SRI International

EXPLAINING EVIDENTIAL ANALYSES

Technical Note No. 430

January 12, 1988

By: Thomas M. Strat and John D. Lowrance

Artificial Intelligence Center
Computer and Information Sciences Division

Approved for public release; distribution unlimited

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Abstract

One of the most highly touted virtues of knowledge-based expert systems is their ability to construct explanations for their lines of reasoning. However, there is a basic difficulty in generating explanations in expert systems that reason under uncertainty using numeric measures. In particular, systems based upon evidential reasoning using the theory of belief functions have lacked all but the most rudimentary facilities for explaining their conclusions. In this paper we review the process whereby other expert system technologies produce explanations, and present a methodology for augmenting an evidential-reasoning system with a versatile explanation facility. The method, which is based on sensitivity analysis, has been implemented, and several examples of its use are described.

Contents

1: Introduction

One of the most highly touted virtues of knowledge-based expert systems is their ability to construct explanations of deduced lines of reasoning. Endowing such systems with an explanation facility has two major advantages [1]. First, an explanation facility contributes to the *transparency* of the program. That is, it allows the user to observe, and perhaps question, the individual inferences that contribute to the conclusions that are reached. This ability to examine a system's inner workings fosters a sense of confidence in the mind of the user; he can become satisfied that the system really "knows" what it is doing and has not just happened upon a plausible conclusion. An explanation capability is thus an important component of user acceptance of a knowledge-based system. Secondly, explanations can be a useful tool for the knowledge engineer. Information gained by questioning the system about its own knowledge base can be valuable for debugging and refining the stored knowledge. Randall Davis' TEIRESIAS is a good example of a system that exploits explanations for the purpose of knowledge engineering [2].

The goal of developing knowledge-based systems that can reason with information that is uncertain or inexact in one way or another has long been a part of artificial intelligence research. Several technologies have been proposed for representing knowledge and deriving consequences from imperfect data: MYCIN's certainty factors [14], Prospector's inference nets [10], fuzzy sets [18], Bayesian nets [8], and Dempster-Shafer belief functions [6] are prominent examples. Individual differences aside, all of these technologies have one thing in common: a basic difficulty in constructing explanations of lines of reasoning.

In this paper we review the process whereby expert systems currently generate explanations, and identify the reasons why explanation generation is difficult in uncertain-reasoning systems. We then propose an explanation facility for one class of automated reasoning systems that does incorporate uncertainty: evidential reasoning. Implementation of this facility results in a knowledge-based system that has both a well-founded representation of uncertainty and a nontrivial ability to explain its inference paths.

In Section 2 we review the state of the art of explanation generation for both Boolean-valued and uncertainty-based expert systems. Section 3 contains an overview of evidential reasoning, developed at SRI International based on the Dempster-Shafer theory of belief functions. Section 4 contains our design for endowing evidential reasoning with an explanation facility based on sensitivity analysis techniques. We conclude with a discussion of the utility of the approach and the feasibility of providing such a facility in uncertain-reasoning systems

based on other technologies. The appendix contains a comprehensive example of the techniques described herein, as implemented in the evidential reasoning system known as Gister[1] [7].

[1]Gister is a trademark of SRI International

2: Explanation Generation

The generation of useful explanations in knowledge-based systems has three main requirements:

1. An effective explanation must be based upon a *recapitulation of actions* taken by a program.
2. The correct *level of detail* of those actions must be chosen.
3. There must be a *shared vocabulary* that makes the program's actions comprehensible to the user.

In simple production systems, these requirements are commonly found to be satisfied without much difficulty. But consider a program that performs inference using a numerical measure of belief. It is difficult to imagine what explanation the system could give if it were queried about a computed probability. A simple recapitulation of all invocations of a combination rule is unlikely to yield an explanation that resembles a user's conscious thought. Although such a sequence may be an effective computational model, there is no easy way to interpret it in a form that can be intuitively understood. This is not a criticism of using belief measures within expert systems; rather it is indicative of the difficulty of generating explanations within any system that employs a numeric measure of belief. A better appreciation of this difficulty can be gained by studying the explanation facilities of several systems based upon various technologies.

2.1 Logic programming

In a logic program, a collection of facts represents known truths about objects, and rules define relationships among objects. A computation is a deduction of logical consequences from a logic program. Rules are statements of the form

$$\text{IF } A_1 \text{ and } A_2 \text{ and } \cdots \text{ and } A_n \text{ THEN } B. \tag{2.1}$$

The law of *modus ponens* says that from (2.1) and the facts $A_1, A_2, \ldots, A_n$ one can deduce B. An existentially qualified goal G is deducible from a program if there is a rule with an instance of the form

$$\text{IF } x_1 \text{ and } x_2 \text{ and } \cdots \text{ and } x_n \text{ THEN } y,$$

such that y is an instance of G, and the x_is are deducible recursively from the program. Unification is used to find a rule whose consequent is an instance of the goal.

For any given goal G that is deducible from the program, one can construct a *proof tree* whose root is G, whose leaves are all instances of facts, and whose structure represents the invocations of rules in a given deduction of G. A proof tree is thus a data structure that can be used to answer queries about a computation. While the proof tree as described is only a conceptual notion, one can construct a proof tree automatically by making use of a meta-interpreter. For example, a partial meta-interpreter for constructing proof trees in Prolog [15] is given in Figure 2.1(a).

Once the proof tree has been constructed, an explanation of a given computation can be generated in a straightforward fashion. Suitable justifications for conclusions derived by modus ponens can be produced by reciting the fact (or collection of facts) that triggered the rule. When additional detail is required, reiterating the rule may also be of use. Figure 2.1(b) provides a portion of a Prolog meta-interpreter that generates a complete explanation by traversing the proof tree. An example of what such an explanation might look like is illustrated in Figure 2.2.

This is the basic mechanism whereby explanations are produced in systems based on logic programming, although its implementation may vary greatly from one system to another.

Mechanisms to control the depth to which the proof tree is explored can be used to better satisfy the second requirement for useful explanations—choosing the correct level of detail. Additionally, a more appropriate vocabulary can be used by augmenting each rule with a descriptive natural language phrase that is displayed in place of the rule itself—thus addressing the third requirement.

2.2 Certainty factors

The need to represent uncertain or inexact information in some applications has forced system developers to implement new formalisms. For example, in the MYCIN system for diagnosis of infectious diseases [14], the standard production system representation was augmented with *certainty factors* to account for the judgmental quality of some rules. On a scale of 1.0 to -1.0, a certainty factor (CF) measures the degree to which a rule's consequent does or does not follow from its premise.

Introducing CFs into a rule-based system can greatly expand the search required to reach a conclusion. In a Boolean-valued logic, any path from the goal to known facts is adequate to assert the truth of the goal, but a rule-based system incorporating uncertainty must invoke all rules that unify with every subgoal in the search tree. While many systems have been written that successfully cope with the additional computation this paradigm requires, it presents substantial

solve(Goal, Tree) — Tree is a proof tree for Goal
clause(A, B) — A<-B is a clause in the program
clause(A, true) — A is a fact

```
solve(true, true).
solve((A, B), (ProofA, ProofB)) <- solve(A, ProofA),
                                   solve(B, ProofB).
solve(A, (A <- Proof)) <- clause(A, B),
                          solve(B, Proof).
```

(a) Constructing a proof tree.

how(Goal) — Explains how the goal was proved

```
how(Goal) <- solve(Goal, Proof), interpret(Proof).

interpret((Proof1, Proof2)) <- interpret(Proof1),
                               interpret(Proof2).
interpret(Proof) <- fact(Proof, Fact),
                    writeln([Fact, 'is a fact in the data base.']
interpret(Proof) <- rule(Proof, Head, Body, Proof1),
                    writeln([Head, ' is proved using the rule']),
      writeln('IF'), write(Body), writeln(['THEN', Head])
                    interpret(Proof1).
fact((Fact<-true),Fact).
rule((Goal<-Proof),Goal,Body,Proof) <- not_equal(Proof, true),
                          extract_body(Proof, Body).
extract_body((Goal<-Proof), Goal).
```

(b) Generating explanations.

Figure 2.1: A partial meta-interpreter for Prolog.

place_in_oven(Dish, Rack) — Dish should be placed in the oven at level Rack for baking

```
place_in_oven(Dish,top) <- pastry(Dish), size(Dish,small).
place_in_oven(Dish,middle) <- pastry(Dish), size(Dish,big).
place_in_oven(Dish,middle) <- main_meal(Dish).
place_in_oven(Dish,low) <- slow_cooker(Dish).

pastry(Dish) <- type(Dish,cake).
pastry(Dish) <- type(Dish,bread).

main_meal(Dish) <- type(Dish,meat).
slow_cooker(Dish) <- type(Dish, milk_pudding).

type(dish1,bread).
type(dish2,meat).
size(dish1,small).
size(dish2,big).
```

(a) A Prolog program for placing dishes in an oven

```
how(place_in_oven(dish1, top))?

place_in_oven(dish1, top) is proved using the rule
IF pastry(dish1) and size(dish1, small)
         THEN place_in_oven(dish1, top)

pastry(dish1) is proved using the rule
IF type(dish1, bread)
         THEN pastry(dish1)

type(dish1, bread) is a fact in the data base.

size(dish1, small) is a fact in the data base.
```

(b) A sample explanation

Figure 2.2: Explanation generation within Prolog

obstacles to the construction of suitable explanations.

Consider what is required to generate an explanation at any level in the proof tree. In a Boolean-valued system, a single rule reduces each subgoal. In MYCIN, several rules may contribute to the CF of a subgoal, and all of those rules must be displayed to construct a complete explanation from a MYCIN inference.

Another difficulty is illustrated by the following MYCIN excerpt [2]:

```
The following rules were used in deducing that the identity of
ORGANISM-1 is pseudomonas-aeruginosa
RULE184
-------
Since [1.1] the category of ORGANISM-1 is not known
      [1.2] the gram stain of ORGANISM-1 is gramneg
      [1.3] the morphology of ORGANISM-1 is rod
      [1.4] the aerobicity of ORGANISM-1 is facultative
There is weakly suggestive evidence (0.3) that the identity
        of ORGANISM-1 is pseudomonas-aeruginosa
```

The low CF associated with the rule calls into question whether the rule is really a reasonable explanation. What if the CF were even lower? What if it were negative, implying that the premises are a counterindication of the consequent?

The conclusion is that systems that use CFs must find a way to select the most important rules used in an inference, if they are to satisfy the second requirement of explanation generation. TEIRESIAS incorporated mechanisms to control the level of detail of explanations generated for MYCIN based upon a measure of information content, but did not attempt to distinguish among the relative contributions when more than one rule was applicable to a given subgoal [2].

2.3 Inference nets

Tracing the arcs of an inference network is the analog of rule backtracing in a rule-based system. As with systems employing certainty factors, several evidence nodes may contribute to the belief in a hypothesis node, so an appropriate explanation may consist of several supporting reasons. An example of an explanation using an inference net from Prospector [10] (Figure 2.3) follows:

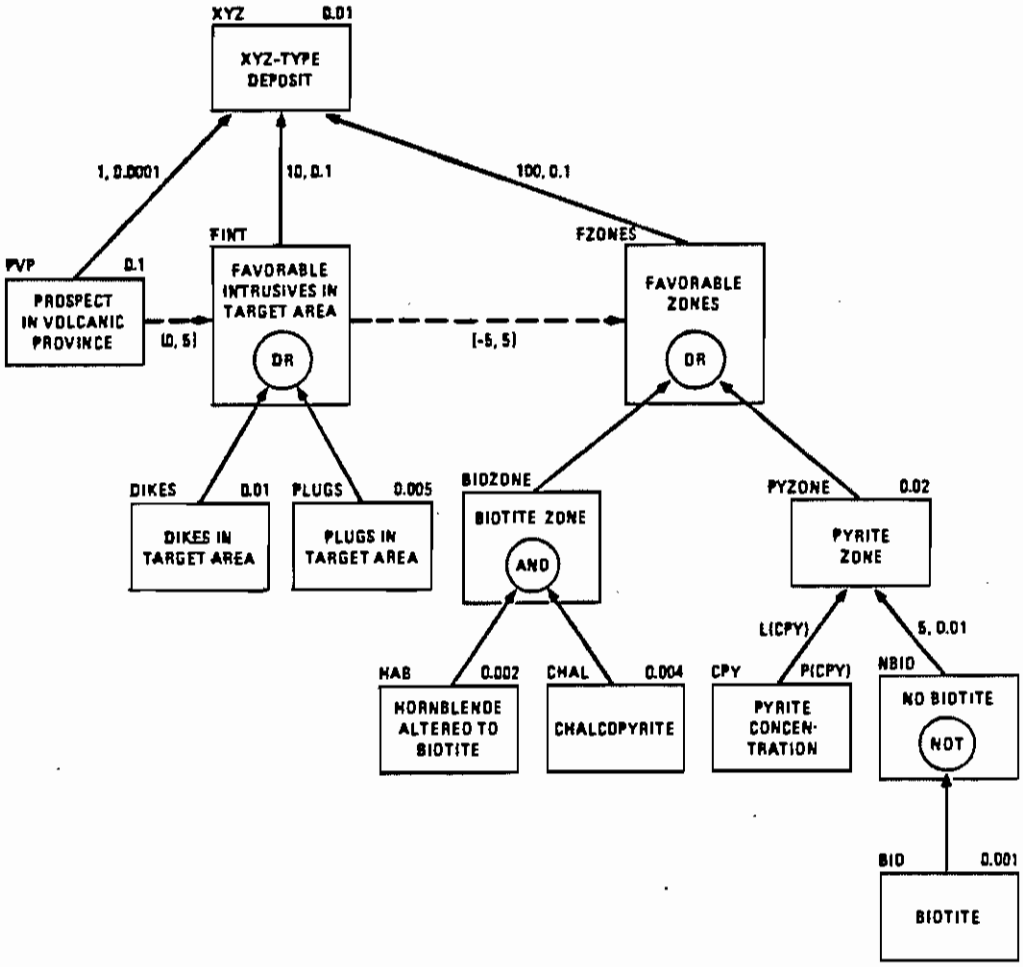


Figure 2.3: An inference network from Prospector.

```
        My certainty in
1.  XYZ-type deposit
 is now 1.21547

Do you wish to see additional information?   YES
----------------------
I suspect that
1 - (* XYZ-type deposit) (1.21547)

There is one favorable factor:
1:  1. Favorable intrusives in target area  4.99999

There is one positive factor whith neutral effect that, if negative,
could have been significant:
1:  2. You were sure that the prospect is in a volcanic province  5.0
```

```
There is one uncertain factor whose score may be subject to change:
1:  3. Favorable zones  0.227085
```

Constructing an explanation in this case is straightforward, because the nodes in Prospector inference nets represent binary predicates (e.g., whether or not there are favorable intrusives in the target area). In Hydro, a derivative system designed for water resource management problems [5], the Prospector model was extended to allow multivalued predicates, and explanation generation became more difficult:

```
On a scale from -5 to 5, my certainty that
6: 1)  INTFW based on soil type and vegetation, corrected for
slope and geology has a value between 0.72 and 1.98 (most likely 1.2825)
(computed by a formula) is now 4.0.

Do you wish to see additional information?  YES
---------------------
There are two favorable factors; in order of importance:
6.1: 1) INTFW based on soil type and vegetation, corrected for slope has
a value between 0.72 and 0.99 (most likely 0.855) (certainty 4.0)

6.1: 2) Correction factor for geology has a value between 1.0 and 2.0
(most likely 1.5) (certainty 3.0)
```

This explanation was constructed by walking the inference net and computing the range of possible values given the evidence collected to that point. While the numeric values in the Prospector explanation made interpretation a chore, in Hydro, the explanation can only be understood by someone very familiar with both the hydrological domain and Hydro's representation of uncertainty. The explanation is barely comprehensible, contradicting the third requirement.

