to a solution. But Michie (1980, 1982) argues convincingly that practical expert systems must be user-friendly, and therefore, out of concern for the validation and intelligibility of results, it would seem that consistency checking ought to be one of their more important functions. Second, the statement of a result as a probability is fine when there is only one result of interest, but can lead to problems in cases such as this when we need a simultaneous reading of several variables. Suppose that the model builder wished to predict the most likely future state from the 2^5 possible in terms of the five indicators. Converting the probabilities to categorical form by threshholding as in Table T3 would lead to the conclusion that

> stocks will fall;
> interest rates will not fall;
> taxes will not be raised;
> bonds will rise; and
> there will be a high deficit.

These conclusions may be individually unsurprising but in combination they violate the "only if" part of assertion A7. This relation has a relatively high validity of 0.8 and so any conclusion that disregards it is suspect. Thus mapping from probabilistic to categorical results for several variables (when called for by the application) may produce conclusions that do not fit with the evidence.

In summary, in order to run our example on Prospector we had to make significant alterations to the formulation of the model and to modify Prospector as well; even so, the results we obtained were

deficient. For all these reasons it would seem that Prospector is not well-suited to this application.

IV. INFERNO

INFERNO [Quinlan, 1982] is another inference network system that was designed around four ideas:

1. General systems for uncertain inference are better off without assumptions such as conditional independence whose universal validity is suspect.
2. On the other hand, it should be possible to assert that particular groups of propositions exhibit relationships such as independence.
3. There should be no restrictions on the direction of information flow in the network. (This was the cause of much of Prospector's difficulty with the example of Sec. II.)
4. The consistency of the given information should be checked and the system should be able to advise on alternative methods of rectifying inconsistencies.

The effect of these requirements has been to lead away from Prospector-style formalisms.

DESCRIPTION OF INFERNO

The first difference comes in the way that the validity of a proposition is represented. Instead of a single point probability, INFERNO uses probability bounds similar to the interval approach of the Dempster-Shafer Theory of Evidence [Barnett, 1981; Garvey et al., 1981]. Every proposition A is characterized by a lower bound <u>t(A)</u> on the

probability P(A) of A and a lower bound <u>f(A)</u> on P(~A), so that

$$t(A) \leq P(A) \leq 1 - f(A).$$

This approach has two features. The uncertainty of our knowledge about A is apparent, being just the difference between t(A) and 1 - f(A). Second, the values of t(A) and f(A) are derived from evidence tending to support and to deny A respectively, and these values are retained and propagated separately.

To achieve the non-directed propagation of inferences as in point (3) above, INFERNO follows WAND [Hayes-Roth, 1981] in viewing relations as establishing constraints on the respective validities of collections of propositions. Changing a probability bound of any proposition in the collection may cause the constraint to be violated, requiring some other bound to be altered. For example, one form of inferential relation, written as

A enables B with strength X

is intended to capture the (uncertain) relation "If A, then B." This relation has two constraints associated with it:

$$t(B) \geq t(A) \times X$$

$$f(A) \geq 1 - (1 - f(B)) / X$$

and thus can cause t(B) to be increased when t(A) is increased, or f(A) to be increased when f(B) is increased. Logical connections among propositions are handled in the same manner. The relation defining A as the conjunction of B_1, B_2, ... , B_n gives four constraints: for all B_i,

$$t(A) \geq 1 - \Sigma_i(1 - t(B_i))$$

$$f(A) \geq f(B_i)$$

$$t(B_i) \geq t(A)$$

$$f(B_i) \geq f(A) - \Sigma_{j\neq i}(1 - t(B_j))$$

These and all other INFERNO constraints can be derived from simple probability identities and do not depend on other assumptions.

This representation also supports a probabilistic concept of consistency. If $t(A) + f(A) > 1$ for some proposition A, the information about A is inconsistent and one or both of the bounds must be incorrect. Since the propagation constraints are provably correct, the inconsistency can only arise from contradictions implicit in the information given to the system. INFERNO can suggest ways to alter the data so as to make it consistent. A <u>change</u> takes the form of lowering the given value of a bound or reducing the strength of one of the inferential relations. A combination of changes that is sufficient to make the bounds on all propositions consistent is called a <u>rectification</u>. INFERNO can generate the best <u>n</u> of the possible rectifications, ranking them under the assumption that those involving the least alteration of the original data are more likely to be acceptable.