Prospector and Hydro both possess additional features to produce a more sophisticated interpretation of the state of their knowledge bases, such as the ability to perform a best and worst-case analysis of the possible effect of a missing piece of evidence. In a later version, a sensitivity analysis was performed by applying Prospector in batch mode to a test case while systematically modifying the input data [11]. This analysis was used primarily to identify areas of disagreement between the expert and the system.

2.4 Belief functions

The theory of belief functions, as originally conceived by Dempster [3] and further developed by Shafer [12], has received considerable attention as a basis for representing uncertainty within expert systems. The theory is a generalization of classical probability theory and provides a representation of degrees of precision

as well as degrees of uncertainty. Its ability to express partial ignorance is of great value in the design of knowledge-based systems for real-world domains.

Currently, one of the most highly developed knowledge-based systems that incorporates Shafer's theory of belief functions for a wide range of application domains is Gister [7]. While Gister performs tasks similar to those of expert systems based on other technologies, like all systems based upon belief functions, it has only a rudimentary explanation capability. In the next section, we present an overview of the evidential-reasoning technology employed by Gister. The derivation of a method for generating explanations within evidential-reasoning systems follows that.

3: Overview of Evidential Reasoning

We now give a brief review of evidential reasoning. The reader is referred to Lowrance *et.al.* [7] for a fuller treatment of this technology.

3.1 Fundamentals

The goal of evidential reasoning is to assess the effect of all available pieces of evidence upon a hypothesis, by making use of domain-specific knowledge. The first step in applying evidential reasoning to a given problem is to delimit a propositional space of possible situations. Within the theory of belief functions, this propositional space is called the *frame of discernment.* A frame of discernment delimits a set of possible situations, exactly one of which is true at any one time. Once a frame of discernment has been established, propositional statements can be represented by subsets of elements from the frame corresponding to those situations for which the statements are true. Bodies of evidence are expressed as probabilistic opinions about the partial truth or falsity of propositional statements relative to a frame. Belief assigned to a nonatomic subset explicitly represents a lack of information sufficient to enable more precise distribution. This allows belief to be attributed to statements whose granularity is appropriate to the available evidence.

The distribution of a unit of belief over a frame of discernment is called a *mass distribution.* A mass distribution, m_Θ, is a mapping from subsets of a frame of discernment, Θ, into the unit interval:

$$m_\Theta : 2^\Theta \mapsto [0, 1],$$

such that

$$m_\Theta(\phi) = 0 \quad \text{and} \quad \sum_{X \subseteq \Theta} m_\Theta(X) = 1.$$

Any proposition that has been attributed nonzero mass is called a *focal element.* One of the ramifications of this representation of belief is that the probability of a hypothesis X is constrained to lie within an interval $[Spt(X),\ Pls(X)]$, where

$$Spt(X) = \sum_{Y \subseteq X} m_\Theta(Y) \quad \text{and} \quad Pls(X) = 1 - Spt(\overline{X}). \tag{3.1}$$

These bounds are commonly referred to as *support* and *plausibility.* A *body of evidence* (BOE) is represented by a mass distribution together with its frame of discernment. A BOE that directly represents one of the available pieces of

evidence is called *primitive*; all other BOEs are *conclusions* or intermediate conclusions.

In evidential reasoning, domain-specific knowledge is defined in terms of *compatibility relations* that relate one frame of discernment to another. A compatibility relation simply describes which elements from the two frames can simultaneously be true. A compatibility relation, $\Theta_{A,B}$ between two frames Θ_A and Θ_B is a set of pairs such that

$$\Theta_{A,B} \subseteq \Theta_A \times \Theta_B,$$

where every element of Θ_A and every element of Θ_B is included in at least one pair.

Evidential reasoning provides a number of formal operations for assessing evidence, including:

1. **Fusion** — to determine a consensus from several bodies of evidence obtained from independent sources. Fusion is accomplished through Dempster's rule of combination:

$$m^3_\Theta(A_h) = \frac{1}{1-k} \sum_{A_i \cap A_j = A_h} m^1_\Theta(A_i) m^2_\Theta(A_j), \tag{3.2}$$

$$k = \sum_{A_i \cap A_j = \phi} m^1_\Theta(A_i) m^2_\Theta(A_j).$$

   Dempster's Rule is both commutative and associative (meaning evidence can be fused in any order) and has the effect of focusing belief on those propositions that are held in common.

2. **Translation** — to determine the impact of a body of evidence upon elements of a related frame of discernment. The *translation* of a BOE from frame Θ_A to frame Θ_B using the compatibility relation $\Theta_{A,B}$ is defined by:

$$m_{\Theta_B}(B_j) = \sum_{\substack{C_{A \mapsto B}(A_k) = B_j \\ A_k \subseteq \Theta_A,\ B_j \subseteq \Theta_B}} m_{\Theta_A}(A_k), \tag{3.3}$$

   where $C_{A \mapsto B}(A_k) = \{b_j | (a_k, b_j) \in \Theta_{A,B}, a_i \in A_k\}$.

3. **Projection** — to determine the impact of a body of evidence at some future (or past) point in time. The *projection* operation is defined exactly as translation, where the frames are taken to be one time-unit apart.

4. Discounting — to adjust a body of evidence to account for the credibility of its source. Discounting is defined as

$$m_{\Theta}^{discounted}(A_j) = \begin{cases} \alpha \cdot m_{\Theta}(A_j), & A_j \neq \Theta \\ 1 - \alpha + \alpha \cdot m_{\Theta}(\Theta), & \text{otherwise} \end{cases} \tag{3.4}$$

where α is the assessed credibility of the original BOE ($0 \leq \alpha \leq 1$).

Several other evidential operations have been defined and are described elsewhere [7].

Independent opinions are expressed by multiple bodies of evidence. Dependent opinions can be represented either as a single body of evidence, or as a network structure that shows the interrelationships of several BOEs. The evidential reasoning approach focuses on a body of evidence, which describes a meaningful collection of interrelated beliefs, as the primitive representation. In contrast, all other technologies described in section 2 focus on individual propositions.

3.2 The analysis of evidence

To make the description more concrete, we trace through the analysis of the following simple problem.

> At 8:00 this morning I left for my office from my house in Palo Alto. At 9:00 I received a phone call from a San Mateo County police officer who informed me that someone in his district found my dog, Rufus, running loose. At 10:00, a coworker arrived and said that he saw, on his way to work, a dog that looked like Rufus cross Hwy 280. Rufus has run away twice before–once I found him in Los Altos and the other time in Menlo Park. Where should I look for Rufus?

In evidential reasoining the first step is to construct the sets of possibilities (the frames of discernment) of each unknown. For example, my dog Rufus could possibly be in any of the following cities:

$$\{Atherton, LosAltos, MenloPark, MountainView, PaloAlto, Sunnyvale\}$$

Other frames could also be constructed; we would probably want one for highways

$$\{Hwy101, Hwy280, elsewhere\},$$

and one for counties

$$\{SanMateo, SantaClara\}.$$

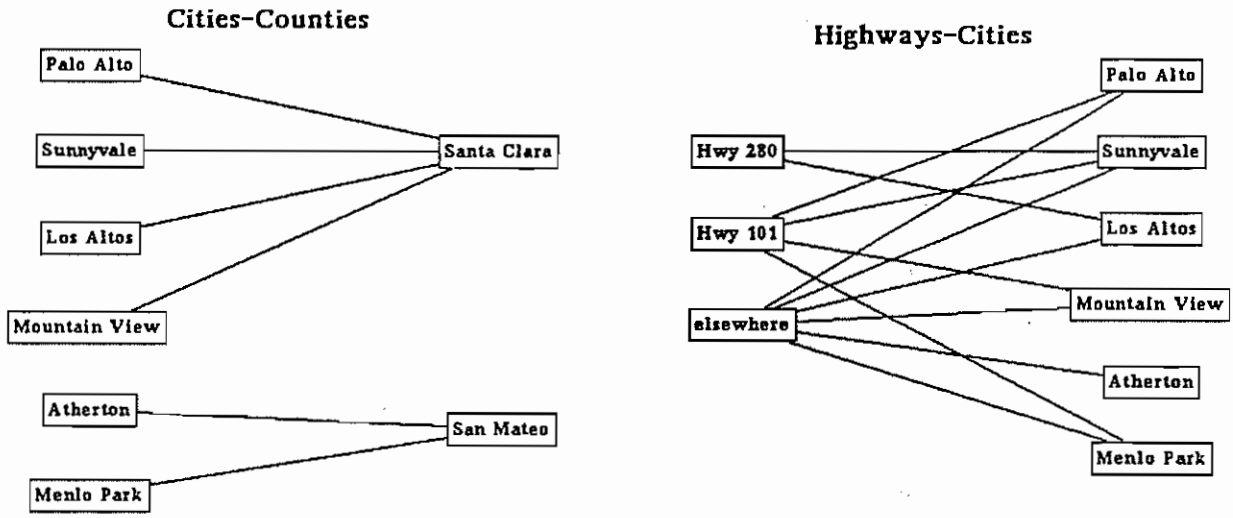


Figure 3.1: Compatibility relations.

The second step is to construct the compatibility relations that define the domain-specific relationships between the frames. Cities and counties are clearly related, so we might define the Cities-Counties relation graphically as shown in Figure 3.1. The relationship between cities and highways is also shown there. A connection between two propositions A_1 and B_1 indicates that they may co-occur (in other words, $(A_1,\ B_1) \in \Theta_{A,B}$).

Time dependencies can also be expressed by compatibility relations. We can construct a state transition diagram describing how far Rufus can wander. For example, suppose that in one hour it is possible for a dog to go from my home in Palo Alto to Los Altos, Menlo Park, or Mountain View. This information, along with travel times between other cities, can be expressed as the state transition graph in Figure 3.2, where the time interval for each arc is one hour. This graph can be interpreted as a compatibility relation, where each arc connects elements of the city frame to those cities where the dog could possibly be one hour later.

Once the frames and compatibility relations have been established, we can analyze the evidence. The goal of the analysis is to establish a line of reasoning from the evidence to determine belief in a hypothesis, (e.g., the present location of Rufus).

The first step is to assess each piece of evidence relative to an appropriate frame of discernment. Each piece of evidence is represented as a mass distribution, which distributes a unit of belief over subsets of the frame. For example, the fact that Rufus was at home when I left at 8:00 is pertinent to the *Cities* frame at 8:00 (*Cities*@8:00), and I would attribute 1.0 to *PaloAlto* to indicate my complete

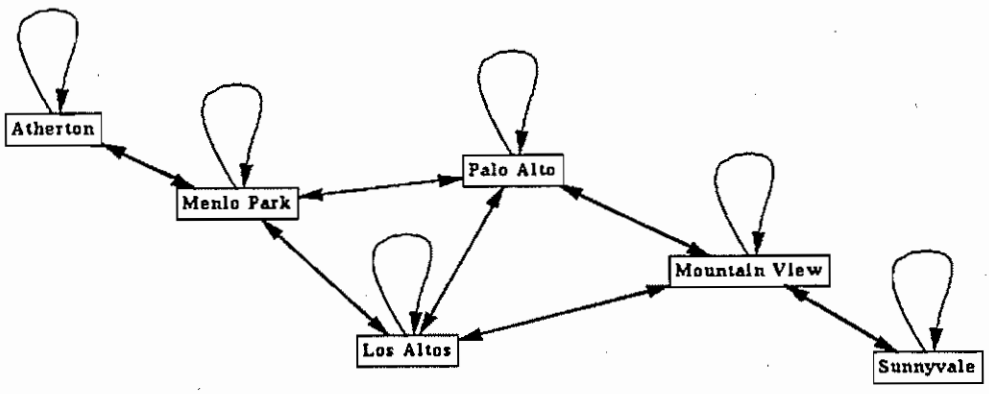


Figure 3.2: Compatibility relation resembling a state transition diagram.

certainty that he was there. The phone call from the policeman gives information about *Counties*@9:00, specifically that Rufus was in *SanMateo* at 9:00. Because this information is not as compelling as my knowledge of Rufus' whereabouts at 8:00, it is discounted to assess its true impact. Assuming that there is a 10% chance that the information is erroneous, we attribute 0.90 mass to *SanMateo*, and 0.10 to "anywhere." The third piece of evidence, that my coworker saw a dog like Rufus cross *Hwy*280, gave information about *Highways*@10:00 and, might be assessed as giving 0.65 support that it was Rufus crossing the road and 0.35 that my coworker couldn't see the dog well enough to identify him. The last piece of evidence (Rufus' previous escapes) weakly suggests that the dog may have returned to either Los Altos or Menlo Park. Each possibility is modeled as a mass distribution giving 0.25 to the city and 0.75 to "anywhere." This evaluation of evidence is quite subjective; however, when objective estimates are not possible, subjective estimates must suffice. For purposes of this paper, it is sufficient to accept some numeric estimate of belief, and we won't further discuss how these assessments should be made.

The final step is to construct the actual analysis of the evidence to determine its impact upon the question at hand. In this case the question can be answered by an assessment of belief over elements in the *Cities* frame at 10:00. The evidential operations can be used to derive a body of evidence providing beliefs about where Rufus might be at 10:00. A good starting point might be to pool the San Mateo

police report with the fact that Rufus was home at 8:00. Before we can combine these two bodies of evidence, we must adjust them to a common frame, say *Cities*@9:00.