APPLYING INFERNO TO THE MODEL

When we wished to apply Prospector to the task of Sec. II we first had to reformulate it to conform to Prospector's architectural restrictions. INFERNO does not impose any such restrictions and the example can be run in its original form.

INFERNO immediately finds the set of assertions to be inconsistent. The explanation of one such inconsistency serves also to illustrate the propagation of bounds in the system. It helps to remember in the following that INFERNO is again using probability as a model of validity.

- Assertion A1 of Table T1 is that stocks will fall with probability .55, and A7 states (with strength .8) that this will happen only if bonds rise and taxes are raised. The conclusion is that the probability of bonds rising and taxes being raised is at least .8 × .55, or .44.
- Clearly, the probability of one component of this conjunction, taxes being raised, must also be at least .44, so the probability that taxes will not be raised is at most .56.
- Assertion A2 gives the probability that taxes will not be raised or stocks will fall and interest rates will fall as .85; the probability of the first term of the disjunction is at most .56, so the probability that stocks will fall and interest rates will fall is at least .29.
- The probability that interest rates will fall must therefore be at least .29, but (by A4) the probability is at most .25.

INFERNO's analysis of the various interdependencies then leads it to propose four alternative rectifications, each of which will correct the above and all other inconsistencies. Each rectification consists of a

single change:

- Reduce the probability that interest rates will not fall (assertion A4) to .71.
- Reduce the probability of assertion A2 to .81.
- Reduce the probability that stocks will fall (assertion A1) to .5.
- Weaken the only-if strength of assertion A7 to .727.

This sort of consistency analysis is intended to permit the user to review selected fragments of the information that he presented to the system with an eye to making it consistent before trusting conclusions based on it. He has the option of ignoring the inconsistencies if he wishes, as he might well do in this case since the probability bounds are in only marginal conflict.

Let us suppose, though, that he elects to remove the inconsistencies by disregarding completely the assertion A1 that stocks will fall; it was after all a relatively weak belief according to its validity measure. The consistent set of probability bounds that INFERNO obtains from A2 through A10 is shown in Table T4. In general it is more difficult to place a categorical interpretation on INFERNO's ranges than

Table T4

RESULTS FROM INFERNO

Proposition	Probability Range	Categorical Interpretation
stocks will fall	.36 - .5	F
interest rates will fall	.138 - .25	F
taxes will be raised	.288 - .4	F
bonds will rise	.288 - 1	?
there will be a high deficit	.45 - .625	T

it was in the case of Prospector's single probabilities, but in this instance the mapping to {T,?,F} seems reasonable. Notice, though, that the categorical interpretation again violates a relatively strong relation (A10) predicting that stocks will fall if there is a high deficit!

To summarize: INFERNO avoids three of the four difficulties that Prospector experienced with the model. It allows assertions and inferences about logical combinations of propositions and is not put out by cycles in the net. It also makes apparent any inconsistencies in the information presented to it and provides helpful aids to reviewing the information. However, an attempt to place categorical interpretations on the results can once more lead to conclusions that are not consistent with the data.

V. PONDEROSA

Ponderosa represents a departure from current plausible inference systems because, although it still deals with uncertain assertions and relations, it does not attempt to propagate validity measures of any kind. Instead, it follows the approach outlined in Sec. I of trying to separate out from the information given to it one or more internally consistent subsets. The merit of any such division is then established as a function of the validities of assertions that were not included.

DESCRIPTION OF THE APPROACH

Consider a set of assertions such as those in Sec. II. Each assertion can be viewed as a well-formed formula (wff) of the propositional calculus with a validity measure attached, or, in the case of the "if and only if" assertions, a pair of such formulas. Let C be a subset of the wffs, where we disregard for the moment each wff's validity measure. C is consistent in the logical sense if there is no wff that can be both proved and disproved from C.[3] A subset is maximally consistent if it is consistent but the addition of any other wff from the original set will make it inconsistent.