Translating the police report to the *Cities* frame yields

$$Police_{Cities@9:00}(x) = \begin{cases} 0.90, & x = \{Atherton,\ MenloPark\} \\ 0.10, & x = \{\Theta_{Cities@9:00}\} \end{cases}$$

Projecting the BOE representing Rufus being at home at 8:00 to the *Cities* frame at 9:00 uses the **DELTA-Cities** relation and yields

$$Home_{Cities@9:00}(x) = 1.0, \quad x = \{LosAltos,\ MenloPark,\ MountainView,\ PaloAlto\}$$

These two independent BOEs are now represented relative to a common frame and can be combined using the *fusion* operation (i.e., Dempster's Rule). Fusing the two previous mass distributions yields:

$$m_{Cities@9:00}(x) = \begin{cases} 0.90, & x = \{MenloPark\} \\ 0.10, & x = \{LosAltos,\ MenloPark,\ MountainView,\ PaloAlto\} \end{cases}$$

The remainder of the evidence is taken into account by translating, projecting, and fusing according to the *analysis graph* shown in Figure 3.3. The result is a mass distribution relative to the *Cities* frame at 10:00, from which conclusions about Rufus' whereabouts can be drawn. Specifically,

$$m_{Cities@10:00}(x) = \begin{cases} 0.63, & x = \{LosAltos\} \\ 0.22, & x = \{LosAltos,\ MenloPark,\ PaloAlto\} \\ 0.08, & x = \{MenloPark\} \\ 0.04, & x = \{LosAltos,\ Sunnyvale\} \\ 0.02, & x = \{Atherton,\ LosAltos,\ MenloPark, \\ & \qquad Mountainview,\ PaloAlto,\ Sunnyvale\} \end{cases}$$

The associated evidential intervals for the atomic propositions in this mass distribution are:

$$\begin{aligned} [Spt(\{LosAltos\}),\ Pls(\{LosAltos\})] &= [0.63,\ 0.92] \\ [Spt(\{MenloPark\}),\ Pls(\{MenloPark\})] &= [0.08,\ 0.32] \\ [Spt(\{Atherton\}),\ Pls(\{Atherton\})] &= [0.00,\ 0.24] \\ [Spt(\{PaloAlto\}),\ Pls(\{PaloAlto\})] &= [0.00,\ 0.24] \\ [Spt(\{Sunnyvale\}),\ Pls(\{Sunnyvale\})] &= [0.00,\ 0.07] \\ [Spt(\{Mountainview\}),\ Pls(\{Mountainview\})] &= [0.00,\ 0.02] \end{aligned}$$

The hypothesis $\{LosAltos\}$ is clearly the most likely of all individual cities.

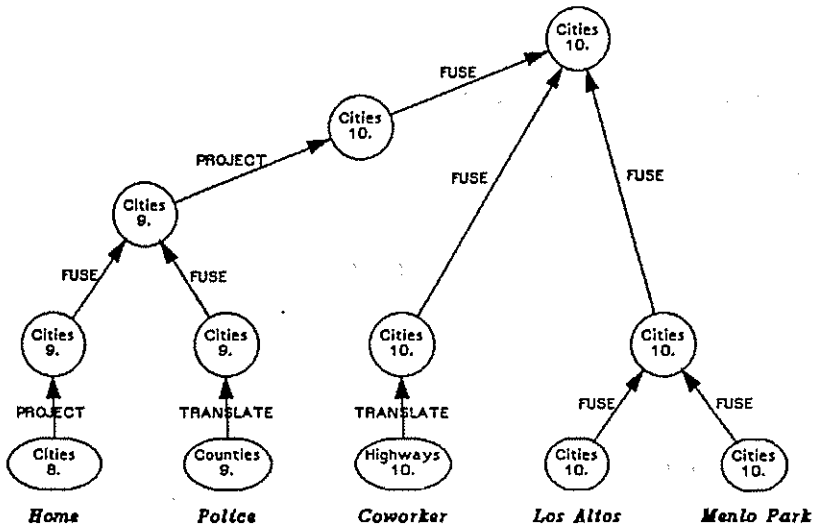


Figure 3.3: The completed analysis graph.

All the operations discussed above have been implemented within Gister. Frames and compatibility relations are represented as graphs, which can be constructed, examined, and modified interactively. Having an automated means to compute a conclusion is necessary. However, without some deeper explanation of why the conclusion is to be believed, it may be difficult to accept.

The completed analysis graph can be seen to be the counterpart of the proof tree of logical deduction. Each node represents an opinion, and the arcs trace the derivation of one opinion from other opinions and the knowledge contained in the compatibility relations. The complete graph shows the derivation of an ultimate conclusion from the primitive bodies of evidence. The next section presents a methodology that makes use of the analysis graph to explain evidential conclusions.

4: Generating Explanations Within Evidential Reasoning

We have already seen how the analysis graph can be construed as the evidential analog of a proof tree. In this section we will use it as a data structure that defines the information flow from primitive sources of evidence to conclusions. The interpretation of an analysis graph as a data-flow model provides intuitive appeal to the discussion that follows.

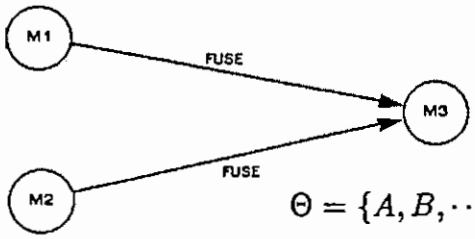
As was done with Hydro, we will use sensitivity analysis as the basis for constructing explanations. Because the belief function representation provides a richer vocabulary for expressing uncertainties than was used in Hydro, we will need a more sophisticated technique to identify the most significant justifications of a conclusion.

Sensitivity analysis requires a systematic variation of inputs to determine a family of solutions in the output space [9]. In Hydro, the probabilities of each piece of evidence are the relevant input parameters. In Gister, this is not feasible because the space of conceivable belief functions is exponentially large. Fortunately, a smaller, more intuitive parameter space is available—one that is motivated by the data-flow interpretation of the analysis graph. In particular, the credibility of each primitive body of evidence can be varied and the effect upon the conclusions of interest ascertained. This is accomplished by means of the *discounting* operation. The updated belief in a hypothesis can be computed by reevaluating the data-flow graph after discounting one (or more) of the primitive bodies of evidence on which it depends.

4.1 Single hypothesis

In this section, we develop the tools to explain why a particular hypothesis was found to be strongly (or weakly) supported. For example, we seek an answer to the question, "Why do you believe Rufus is in Los Altos at 10:00?"

The simplest case to consider is the fusion of two bodies of evidence, as shown below:



$$\Theta = \{A, B, \cdots\}$$

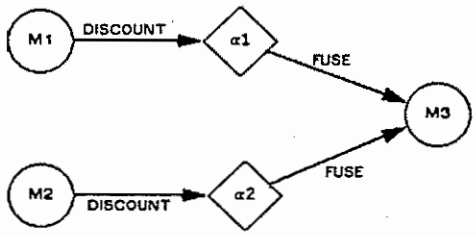
$$M1(x) = \begin{cases} 0.8, & x = A \\ 0.2, & x = \theta \end{cases} \qquad M3(x) = \begin{cases} 0.74, & x = A \\ 0.08, & x = B \\ 0.18, & x = \theta \end{cases}$$

$$M2(x) = \begin{cases} 0.3, & x = B \\ 0.7, & x = \theta \end{cases}$$

$$\begin{aligned} [Spt(A), Pls(A)]_{M3} &= [0.74, 0.92] \\ [Spt(B), Pls(B)]_{M3} &= [0.08, 0.26] \\ [Spt(A \vee B), Pls(A \vee B)]_{M3} &= [0.82, 1.00] \end{aligned}$$

To perform a sensitivity analysis of this graph, we insert a discounting node after each BOE representing primitive evidence. For each such BOE_i, we define α_i to be the credibility of that evidence, so that

$$\begin{aligned} \alpha_i = 1 &\implies \text{full impact of BOE}_i \\ \alpha_i = 0 &\implies \text{BOE}_i \text{ is ignored.} \end{aligned}$$



Obviously, if $\forall i, (\alpha_i = 1)$, then the computation in the modified analysis graph is the same as the ordinary fusion defined by the original graph. We are now in a position to answer "Why do you believe $[Spt(A), Pls(A)] = [0.74, 0.92]$?" The process consists of two steps:

1. Compute for each BOE_i

$$\widehat{Spt}_i(A) \doteq \left.\frac{\partial Spt(A)}{\partial \alpha_i}\right|_{\alpha_i=1} \qquad \widehat{Pls}_i(A) \doteq \left.\frac{\partial Pls(A)}{\partial \alpha_i}\right|_{\alpha_i=1}. \tag{4.1}$$

   Here, $\widehat{Spt}_i(A)$ is interpreted as the sensitivity of the support for A to BOE_i, and likewise for plausibility.

2. Identify those BOE_i with the extreme values.

The quantities in the preceding equations indicate the change in the support or plausibility relative to a change in the credibility of an evidence source. The partial derivative is evaluated at $\alpha_i = 1$ to assess the sensitivity of the conclusion, which was computed at $\alpha_i = 1$.

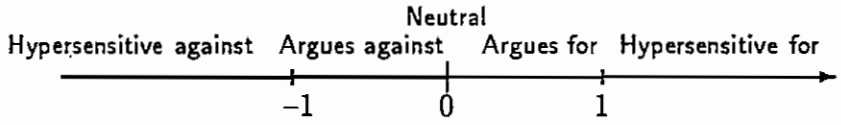
In theory these quantities can be computed algebraically or numerically; in practice numeric techniques are typically more practical. Returning to the previous example, we find

$$\widehat{Spt}_1(A) = \left.\frac{.8-.24\alpha_2}{(1-.24\alpha_1\alpha_2)^2}\right|_{\alpha_i=1} = 0.97$$

$$\widehat{Spt}_2(A) = \left.\frac{.192\alpha_1^2-.24\alpha_1}{(1-.24\alpha_1\alpha_2)^2}\right|_{\alpha_i=1} = -0.08.$$

From this information, it is apparent that BOE_1 is strong evidence in support of A, and BOE_2 weakly detracts from its support.

In general, the quantities $\widehat{Spt}_i(A)$ and $\widehat{Pls}_i(A)$ can be compared on the following scale



It can also be informative to interpret $\widehat{Spt}_i(A)$ and $\widehat{Pls}_i(A)$ with the aid of a *sensitivity space*, as illustrated in Figure 4.1. Plotting $\widehat{Spt}_i(A)$ and $\widehat{Pls}_i(A)$ in this space for each i yields a scatter plot that can be used to further analyze the results of the sensitivity computation. The farther a point is from the origin of sensitivity space, the greater the impact of the BOE that that point represents upon the conclusion. Entries in the northeast quadrant identify BOEs that support the proposition, A, while BOEs in the southwest quadrant argue against A. Points in the northwest signify BOEs that add to the confusion about the hypothesis, while the southeast quadrant identifies BOEs that argue both for and against the hypothesis.

So far, we have given examples only of a sensitivity analysis for a single fusion node. The techniques can be extended in a straightforward manner to apply across the full extent of an analysis graph. For example, the analysis in Figure 3.3 can be augmented with discounting nodes after each primitive evidence node. When the resulting analysis graph is viewed as a data-flow model, the discounting nodes can be seen to act as "valves," where lowering the α-value serves to diminish the flow of information through the valve.

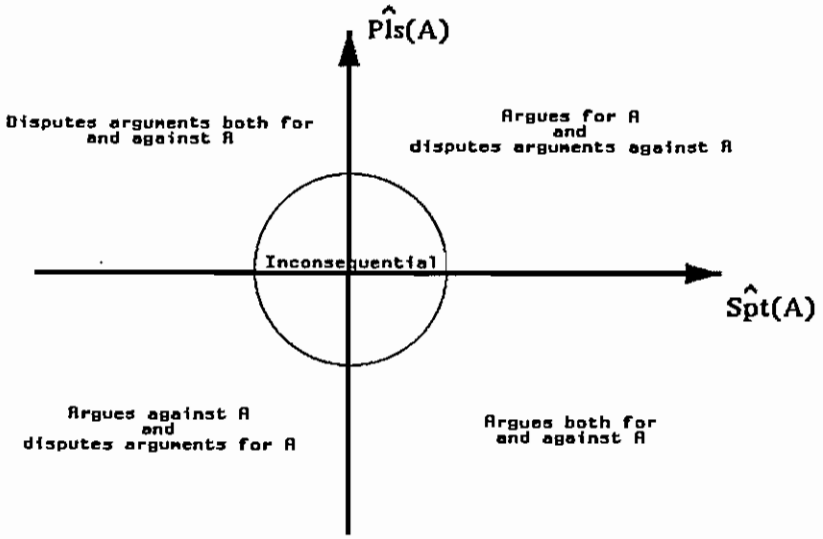


Figure 4.1: Sensitivity space for support and plausibility.

A discrete approximation to the quantities $\widehat{Spt}_i(A)$ and $\widehat{Pls}_i(A)$ can be obtained for any proposition A by systematically varying each of the α_is and reevaluating the data flow. This information then indicates the relevant import of each piece of primitive evidence. Plotting each point in sensitivity space yields a graphic illustration of the effect each body of evidence has upon the belief in a proposition.