Suppose now that the original set of wffs has been divided into a maximally consistent subset C and the remainder $R = \{R_1, R_2, \ldots, R_n\}$, and let $V(R_i)$ be the validity measure of R_i. One way of assessing the situation (which will be called an interpretation) would be to accept

[3] This notion of consistency is stronger than the one used for INFERNO in which it is permissible to infer both A and ~A so long as the sum of the upper bounds of P(A) and P(~A) does not exceed 1.

the wffs in C together with all their (consistent) inferences and to ignore the wffs in R as being either erroneous (e.g., resulting from faulty observation) or general default assertions that do not apply in this case. How plausible is this interpretation? If it is to be correct, each individual R_i must be incorrect or inapplicable. The probability that this interpretation is incorrect is then the probability of the disjunction of the elements of R, i.e.,

$$P(R_1 \vee R_2 \vee \ldots \vee R_n)$$

If we again treat validity measures as probabilities and use the identity

$$P(A),\ P(B) \leq P(A \vee B) \leq P(A) + P(B)$$

we obtain the probability P(C,R) that the interpretation is correct as

$$1 - \Sigma_i V(R_i) \leq P(C,R) \leq \min_i (1 - V(R_i))$$

Since we are identifying validity measures with probabilities, P(C,R) represents the validity of the interpretation dividing the original set of wffs into C and R.

Of course, the number of potential splits of a set of wffs into a maximally consistent subset and a remainder grows exponentially with the size of the set. The validity measures attached to propositions, however, provide methods of reducing the computational load. First, we are clearly uninterested in any interpretation whose validity is zero. If any wff in R has a validity of 1 and thus is categorically correct, the inequality above gives a zero upper bound on the validity of that

interpretation. Consequently, we need consider only interpretations in which all categorical assertions are included in the consistent subset C. Second, we do not wish to swamp the user with all possible divisions, but rather to display only the best n of them for some small, fixed n (currently 10). So we do not need to generate all possible interpretations provided that the ones omitted are inferior to the ones displayed. Since the validity of an interpretation is known only as a range and thus two interpretations cannot be compared directly, the midpoint of the range is used for ranking them.

The innards of Ponderosa can be sketched as follows.[4] Each proposition A is broken into two findings "A is true" and "A is false." Associated with each finding is a collection of justifications for the finding, where a null justification indicates that there is no reason to believe the finding. Each justification for the finding is either that the finding is an explicit assertion given to the system, or that the finding is an inference from a relation and one or more other findings with non-null justifications. For instance, the finding "B is false" and the relation "A implies B" together justify the finding "A is false," and the logical relation "X is the disjunction of A and B" together with both these findings justifies "X is false."

Every datum is either a relation or an assumed finding and all findings depend ultimately on the data. Ponderosa keeps with each finding a removal plan in the form of a collection of sets of data, the idea being that all justifications for this finding would evaporate if,

[4] The algorithms used in the current Ponderosa have been heavily influenced by the fact that it was implemented as a rewrite of INFERNO, and are almost certainly not the best that could be developed.

and only if, any one of these sets of data were removed. The algorithms for removal sets below depend on the observation that a removal plan is isomorphic to a logical expression in disjunctive normal form. Let us map each datum D to the predicate "D is excluded" and the removal plan

$$\{ \{D_{11}, D_{12}, \ldots\}, \{D_{21}, D_{22}, \ldots\}, \ldots \}$$

to the logical expression

$$\begin{aligned} &(D_{11} \text{ is excluded and } D_{12} \text{ is excluded and } \ldots) \\ \text{or } &(D_{21} \text{ is excluded and } D_{22} \text{ is excluded and } \ldots) \\ \text{or } &\ldots \end{aligned}$$

Then the expression is true if, and only if, one of the sets of data making up the removal plan has been discarded, in which case the plan is satisfied.

The computation of removal plans keeps pace with the propagation of inferences. Initially the only findings with justifications are those that appear in the data, and the removal plan for such a finding is {{itself}}. Suppose now that a new justification for finding F has been inferred from a relation R and findings $\{S_i\}$. This justification could be removed if either R or any of the S's could be removed, as given by the plan (in disjunctive form)

$$X = R \lor \text{removal plan}(S_1) \lor \text{removal plan}(S_2) \lor \ldots$$

But previous justifications may have been found for F and removal of F would require removal of them as well. In this case, the new removal plan for F becomes the conjunction of the old removal plan and X.

When the data are inconsistent there will be one or more contradictory propositions $\{A_i\}$ that can be both proved and disproved, i.e., one or more pairs of findings "A_i is true" and "A_i is false," both with non-null justifications and removal plans. Clearly, the data would become consistent if, and only if, one of each such pair of findings could be removed. When put into disjunctive normal form, the removal plan obtained as the conjunction over i of

removal plan ("A_i is true") v removal plan ("A_i is false")

is then just the set of remainders corresponding to all possible maximally consistent sets. Ponderosa computes this overall plan in a depth-first way so that, if a partial remainder is generated that is already more implausible than the best <u>n</u> complete remainders found so far, all possible remainders containing the partial one are omitted.

As a small but pathological illustration of all this, consider the inconsistent assertions

1. A implies B
2. B implies C
3. C implies A
4. A is true
5. B is true
6. C is false

Ponderosa notes the various findings and their associated removal plans as follows, with data referenced by the above numbers:

A is true:	{ {2,4}, {3,4}, {4,5} }
A is false:	{ {1}, {2}, {6} }
B is true:	{ {1,5}, {4,5} }
B is false:	{ {2}, {6} }
C is true:	{ {1,5}, {2}, {4,5} }
C is false:	{ {6} }

The disjunctions for each contradictory proposition are then

A: { {1}, {2}, {3,4}, {4,5}, {6} }
B: { {1,5}, {2}, {4,5}, {6} }
C: { {1,5}, {2}, {4,5}, {6} }

and the conjunction of these representing all possible remainders is

{ {1,5}, {2}, {4,5}, {6} }

APPLYING PONDEROSA TO THE MODEL

As was the case with INFERNO, Ponderosa contains no restrictions that would require the model of Sec. II to be reformulated. Once again the information in assertions A1 through A10 is found to be inconsistent and Ponderosa generates the six possible divisions of the corresponding wffs into a maximally consistent subset and a remainder. The six remainders are displayed in Table T5 together with the bounds on the validity of the divisions and the midpoints of these ranges. Notice that, whereas INFERNO would accept the weakening of just "stocks will fall" as sufficient to remedy the inconsistencies, Ponderosa uses a stronger, categorical definition of consistency and finds that removal of assertion A1 alone is not enough.

Ponderosa does not automatically select the "best" or any other maximally consistent subset as being correct. Its function stops with pointing out to the user the possibilities that exist for making his information consistent, using the validity ranking only as a filter and heuristic guide. The user's specialist knowledge may place a value on various subsets of the information that differs from this simple plausibility model. In this instance, let us suppose that the fourth

Table T5

PONDEROSA REMAINDERS

	Remainder	Validity Low	Mid	High
A7b	stocks- only if bonds+ & taxes+	.2	.2	.2
A2	~taxes+ v stocks- & interest-	.15	.15	.15
A1 A4	stocks- ~interest-	0	.125	.25
A1 A10	stocks- if high deficit then stocks-	0	.1	.2
A1 A5	stocks- taxes+ v high deficit	0	.075	.15
A4 A8	~interest- if interest- then ~stocks- v ~bonds+	0	.025	.05

remainder (assertions A1 and A10) is selected as the least valuable of those possible. When these assertions are deleted, the remainder form a consistent subset whose implications for the five indicators appear in Table T6. Ponderosa ignores the validity of wffs when it generates inferences, so all inferences are categorical and cannot individually or collectively fail to mesh with the evidence in the consistent subset.