Returning to the Rufus example, sensitivity analysis shows

$$\begin{array}{ll}
\widehat{Spt}_{Home}(LosAltos) = 0 & \widehat{Pls}_{Home}(LosAltos) = 0 \\
\widehat{Spt}_{Police}(LosAltos) = 0.40 & \widehat{Pls}_{Police}(LosAltos) = 0 \\
\widehat{Spt}_{Coworker}(LosAltos) = 0.43 & \widehat{Pls}_{Coworker}(LosAltos) = 0.12 \\
\widehat{Spt}_{MenloPark}(LosAltos) = -0.06 & \widehat{Pls}_{MenloPark}(LosAltos) = -0.08 \\
\widehat{Spt}_{LosAltos}(LosAltos) = 0.11 & \widehat{Pls}_{LosAltos}(LosAltos) = .02
\end{array}$$

From this information, which is plotted in Figure 4.2, we can conclude that my knowing that Rufus was at home at 8:00 had no bearing on the conclusion that he is probably in Los Altos now, while the information provided by the police and my coworker were the strongest pieces of evidence supporting Los Altos. Only the fact that Rufus had gone to Menlo Park once before argues against him being in Los Altos now. This information can be used to construct explanations to user queries:

```
Why do you believe Spt(Los Altos) = 0.63?
```

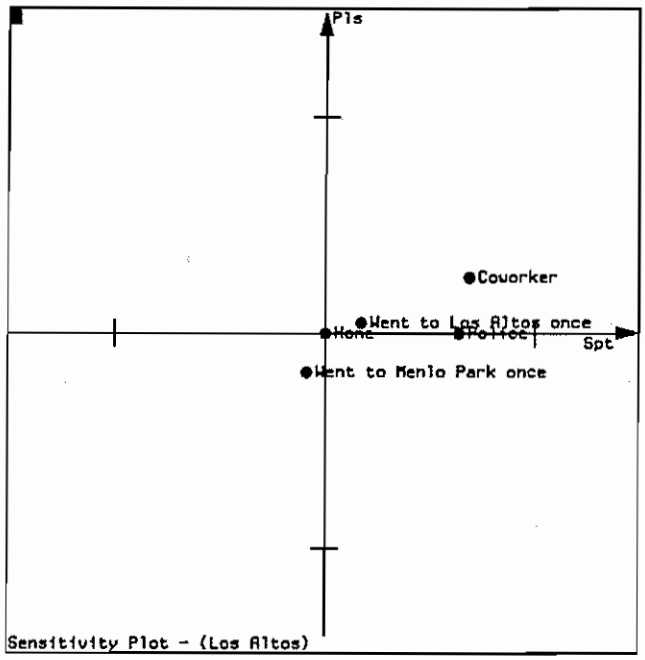


Figure 4.2: Plot of the sensitivities of Spt(Los Altos) and Pls(Los Altos) from the lost-dog story.

```
Because
the police reported that Rufus was seen in San Mateo County at 9:00, and
my coworker reported seeing a dog that looks like Rufus along Highway 280, and
Rufus went to Los Altos once before.
```

Another example uses negativity of $\widehat{Pls}_i(LosAltos)$ to answer a question:

```
Is there any reason to believe that Rufus is not in Los Altos?

Yes.
Rufus went to Menlo Park once before.
```

If the user desires a more complete response than this, we can construct an explanation from those compatibility relations that were used along any particular path in the graph. A natural language text that describes what the compatibility relation encodes might suffice (e.g., **DELTA-Cities** is "the limits on how far a dog can travel in one hour"); otherwise, the identification of particular links in the relation (perhaps graphically) can help pinpoint a reason.

This analysis indicates only the effect of each primitive piece of evidence individually; the joint effect of multiple bodies of evidence is not determined. Computing joint effects numerically, while straightforward theoretically, requires exploration of a combinatorically large parameter space.

4.2 Entire body of evidence

Explanations of a single hypothesis (such as those derived in the preceding section) are quite similar to those produced in systems based on certainty factors or inference nets. The notion of a body of evidence that is used in evidential reasoning permits a higher-level description of an inference chain. Rather than asking a question about a belief in a particular proposition, the user can pose questions that search for the primitive pieces of evidence that were the most influential in general.

There have been numerous proposals for characterizing BOEs [4] that can be used as the basis for selecting informative explanations. While nearly any sound characterization will suffice for our present purposes, we will make use of several due to Yager [17].

We have already noted that the theory of belief functions allows representation of varying degrees of precision as well as uncertainty. The relative precision of a BOE can be characterized by the following expression for *specificity*:

$$Spec(m_\Theta) = \sum_{A_j \subseteq \Theta} \frac{m_\Theta(A_j)}{\|A_j\|}, \text{ where } \|A_j\| \text{ is the cardinality of the subset } A_j. \tag{4.2}$$

It is easy to show that

$$0 < \frac{1}{\|\Theta\|} \leq Spec(m_\Theta) \leq 1, \text{ for any mass distribution } m_\Theta.$$

Roughly speaking, $Spec(m_\Theta)$ measures the degree of commitment of a belief function to precise propositions, assuming that each element of Θ is equally precise. The vacuous belief function, $m_\Theta : m_\Theta(\Theta) = 1$, has the smallest possible specificity for any frame Θ. A mass distribution whose specificity is 1 is a classical probability distribution as well.

The relative uncertainty of a BOE can be characterized by an entropy-like measure. Yager defines

$$Ent(m_\Theta) \doteq -\sum_{A_j \subseteq \Theta} m_\Theta(A_j) \cdot \ln Pls(A_j) \tag{4.3}$$

and shows that $Ent(m_\Theta)$ is just Shannon's measure of entropy in the special case when m_Θ is a probability distribution. To use this measure to generate explanations, it will be more convenient to work instead with a measure of *consonance*:

$$Cons(m_\Theta) \doteq \frac{1}{1 + Ent(m_\Theta)}, \tag{4.4}$$

so that

$$0 < Cons(m_\Theta) \leq 1.$$

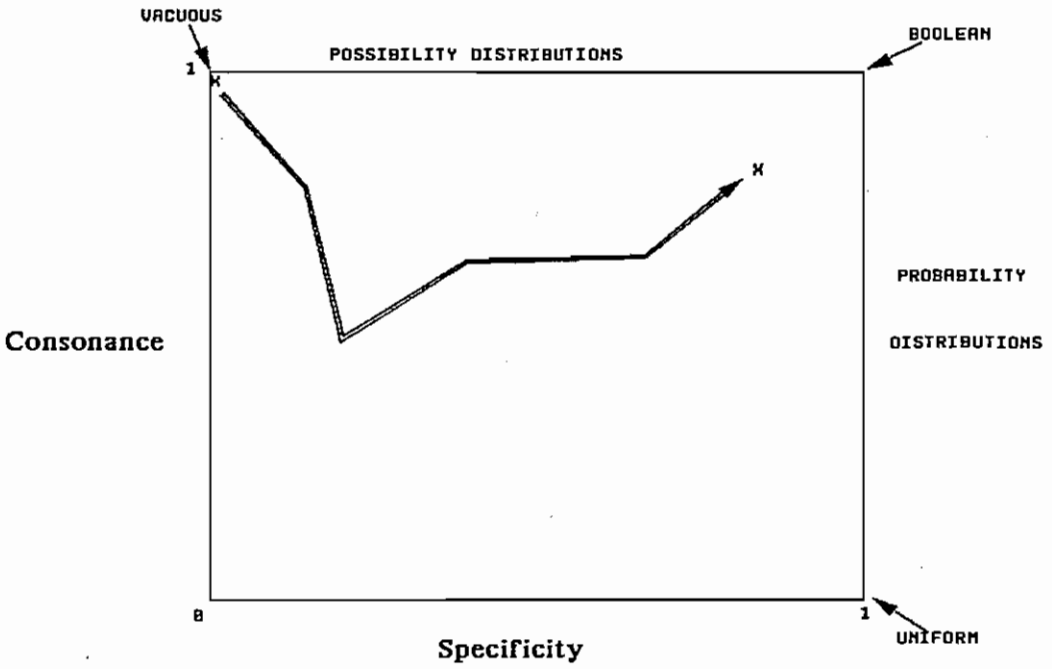


Figure 4.3: The characterization of mass distributions in terms of specificity and consonance.

Minimal consonance is thus maximal entropy, and exists whenever the focal elements of a mass distribution are mutually exclusive. Consonance equal to 1 occurs when all the focal elements are nested and thus represents a possibility distribution as defined by fuzzy set theory [13], [17], [18].

To gain some intuition, it is useful to note that any BOE is characterized by a point in the unit square shown in Figure 4.3. The special cases of possibility distributions and probability distributions lie on the boundaries of the square. A Boolean statement has $Cons(m) = Spec(m) = 1$. The vacuous belief function has $Cons(m) = 1$ and $Spec(m) = 0$ and is represented by the upper-left corner of the square. Starting with no information and gradually fusing pieces of evidence as they become available, we trace a path in the square that starts at the upper-left corner and wanders through the space. The ideal analysis would reach a Boolean conclusion (upper-right corner), but typically the path stops somewhere short. The intuition, then, is that pieces of evidence that move the path closer to the upper-right corner are the most important ones for making decisions.

We are now in a position to select pieces of evidence as justification for an evidential-reasoning inference chain. As before, we will perform a sensitivity anal-

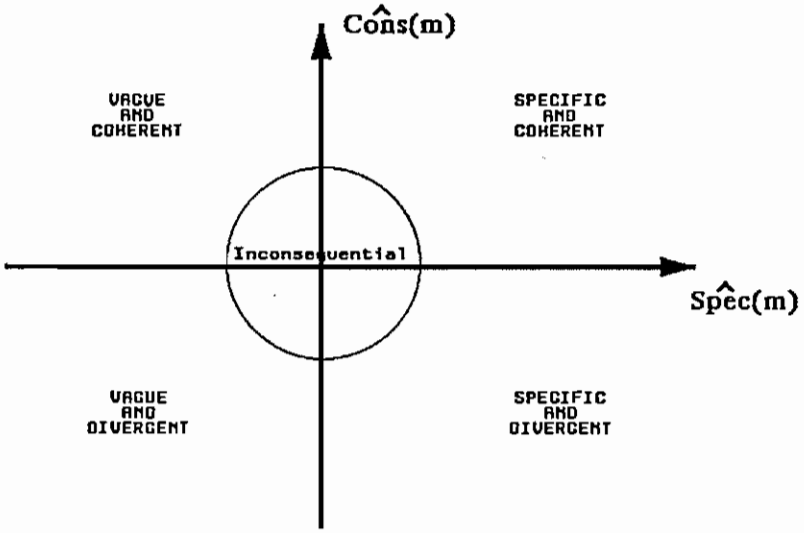


Figure 4.4: Sensitivity space for characterizations of a body of evidence.

ysis to choose the components of the explanation, but this time we will measure the change in our two characterizations of a BOE. We define

$$\widehat{Spec}_i(m) \doteq \left.\frac{\partial Spec(m)}{\partial \alpha_i}\right|_{\alpha_i=1} \quad \text{and} \quad \widehat{Cons}_i(m) \doteq \left.\frac{\partial Cons(m)}{\partial \alpha_i}\right|_{\alpha_i=1} \tag{4.5}$$

as the sensitivity of specificity and consonance respectively, where α_i is the credibility of BOE_i as before. Once again, these measures can be computed for each item of primitive evidence and plotted in sensitivity space for comparison (see Figure 4.4).

In this graph, the northeast quadrant represents those BOEs whose inclusion in an analysis forces the path to the upper-right (the Boolean case) and are therefore important for making decisions. The southwest quadrant contains BOEs whose inclusion decreases both the consonance and specificity—these are pieces of evidence that run counter to the consensus, and may be suggestive of an errorful source or a need to use case-based reasoning by maintaining multiple analysis paths. The other quadrants can be interpreted as labeled. Once again, distance from the origin indicates the relative contribution of evidence to the conclusion. Appendix A contains examples illustrating evidence that falls in each of the four quadrants.

Sensitivity analysis for the BOE that represents the conclusion from the lost-

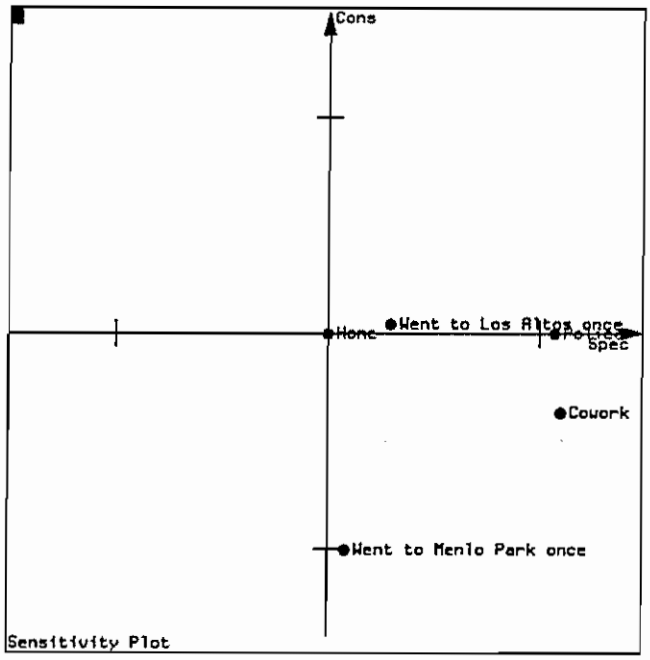


Figure 4.5: Plot of the sensitivity of specificity and consonance from the lost-dog story.

dog story reveals

$$\begin{array}{ll} \widehat{Spec}_{Home}(m) = 0.00 & \widehat{Cons}_{Home}(m) = 0.00 \\ \widehat{Spec}_{Police}(m) = 0.22 & \widehat{Cons}_{Police}(m) = 0.00 \\ \widehat{Spec}_{Coworker}(m) = 0.25 & \widehat{Cons}_{Coworker}(m) = -0.05 \\ \widehat{Spec}_{MenloPark}(m) = 0.02 & \widehat{Cons}_{MenloPark}(m) = -0.13 \\ \widehat{Spec}_{LosAltos}(m) = 0.06 & \widehat{Cons}_{LosAltos}(m) = 0.01 \end{array}$$

From these sensitivities (plotted in Figure 4.5) it is clear that the fact that Rufus was at home at 8:00 did not contribute to the conclusion. The sensitivities of specificity indicate that the conclusion would have been much less specific if either the coworker's report or the police report were excluded from the analysis—these pieces of evidence were instrumental in establishing Los Altos as the strongest possibility. Looking at the consonance values, the fact that Rufus had gone to Menlo Park once before argues against the consensus most strongly, while the coworker's report disputes the argument for Menlo Park, which would have been the consensus of the other four reports. These results can be used to answer questions about the analysis:

```
Which reports can be safely ignored?

Rufus was at home at 8:00.
```

```
Which reports argue against the consensus?

Rufus went to Menlo Park once before, and
my coworker reported seeing a dog that looks like Rufus along Highway 280.
```

4.3 Using sensitivity results to generate explanations

With these tools in hand, a number of different questions about an analysis can be answered. In the following, Q indicates a question that might be posed by the user, and P indicates a procedure that can be used to construct an answer to that question.

Q: Why do you strongly believe A?

(I.e., Which report argues most strongly for A?)

P: Choose the BOE_i for which $\widehat{Spt}_i(A)$ is greatest.

Q: Why don't you believe B?

(I.e., Which report argues most strongly against B?)