Table T6

CONSISTENT INFERENCES

Proposition	Validity
stocks will fall	F
interest rates will fall	F
taxes will be raised	F
bonds will rise	T
there will be a high deficit	T

VI. CONCLUSION

This paper has focused on a class of plausible reasoning tasks with three characteristics: inconsistent data, non-hierarchical interaction of concepts, and the need to obtain simultaneous readings on several hypotheses. A simple model with these attributes was used to demonstrate that existing systems for inexact inference are not suited to this kind of task. We first examined Prospector as the quintessential example of a Bayesian system and showed that both the model and Prospector itself would have to be altered to get any results at all. Even then, the inconsistency inherent in the given model was not made evident and a straightforward interpretation of the results turned out to be at variance with the model. INFERNO, a more tolerant non-Bayesian system, fared better in that the model did not have to be changed and its inconsistencies were discovered, but once more the attempt to wring a categorical interpretation from the results produced an anomaly. Ponderosa was introduced as a system to perform uncertain inference by finding consistent subsets of the model, leading to results that are always categorical and that agree with whatever reduced model is used.

There are clearly other classes of plausible reasoning tasks to which Ponderosa is unsuited. If all the data is consistent or if there is a single proposition about which information is sought, the probability-bounding approach of INFERNO gives a better appraisal of the confidence with which the results can be accepted. This suggests an interesting possibility for combining the talents of Ponderosa and

INFERNO. First, Ponderosa would be used to find whether the data is categorically consistent and, if not, to help the user choose a maximally consistent subset of it. INFERNO could then be run with this subset to supplement Ponderosa's categorical inferences with probability bounds. For instance, in the previous section we selected a maximally consistent subset A2 through A9 of the assertions in Table T1. The analysis of this subset with INFERNO is shown in Table T7. It now becomes apparent that, while categorical inferences from the subset justify both the predictions that bonds will rise and that there will be a high deficit, the former conclusion has weaker probability bounds as a consequence of its derivation from less valid assertions.

In the abstract of their 1978 paper, Szolovits and Pauker state that

> "... a program which can demonstrate expertise in the area of medical consultation will have to use a judicious combination of categorical and probabilistic reasoning--the former to establish a sufficiently narrow context and the latter to make comparisons among hypotheses and eventually to recommend therapy."

Table T7

COMBINING INFERNO AND PONDEROSA

Proposition	Categorical Validity	Probability Bounds
stocks will fall	F	0 - .5
interest rates will fall	F	.25 - .25
taxes will be raised	F	.15 - .4
bonds will rise	T	.11 - 1
there will be a high deficit	T	.45 - .79

Their PIP system used categorical reasoning to generate hypotheses whose validity was then investigated probabilistically. In some ways, the proposed partnership of INFERNO and Ponderosa suggests another way of arriving at the combined approach advocated above. Ponderosa would establish a context, in the form of a subcollection of the evidence that hangs together, within which INFERNO would be used to carry out probabilistic reasoning.

Ponderosa has been implemented in Pascal and C for a VAX 11/780 minicomputer, based on a similar implementation of INFERNO. The prototype has been applied only to small tasks with less than 100 relations and propositions, and on these it is fast enough to be useful but considerably slower than INFERNO. For comparison, the CPU times consumed by the runs of Sec. IV and V were just over one second for INFERNO versus about 6 seconds for Ponderosa.

ACKNOWLEDGMENTS

I am indebted to Donald Michie of the University of Edinburgh and to Norman Shapiro of Rand, both for their comments on this paper and their suggestions for future improvements to Ponderosa.

REFERENCES

1. Barnett, J.A., "Computational Methods for a Mathematical Theory of Evidence," Proc. 7th International Joint Conf. Artificial Intelligence, Vancouver, 1981, pp. 868-875.

2. Buchanan, B.G., "New Research on Expert Systems," Machine Intelligence 10 (J.E. Hayes, D. Michie, and Y-H. Pao, eds.), London: Ellis Horwood, 1982.

3. Campbell, A.N., V. Hollister, R.O. Duda, and P.E. Hart, "Recognition of a Hidden Mineral Deposit by an Artificial Intelligence Program," Science, Vol. 217, September 1982, pp. 927-929.