P: Choose the BOE_i for which $\widehat{Pls}_i(B)$ is most negative.

Q: Which reports serve to focus the conclusion more precisely?

(I.e., Which bodies of evidence cause the conclusion to become more specific and more coherent?)

P: Choose those BOE_i for which $\widehat{Spec}_i(m)$ and $\widehat{Cons}_i(m)$ are both positive.

Q: Which report most strongly disagrees with the consensus?

(I.e., Which body of evidence serves to make the conclusion the most divergent?)

P: Choose the BOE_i for which $\widehat{Cons}_i(m)$ is most negative.

Q: Which reports can be safely ignored?

(I.e., Which bodies of evidence are inconsequential?)

P: Choose those BOE_i for which $\widehat{Spec}_i(m) \approx \widehat{Cons}_i(m) \approx 0$.

Q: What are the most influential reports that impinge upon this conclusion?

(I.e., Which bodies of evidence are farthest from the origin in sensitivity space?)

P: Choose those BOE_i for which $\left(\widehat{Spec}_i(m)\right)^2 + \left(\widehat{Cons}_i(m)\right)^2$ is greatest.

Throughout this paper we have used the Rufus example to illustrate our use of sensitivity analysis for constructing explanations of evidential analyses. A more thorough example of these explanation techniques, applied to a more complicated evidential domain, is detailed in Appendix B.

5: Discussion

The three requirements of explanation generation from section 2 have been satisfied by our approach:

1. The difficulty of recapitulating program actions within systems that use a numeric measure of uncertainty has been overcome by the use of sensitivity analysis. By varying the input parameters and recomputing an analysis, the system can explain the interaction of the evidence and its impact upon selected conclusions. Focusing on the credibility of bodies of evidence, instead of probabilities of individual propositions, makes sensitivity analysis of belief functions feasible.

2. The correct level of detail can be controlled in two ways. First, the depth of exploration can be confined to a selected subtree within the analysis. When additional detail is requested, the subtree can be expanded to reveal previously hidden details. Second, the number of justifications to be provided is adjusted by rank ordering the sensitivities and choosing the most important ones.

3. A shared vocabulary is also provided in two forms. As with the other technologies, natural language text is associated with each primitive evidence node and displayed in place of the machine representation. Second, the vocabulary is in terms of the high-level constructs of a set of related beliefs represented by bodies of evidence, instead of each proposition and its belief individually.

The use of evidential reasoning provides a richer vocabulary for expressing belief about uncertain events than is available in most other technologies, but confounds the construction of suitable explanations of chains of inference. The use of sensitivity analysis as described here not only permits the customary forms of explanation characteristic of rule-based systems, but also enables a variety of additional queries to be posed and answered.

The tools presented in this paper have several uses in addition to that of constructing explanations for a user. Sensitivity information can be an important component of decision analysis. Knowledge of the sensitivity of conclusions can suggest whether sufficient information is available, or whether additional information should be sought. It can also be used to focus information-collection efforts. By hypothesizing the information that might be collected by a particular source, one can determine whether it could possibly have sufficient impact on the

hypothesis to alter a pending decision. These ideas, while promising, have not yet been investigated.

6: Conclusions

We have presented an approach to constructing an answer to various kinds of questions that can be asked about a conclusion derived through evidential reasoning. We have argued that the technique satisfies the three requirements for explanations. It also has the generality to be able to provide a variety of information about an evidential inference chain and can be used to further insulate the user from the cryptic numbers that are manipulated by the machine. Coupling this mechanism with the evidential-reasoning techniques already developed allows the creation of a powerful knowledge-based system for reasoning under uncertainty that can explain its behavior in understandable terms.

7: Acknowledgments

The authors wish to thank the members of SRI's Artificial Intelligence Center who read and critiqued earlier drafts of this paper. In particular, Thomas Garvey provided keen insight on the calculation and interpretation of the sensitivity measures. Discussions with Steven Lesh were valuable for understanding the role sensitivity analysis might play in decision analysis, and Leonard Wesley illuminated the relationship between this paper and his thesis on evidential control.

Bibliography

[1] Barr, Avron, and Feigenbaum, Edward A., ed., *The Handbook of Artificial Intelligence*, Vol. 1, William Kaufmann, Inc., Los Altos, California, 1981.

[2] Davis, Randall, and Lenat, Douglas B., *Knowledge-Based Systems in Artificial Intelligence*, McGraw-Hill, New York, 1982.

[3] Dempster, Arthur P., "A Generalization of Bayesian Inference," *Journal of the Royal Statistical Society* **30**(Series B), 1968, pp. 205–247.

[4] Dubois, Didier, and Prade, Henri, "Properties of Measures of Information in Evidence and Possibility Theories," Actes Journées *Analyse de problèmes décisionelles dans un environment incertain et imprécis*, Reims, France, July 11-13, 1985.

[5] Gaschnig, John, Reboh, Rene, and Reiter, John, "Development of a Knowledge-Based Expert System for Water Resource Problems," Final Report, SRI Project 1619, August 1981.

[6] Lowrance, John D., and Garvey, Thomas D., "Evidential Reasoning: A Developing Concept," *Proceedings of the IEEE International Conference on Cybernetics and Society*, October 1982, pp. 6–9.

[7] Lowrance, John D., Garvey, Thomas D., and Strat, Thomas M., "A Framework for Evidential-Reasoning Systems," *Proceedings AAAI-86*, Philadelphia, Pennsylvania, August 1986.

[8] Pearl, Judea, "Fusion, Propagation, and Structuring in Bayesian Networks," Tech. Report CSD-850022, Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles, June 1985.

[9] Radanovic, L., ed., *Sensitivity Methods in Control Theory*, Proc. International Symposium, Pergamon Press, Dubrovnik, Yugoslavia, September 1964.

[10] Reboh, Rene, "Knowledge Engineering Techniques and Tools in the Prospector Environment," Technical Note 243, Artificial Intelligence Center, SRI International, Menlo Park, California, June 1981.

[11] Reboh, Rene, "Extracting Useful Advice from Conflicting Expertise," *Proceedings IJCAI-83*, Karlsruhe, Federal Republic of Germany, August 1983, pp. 145–150.

[12] Shafer, Glenn A., *A Mathematical Theory of Evidence*, Princeton University Press, New Jersey, 1976.

[13] Shafer, Glenn A., "Belief Functions and Possibility Measures," in *Analysis of Fuzzy Information – Volume I,* Bezdek, James C., ed., CRC Press, Boca Raton, Florida, 1986.

[14] Shortliffe, Edward H., *Computer-Based Medical Consultations: MYCIN*, American Elsevier, New York, 1976.

[15] Sterling, Leon, and Shapiro, Ehud, *The Art of Prolog*, MIT Press, Cambridge, Massachusetts, 1986.

[16] Swartout, William R., "Explaining and Justifying Expert Consulting Programs" *Proc. 7th IJCAI*, Vancouver, B.C., 1981, pp. 815–822.

[17] Yager, Ronald R., "Entropy and Specificity in a Mathematical Theory of Evidence" *Int. J. General Systems*, Vol. 9, 1983, pp. 249–260.

[18] Zadeh, Lotfi A., "Fuzzy Sets as a Basis for a Theory of Possibility," *Fuzzy Sets and Systems*, Vol. 1, 1978, pp. 3–28.

A: An Exploration of Sensitivity Space

In this appendix we give examples of evidence that falls in each of the four quadrants of sensitivity space (Figure A.1). By studying each example, one can gain a better understanding of the meaning of sensitivities for more complicated analyses.[1]

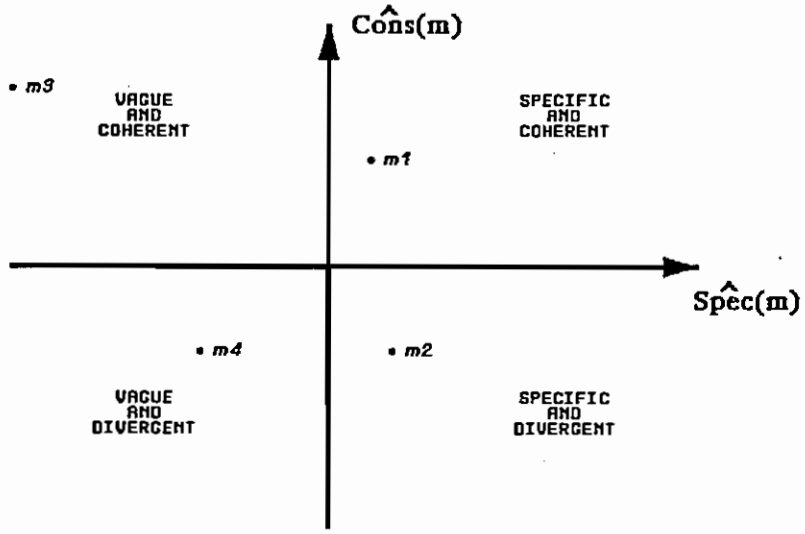


Figure A.1: Sensitivity space plot for automobile example.

The question under consideration is the make of car that Fred is going to buy. He has gone to a used car dealer who only has 5 cars available. The frame of discernment is the set containing the make of each car:

{Mercedes 280SL, Porsche 944, Ford Escort, Ford Mustang, Ford Taurus}.

In each case, we will fuse a mass distribution defined over this frame with the following mass distribution, which reflects our initial assessment of Fred's choice:

$$m0(x) = \begin{cases} 0.90, & x = \{\text{Porsche 944}\} \\ 0.10, & x = \{\text{Ford Escort, Ford Mustang, Ford Taurus}\}. \end{cases}$$

[1]This plot shows superimposed the sensitivities of the four mass distributions relative to four different conclusions. The other sensitivity plots in this paper show sensitivities of BOEs relative to a single conclusion.

For $m0$, we have $Spec(m0) = 0.933$ and $Cons(m0) = 0.755$; i.e., we strongly believe Fred will get the Porsche 944, we weakly believe that he might get one of the Fords, and we're sure that he won't get the Mercedes.

A.1 Northeast Quadrant: Specific and Coherent

We fuse $m0$ with $m1$, a mass distribution independently assessed by a neighbor, that just happens to be equivalent to $m0$. This fusion yields

$$m01(x) = \begin{cases} 0.99, & x = \{\text{Porsche 944}\} \\ 0.01, & x = \{\text{Ford Escort, Ford Mustang, Ford Taurus}\} \end{cases}$$

$$Spec(m01) = 0.992 \text{ and } Cons(m01) = 0.938$$

We compute $\widehat{Spec}_{m1}(m01) = 0.059$ and $\widehat{Cons}_{m1}(m01) = 0.183$. It can be seen that inclusion of $m1$ makes the conclusion more specific and more coherent, as expected.

A.2 Southeast Quadrant: Specific and Divergent

We will fuse $m0$ with the following mass distribution assessed by a different neighbor, who strongly believes Fred will get the Mustang and weakly suggests that he will get one of the foreign cars.

$$m2(x) = \begin{cases} 0.90, & x = \{\text{Ford Mustang}\} \\ 0.10, & x = \{\text{Porsche 944, Mercedes 280SL}\} \end{cases}$$

Their fusion yields equal probability for it being either a Mustang or a Porsche:

$$m02(x) = \begin{cases} 0.50, & x = \{\text{Ford Mustang}\} \\ 0.50, & x = \{\text{Porsche 944}\} \end{cases}$$

$$Spec(m02) = 1.0 \text{ and } Cons(m02) = 0.591$$

We compute $\widehat{Spec}_{m2}(m02) = 0.067$ and $\widehat{Cons}_{m2}(m02) = -0.164$. It can be seen that inclusion of $m2$ makes the conclusion more specific (since all the belief is now attributed to individual cars) but less coherent (because the belief is divided evenly among disjoint propositions) than the original beliefs, $m0$.

A.3 Northwest Quadrant: Vague and Coherent

We fuse $m0$ with the following mass distribution, which strongly supports Mercedes 280SL and weakly supports Ford:

$$m3(x) = \begin{cases} 0.90, & x = \{\text{Mercedes 280SL}\} \\ 0.10, & x = \{\text{Ford Escort, Ford Mustang, Ford Taurus}\} \end{cases}$$

Their fusion yields complete belief in Ford: (Although neither mass distribution held that Ford was very likely, it was the only proposition that both admitted as a possibility.)

$$m03(x) = \begin{cases} 1.0, & x = \{\text{Ford Escort, Ford Mustang, Ford Taurus}\} \end{cases}$$

$$Spec(m03) = 0.333 \text{ and } Cons(m03) = 1.0$$

This results in $\widehat{Spec}_{m3}(m03) = -0.600$ and $\widehat{Cons}_{m3}(m03) = 0.245$. It can be seen that inclusion of $m3$ makes the conclusion less specific (because mass has "moved" from a proposition containing only one element to a proposition that is a disjunction of three elements) but more coherent (because now all the mass is focused upon a single proposition).

A.4 Southwest Quadrant: Vague and Divergent

We fuse $m0$ with the following mass distribution, which indicates strong belief for Mercedes and weak belief equally divided between Mustang and the other Fords:

$$m4(x) = \begin{cases} 0.90, & x = \{\text{Mercedes 280SL}\} \\ 0.05, & x = \{\text{Ford Mustang}\} \\ 0.05, & x = \{\text{Ford Escort, Ford Taurus}\} \end{cases}$$

Their fusion results in complete belief in Ford, with equal probability of whether it will be a Mustang or not:

$$m04(x) = \begin{cases} 0.50, & x = \{\text{Ford Mustang}\} \\ 0.50, & x = \{\text{Ford Escort, Ford Taurus}\} \end{cases}$$

$$Spec(m04) = 0.750 \text{ and } Cons(m04) = 0.591$$

This yields $\widehat{Spec}_{m4}(m04) = -0.183$ and $\widehat{Cons}_{m4}(m04) = -0.164$. It can be seen that inclusion of $m4$ makes the conclusion less specific (because now only half the belief is attributed to an individual car) and less coherent (because the belief is evenly divided between two disjoint propositions).

B: Detective Example

In this appendix we provide an extensive collection of explanations constructed as answers to various questions by using the techniques described in the body of the paper. The figures are taken directly from the screen images produced by the implementation of these techniques within Gister. As of this writing, emphasis has been placed on providing a versatile explanation facility rather than one in which the operations have been prepackaged into their most user-friendly form. It is hoped that experience gained with the present implementation will lead to a better understanding of the important criteria for evidential explanations. For this reason, the actual facilities described should be viewed as a substrate upon which a user-friendly explanation system could be constructed.

The following detective story is used throughout the appendix for the sake of illustrating the range of explanatory information available. It has been chosen on the basis of being small enough for the casual reader to understand, yet contains enough richness to illustrate the scope of the explanation techniques. Of course, the need for explanation facilities is much more acute when constructing large analyses.

The Case of the Sweetshop Burglary

When Mike arrived to open the sweetshop on Thursday morning, he found the safe in the office open and the receipts from the previous day ($645) missing. The burglar apparently used a key to enter the shop and knew the combination of the safe. A witness who lives across the street saw what appeared to be a man entering the sweetshop at the approximate time of the burglary.

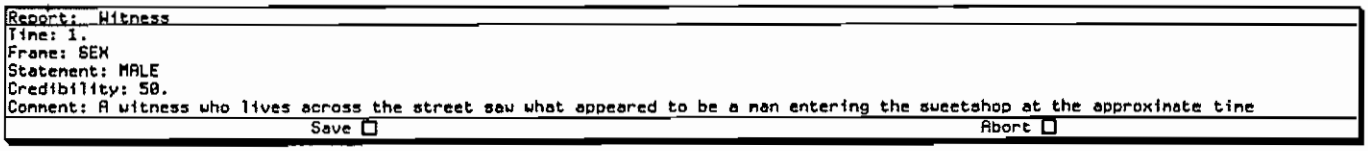
Upon further investigation, we find that:

- The only people with keys and knowledge of the combination are the sweetshop employees, Ann, David, Frank, Judy, and Mike.
- Ann is over-extended on her credit cards.
- Frank broke his leg and is still using crutches to get around.
- Mike recently started dating Judy, who was David's former girlfriend.
- Judy's mother says that Judy was home all night.
- The police find that although the fingerprints on the safe are smudged, it can be determined that they are from someone's left hand. However, they cannot be sure that the prints are from the burglar. David and Ann are the only left-handed employees.

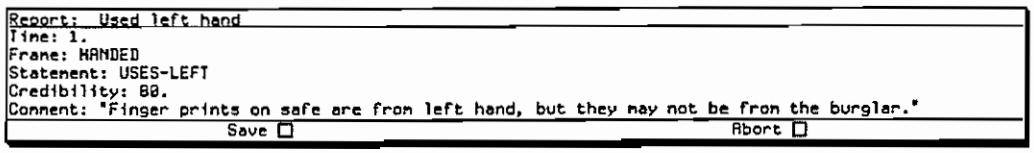
Who is the Burglar?

The remainder of this appendix illustrates the result of analyzing this case using the evidential reasoning technology embodied in Gister. For our current purposes, we are not as concerned with accurate numerical assessment of uncertainties as we are with deriving sound inferences given those uncertainties and extracting meaningful explanations based upon those derivations.

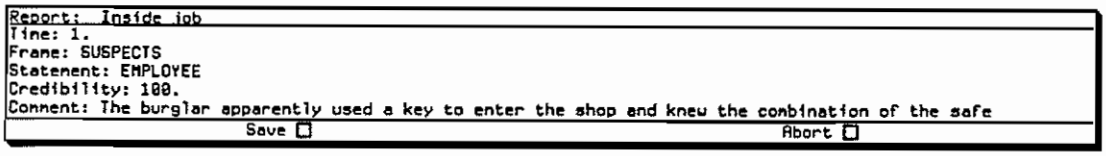
Figure B.1: Using Gister, the detective enters his assessment of each piece of available evidence. A summary of his thoughts is provided for each assessment.



(a) The detective feels there is a fifty-fifty chance that the man who was seen was not the burglar. This yields the mass distribution $\{m(Male) = 0.5,\ m(Male \vee Female) = 0.5\}$ for the sex of the burglar.



(b) The fingerprints are certainly from someone's left hand, but they may or may not be from the burglar. The probability that they are from the burglar is estimated to be 0.8.



(c) Because there was no sign of forced entry, we assume that the crime was perpetrated by someone with a key and knowledge of the safe's combination (i.e., an employee).

```
Primitive:  Financial motive
FOD: (SUSPECTS 1.)
USER-MASS: (((OTHER ANN) 50.) ((MIKE OTHER) 20.) ((OTHER JUDY) 10.) ((OTHER FRANK) 10.) ((OTHER DAVID) 10.))
MASS: (((OTHER ANN) 0.5) ((MIKE OTHER) 0.19999999) ((OTHER JUDY) 0.099999994) ((OTHER FRANK) 0.099999994) ((OTHER DAVID) 0.099999994))
COMMENT: "Ann is over extended on her credit cards.
                Save ☐                                          Abort ☐
```

(d) Ann is overextended on her credit cards so she certainly had a financial motive. But the other employees had financial motives as well, although to lesser degrees. Clearly, others (non-employees) had financial motives of varying degrees besides.

```
Primitive:  Revenge motive
FOD: (SUSPECTS 1.)
USER-MASS: (((OTHER DAVID) 40.) ((OTHER MIKE JUDY FRANK DAVID ANN) 60.))
MASS: (((OTHER MIKE JUDY FRANK DAVID ANN) 0.6) ((OTHER DAVID) 0.4))
COMMENT: "Mike recently started dating Judy, who was David's previous girlfriend."
                Save ☐                                          Abort ☐
```

(e) David may have burgled the safe to get back at Mike, who is now dating David's former girlfriend. This tends to weakly (0.4) impugn David. The other employees are less likely to have revenge motives.

```
Report:  Judy's alibi
Time: 1.
Frame: SUSPECTS
Statement: (NOT JUDY)
Credibility: 20.
Comment: "Mother says that she was home all night."
                Save ☐                                          Abort ☐
```

(f) Although Judy's mother claims Judy was home all night, she may well be lying to protect Judy. It is assessed to be only 20% credible as an alibi for Judy.

```
Report:  Frank's injury
Time: 1.
Frame: SUSPECTS
Statement: (NOT FRANK)
Credibility: 80.
Comment: "Frank is still using crutches to get around; he broke his leg."
                Save ☐                                          Abort ☐
```

(g) It seems highly unlikely that Frank could have done it because he has a broken leg and is still on crutches.

Figure B.2: Analysis graph constructed for the case of the sweetshop burglary (next page).

The menus on the left contain the evidential reasoning operations used by the detective in constructing the analysis graph. Elliptical nodes encode individual bodies of evidence. Circular nodes represent derived conclusions. The diamond marked "Int" computes the evidential intervals of some selected propositions from "Conclusion." These intervals are displayed in the lower portion of the screen. The numbers in brackets are [Support, Plausibility] for each suspect—the shaded areas reflect these bounds on a scale from 0.0 to 1.0.

In constructing this analysis, the detective first combined Judy's alibi and Frank's injury to form an intermediate conclusion based on alibis. Next he discounted by 20% the financial motive (to account for the fact that the amount of money in the safe may have been too small to motivate the burglar) and combined it with the evidence representing the revenge motive to form the intermediate conclusion based on motives. Next he assembled the physical evidence by translating the witness report from the SEX frame of discernment to the SUSPECTS frame; translating the fingerprint evidence from the HANDEDNESS frame to the SUSPECTS frame; and combining them with the evidence regarding the burglary as an inside job. Finally, a conclusion was drawn by combining this physical evidence with the intermediate conclusions involving motives and alibis. Of course, the evidence could have been assembled in any order because the fusion operation is commutative and associative, but this choice does allow some meaningful intermediate conclusions.

From this information, the detective concludes that David is the most likely suspect. But which pieces of evidence fingered him, and how sensitive is this conclusion to the assessments made by the detective?

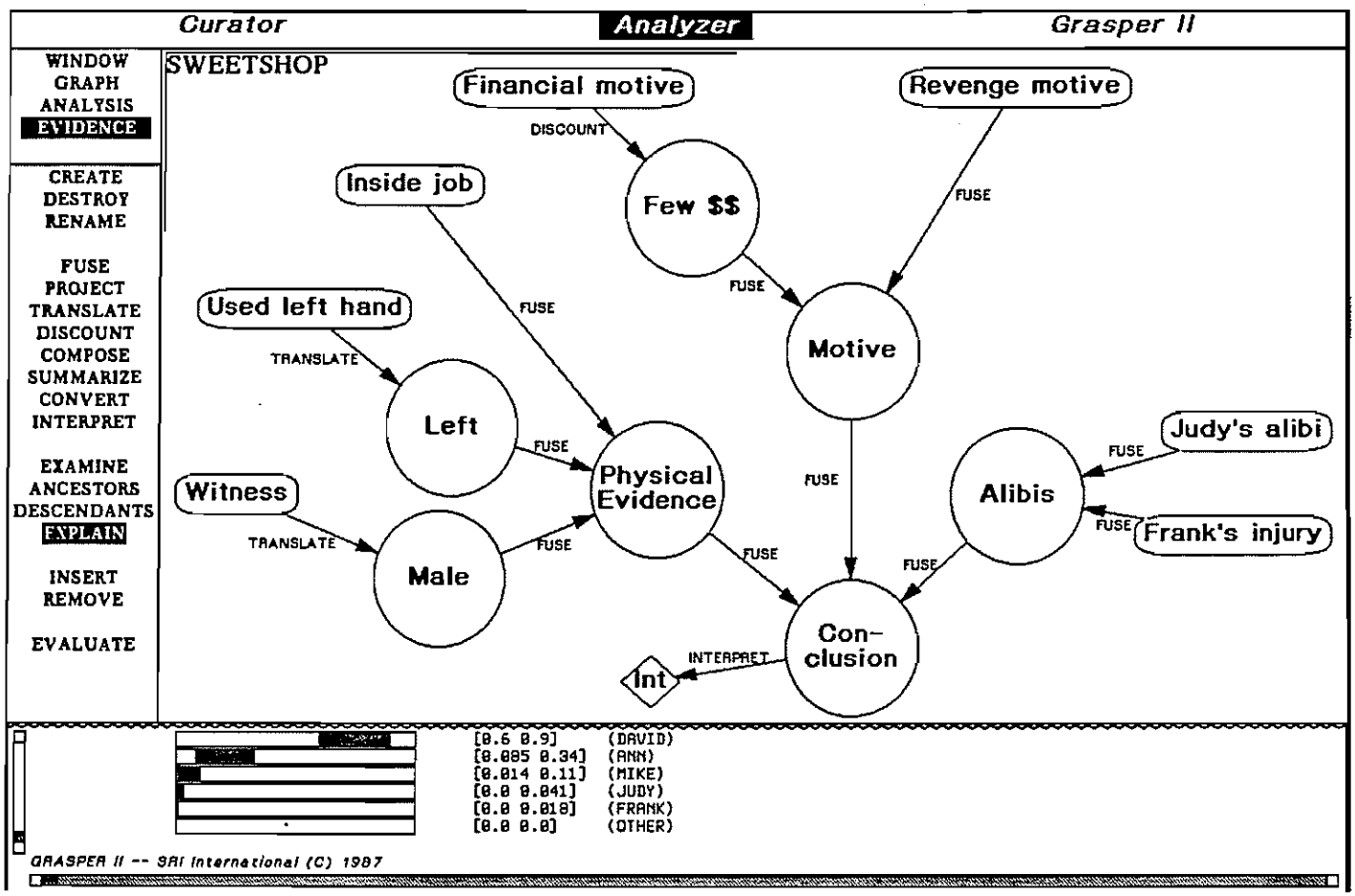
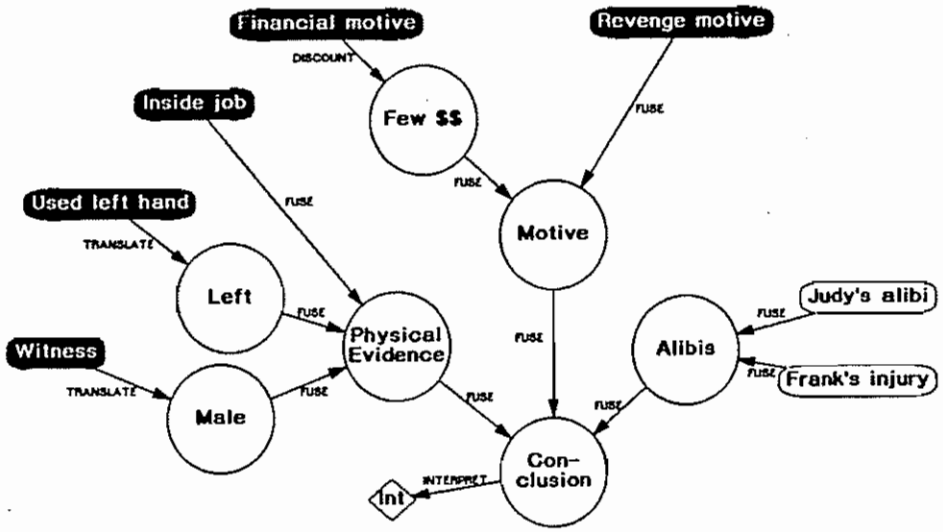
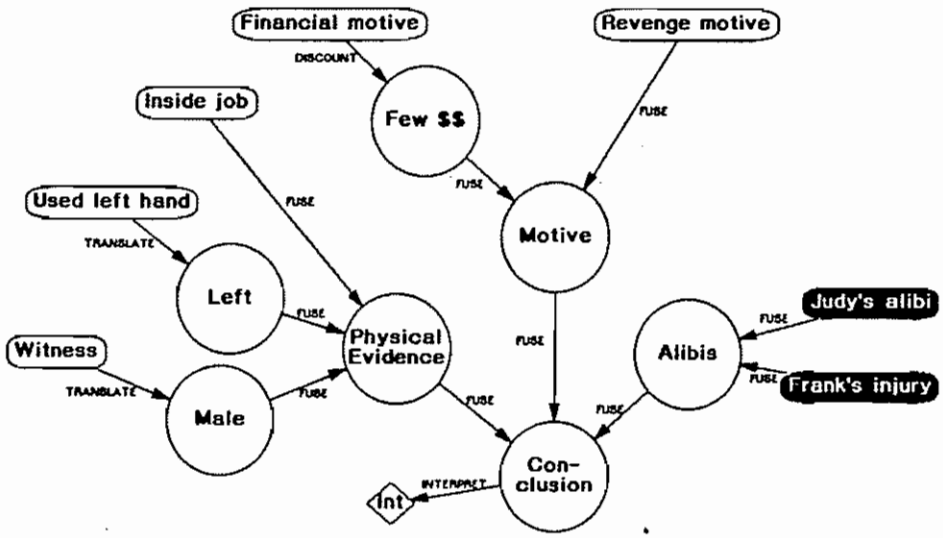