4. Duda, R.O., P.E. Hart, and Nils Nilsson, Subjective Bayesian Methods for Rule-Based Inference Systems, Technical Note 124, Artificial Intelligence Center, SRI International, 1976.

5. Duda, R.O., P.E. Hart, N.J. Nilsson, R. Reboh, J. Slocum, and G.L. Sutherland, Development of a Computer-Based Consultant for Mineral Exploration, SRI International, 1977.

6. Duda, R.O., J. Gaschnig, and P.E. Hart, "Model Design in the Prospector Consultant System for Mineral Exploration," Expert Systems in the Micro Electronic Age (D. Michie, ed.), Edinburgh University Press, 1979.

7. Duda, R.O., P.E. Hart, K. Konolige, and R. Reboh, A Computer-Based Consultant for Mineral Exploration, SRI International, 1979.

8. Garvey, T.D., J.D. Lowrance, and M.A. Fischler, "An Inference Technique for Integrating Knowledge from Disparate Sources," Proc. 7th International Joint Conf. Artificial Intelligence, Vancouver, 1981, pp. 319-325.

9. Gaschnig, J., Development of Uranium Exploration Models for the Prospector Consultant System, SRI International, 1980.

10. Gaschnig, J., "Prospector: An Expert System for Mineral Exploration," State of the Art Report on Machine Intelligence (A. Bond, ed.), London: Pergamon Infotech, 1981.

11. Hayes-Roth, F., "Probabilistic Dependencies in a System for Truth Maintenance and Belief Revision," unpublished working paper, The Rand Corporation.

12. Hayes-Roth, F., D.A. Waterman, and D.B. Lenat, "Principles of Pattern-Directed Inference Systems," Pattern-Directed Inference Systems (D.A. Waterman and F. Hayes-Roth, eds.), Academic Press, 1978.

13. Konolige, K., "An Information-Theoretic Approach to Subjective Bayesian Inference in Rule-Based Systems," unpublished draft, SRI International.

14. Michie, D., "Expert Systems," Computer Journal, Vol. 23, No. 4, 1980.

15. Michie, D., "The State of the Art in Machine Learning," Introductory Readings in Expert Systems (D. Michie, ed.), London: Gordon & Breach, 1982.

16. Paterson, A., AL/X User Manual, Intelligent Terminals Ltd., Oxford, 1981.

17. Pearl, J., Distributed Bayesian Processing for Belief Maintenance in Hierarchical Inference Systems, Cognitive Systems Laboratory, UCLA, Report UCLA-ENG-CSL-82-11, 1982.

18. Pednault, E.D.P., S.W. Zucker, and L.V.Muresan, "On the Independence Assumption Underlying Subjective Bayesian Inference," Artificial Intelligence, Vol. 16, 1981, pp. 213-222.

19. Quinlan, J.R., INFERNO: A Cautious Approach to Uncertain Inference, N-1898-RC, The Rand Corporation, 1982.

20. Reiter, J., AL/X: An Expert System using Plausible Inference, Intelligent Terminals Ltd., Oxford, 1980.

21. Reiter, J., "AL/X: An Inference System for Probabilistic Reasoning," M.Sc. Thesis, Department of Computer Science, University of Illinois at Urbana-Champaign, 1981.

22. Shortliffe, E.H., and B.G. Buchanan, "A Model of Inexact Reasoning in Medicine," Mathematical Biosciences, Vol. 23, 1975, pp. 351-379.

23. Szolovits, P.S., and S.G. Pauker, "Categorical and Probabilistic Reasoning in Medical Diagnosis," Artificial Intelligence, Vol. 11, 1978, pp. 115-144.

24. van Melle, W., "A Domain-Independent Production Rule System for Consultation Programs," Proc. 6th International Joint Conf. Artificial Intelligence, Tokyo, 1979.

25. van Melle, W., A Domain-Independent System that Aids in Constructing Knowledge-Based Consultation Programs, Department of Computer Science, Stanford University, Report STAN-CS-80-820, 1980.