Curator
Analyzer
Grasper II
WINDOW
GRAPH
ANALYSIS
EVIDENCE
CREATE
DESTROY
RENAME
FUSE
PROJECT
TRANSLATE
DISCOUNT
COMPOSE
SUMMARIZE
CONVERT
INTERPRET
EXAMINE
ANCESTORS
DESCENDANTS
EXPLAIN
INSERT
REMOVE
EVALUATE
SWEETSHOP
Financial motive
Revenge motive
DISCOUNT
Few $$
Inside job
FUSE
FUSE
Used left hand
Motive
TRANSLATE
Left
FUSE
Judy's alibi
FUSE
Physical Evidence
FUSE
Alibis
Witness
FUSE
Frank's injury
TRANSLATE
Male
FUSE
FUSE
FUSE
Con-clusion
INTERPRET
Int
[0.6 0.9] (DAVID)
[0.085 0.34] (ANN)
[0.014 0.11] (MIKE)
[0.0 0.041] (JUDY)
[0.0 0.018] (FRANK)
[0.0 0.0] (OTHER)
GRASPER II -- SRI International (C) 1987

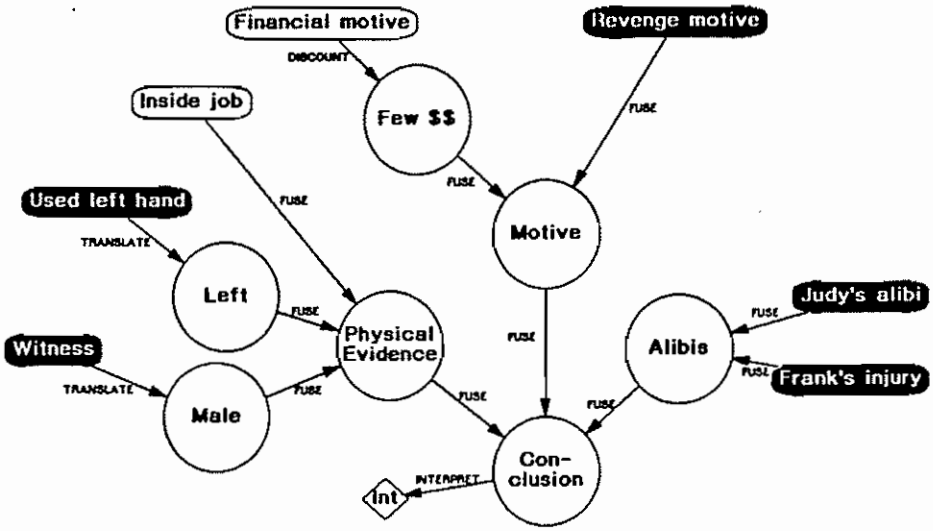
Figure B.3: The sensitivity tools allow the detective to divide the pieces of evidence into useful categories. Each set is constructed by computing the sensitivity of the conclusion for each piece of evidence; those with sensitivities that exceed a threshold are selected.



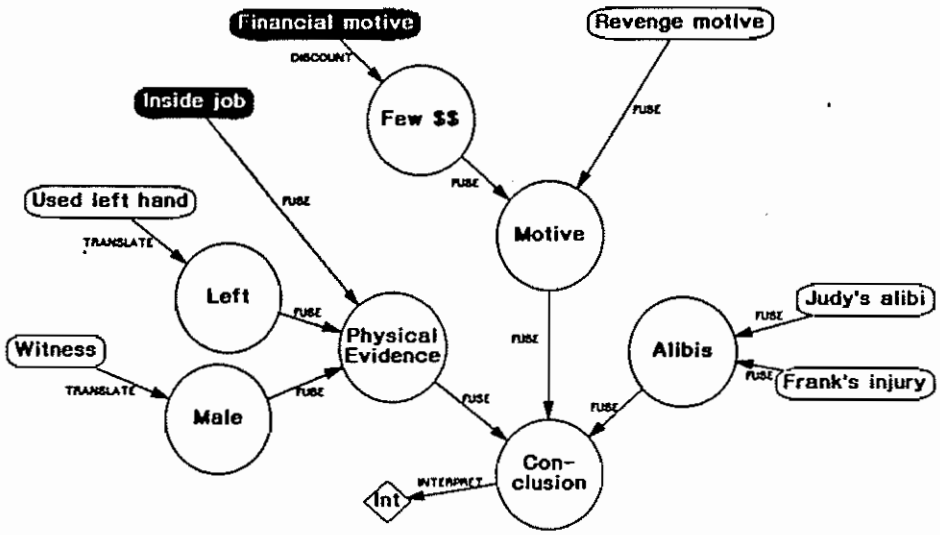
(a) These reports are the most *influential*—they have the largest impact on the specificity and consonance of the conclusion.



(b) Judy's alibi and Frank's injury were the most *inconsequential* pieces of evidence. Alibis are inconsequential unless other evidence incriminates.

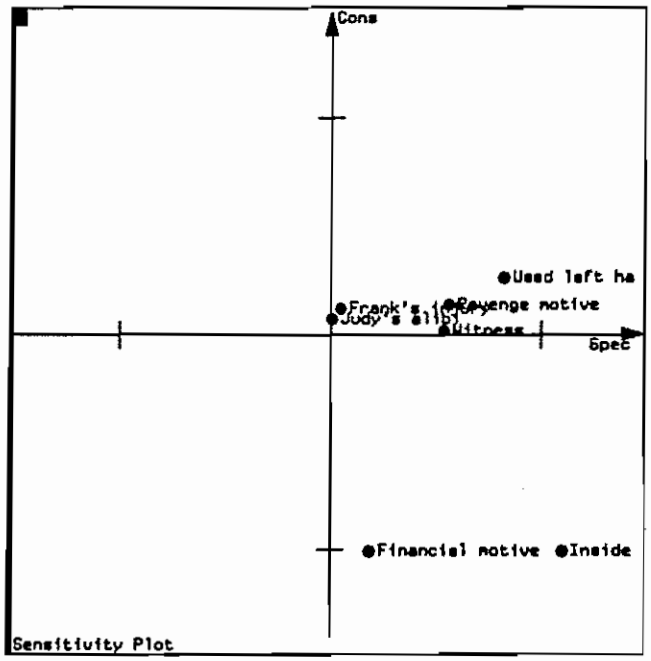


(c) These reports tended to *confirm* the conclusion. That is, they all tended to increase the consonance of the conclusion.



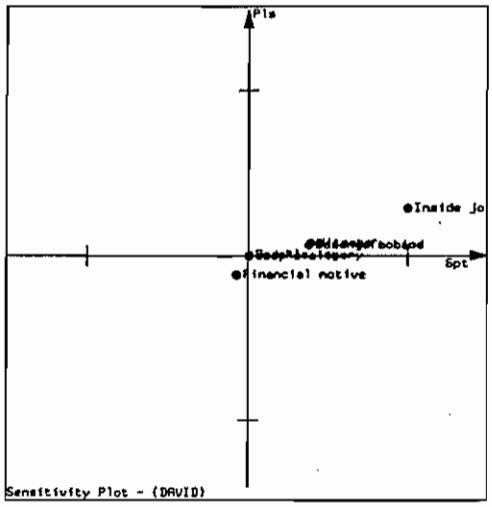
(d) These reports decreased the consonance of the conclusion. Financial motive gave some support to employees other than David; while Inside job ruled out "other," (i.e., non-employees) which was allowed by every other report.

Figure B.4: Scatter plot in Sensitivity Space.

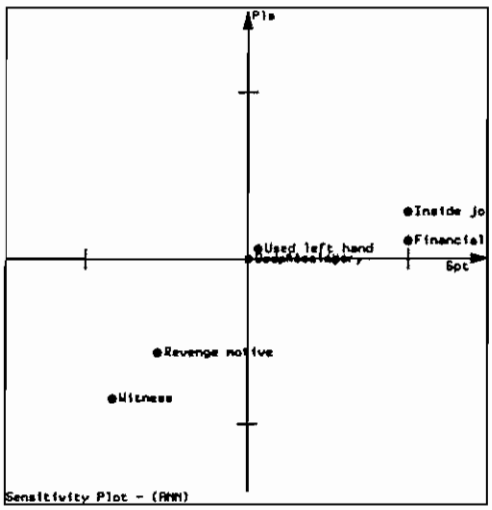


$\widehat{Spec}_i$ and $\widehat{Cons}_i$ of the conclusion are computed for each piece of evidence and plotted here. Those reports near the origin (Frank's injury, Judy's alibi) are inconsequential. Those in the NE quadrant tend to agree with the consensus and also to narrow the set of suspects. Financial motive lends support to suspects other than David, and is therefore dissonant. Inside job also is dissonant because it rules out a possibility ("other") that all the other pieces of evidence admit. See Figure 4.4 for further explanation of each quadrant.

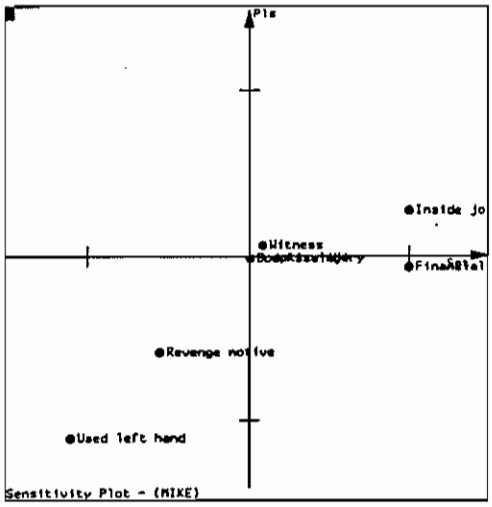
Figure B.5: Sensitivity of Support and Plausibility. The scatter plots show the sensitivity of the evidential interval for each suspect.



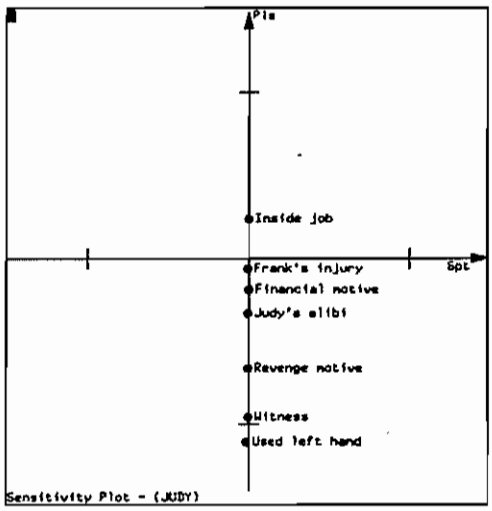
(a) Inside job, Used left hand, Revenge motive, and Witness all tend to incriminate David. Only the financial motive (of Ann) argues against David's guilt.



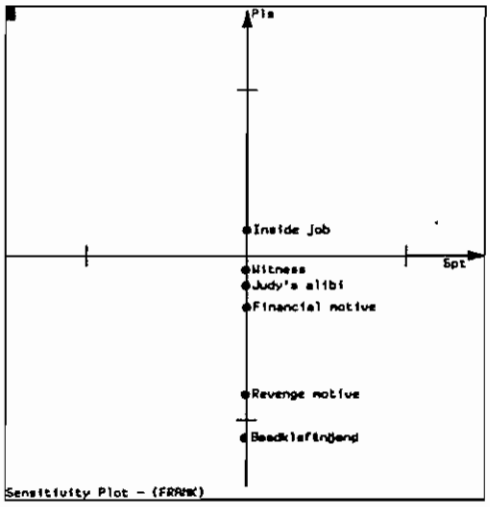
(b) Ann is incriminated by her possible financial motive and by the inside job evidence. The witness report and David's revenge motive tend to absolve her.



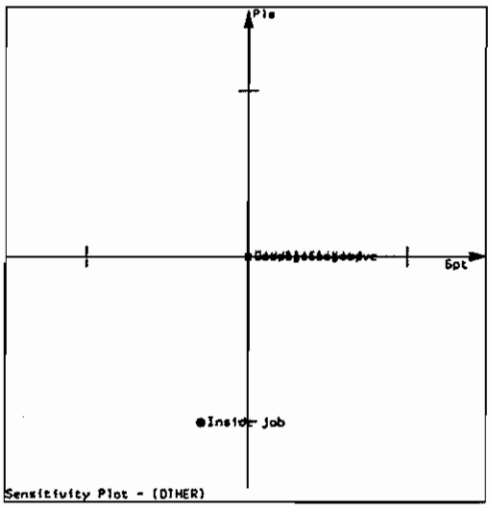
(c) Mike was right-handed, so the fingerprints (i.e, the "Used left hand" evidence) argues against his guilt. The financial motive adds some support to Mike, but also decreases his plausibility slightly. Since Mike had more of a financial motive than David, Frank, or Judy, but less than Ann, it both supports and refutes Mike's guilt.



(d) Nothing in the collection of evidence points directly to Judy, so $Spt(Judy) = 0.0$, even if one piece of evidence were removed. In fact, only it being an Inside job is suggestive of Judy.

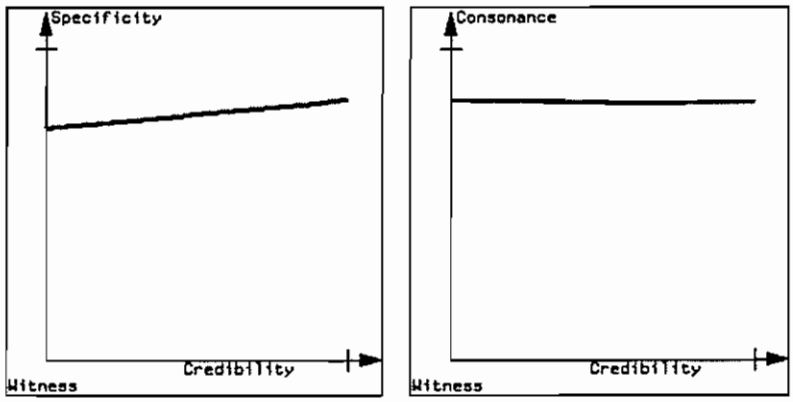


(e) Same reasoning as for Judy.

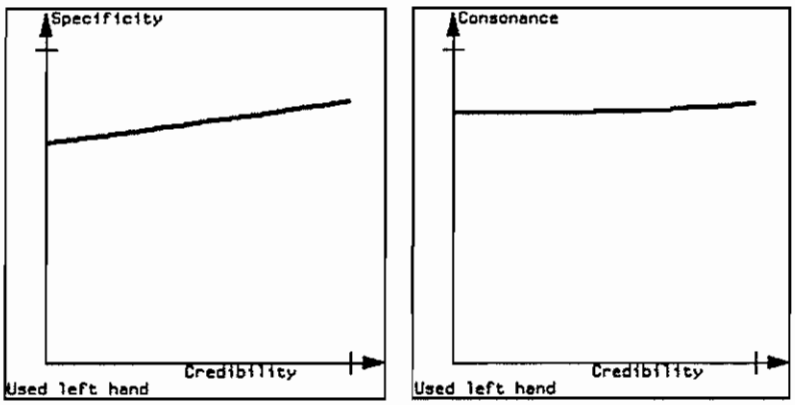


(f) Inside job completely rules out the possiblity of "other." Therefore, it decreases both the support and plausibility of "other." The remaining reports provide no information either for or against any non-employees.

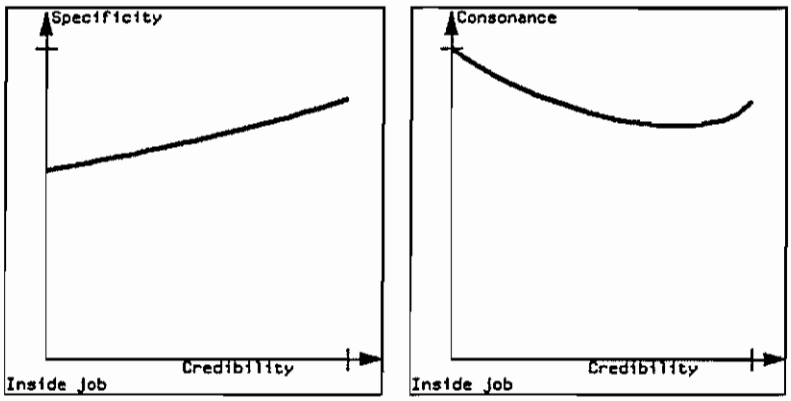
Figure B.6: The following graphs portray how the specificity and consonance of the conclusion vary as the credibility of each piece of evidence is varied.
The origin indicates the measure when the report is completely disregarded; the hash mark near the right end of the abscissa corresponds to the initial assessment made by the detective. Intermediate values reflect various discounting factors. If the slope of the curve is positive, the evidence increases the specificity/consonance of the conclusion. The greater the slope, the more significant is the evidence.



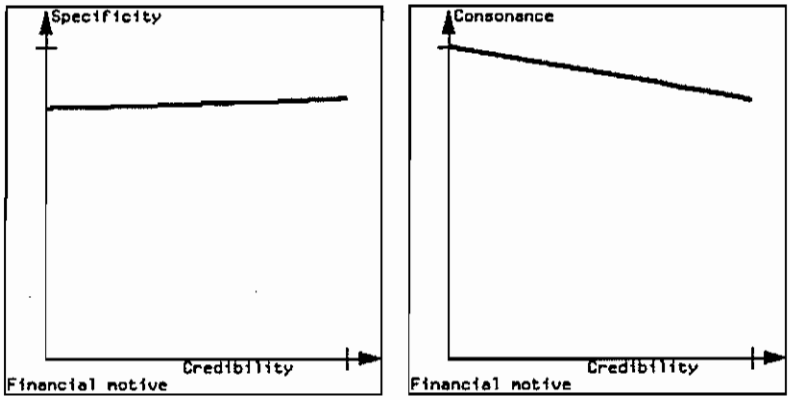
(a) The witness' report makes the conclusion more specific by increasing belief in male suspects.



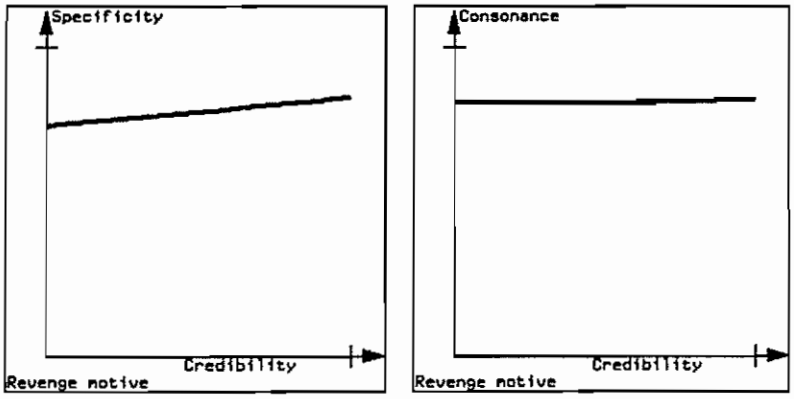
(b) The fingerprints increase both specificity and consonance by focusing on Ann and David.



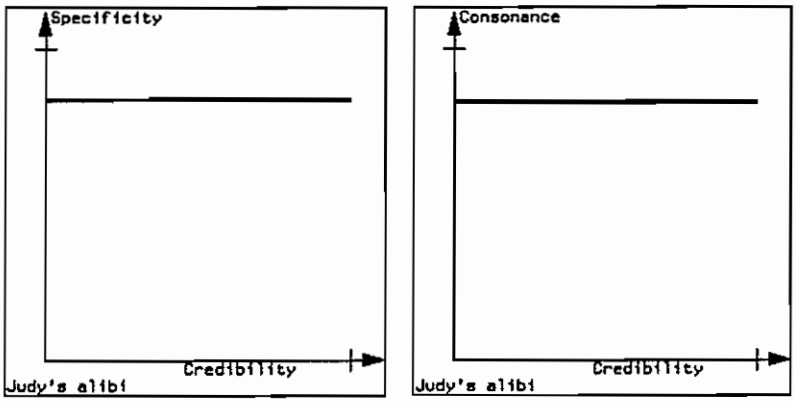
(c) Believing that it may have been an inside job decreases consonance initially because the conclusion is totally consonant without that information (focused on "other"). But when it's credibility nears certainty, the consonance increases as the consensus switches to David.



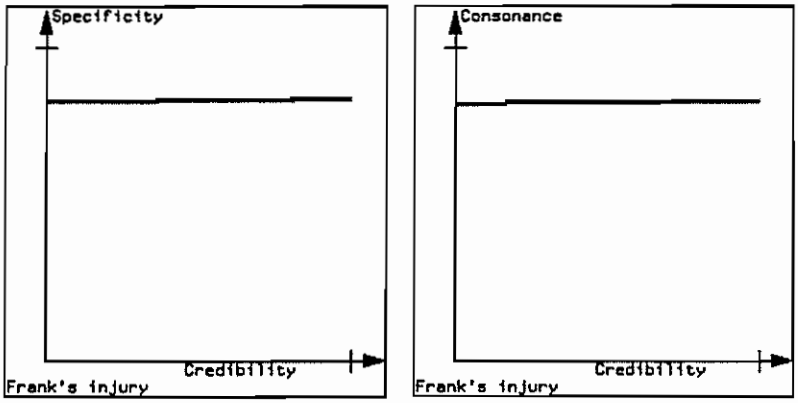
(d) Financial motive is inherently dissonant because it points the finger (to varying degrees) at each employee. Thus, the more we believe it, the more dissonant is our conclusion.



(e) Revenge motive slightly increases specificity by focusing even more on David.

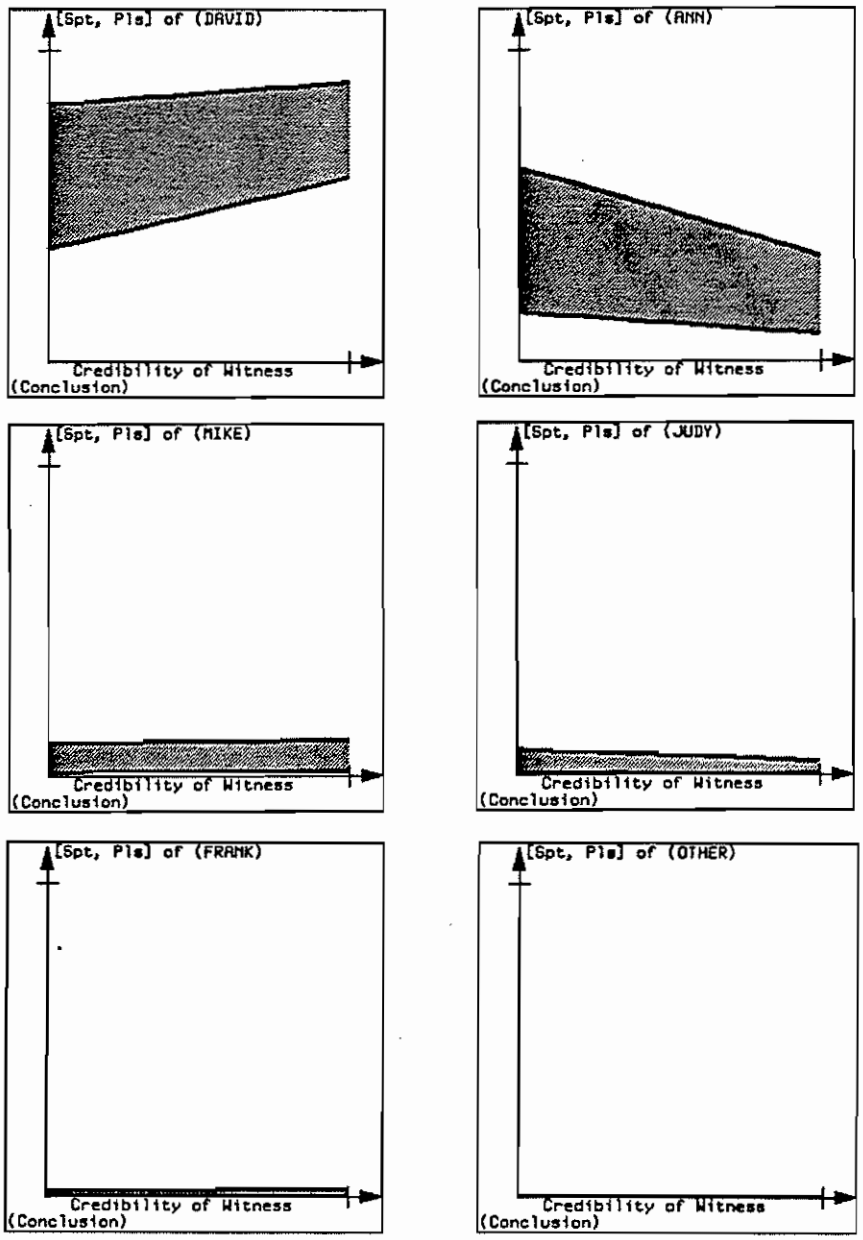


(f) Judy's alibi has no noticeable effect since her guilt is not suggested by the other evidence.

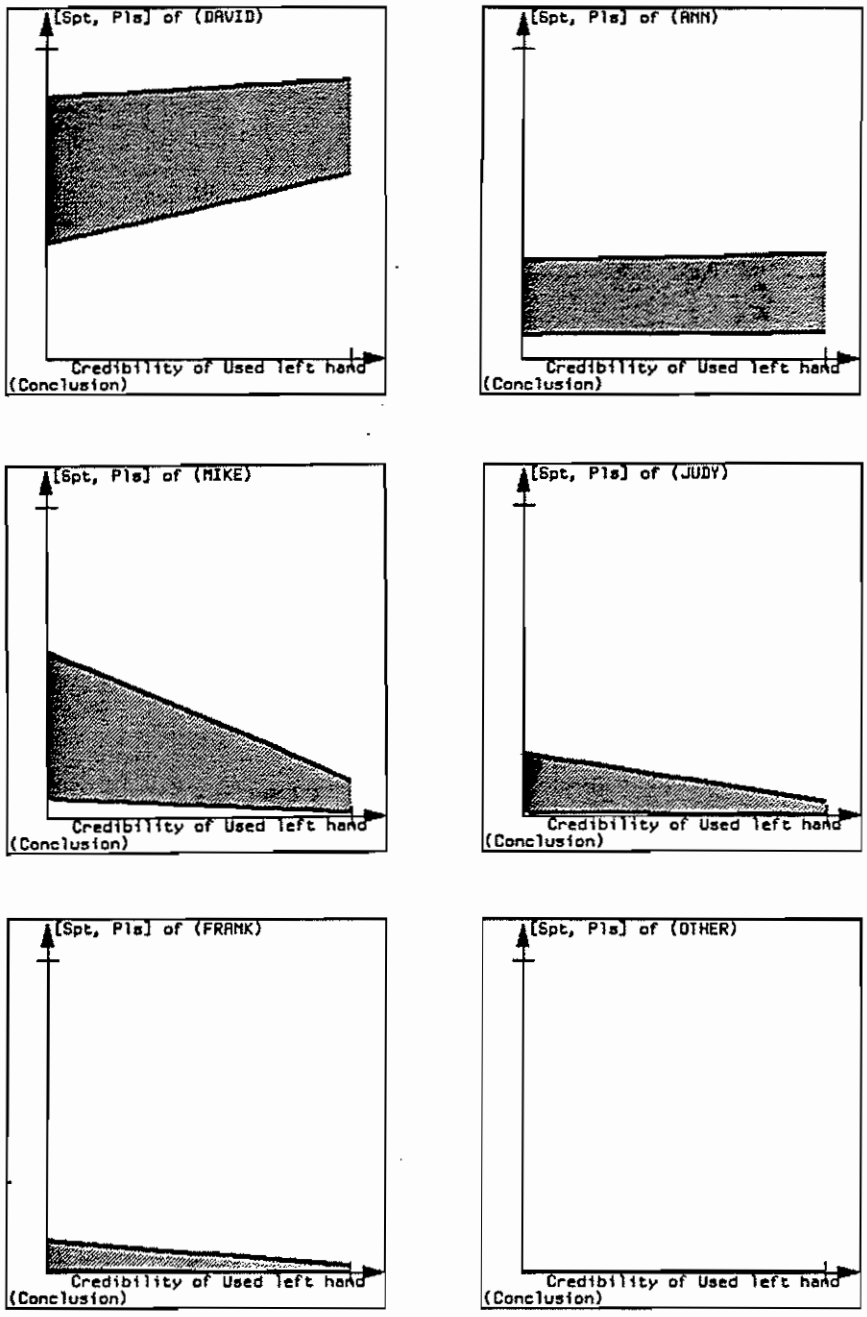


(g) Frank's injury has little effect since the other evidence points only weakly to his possible guilt.

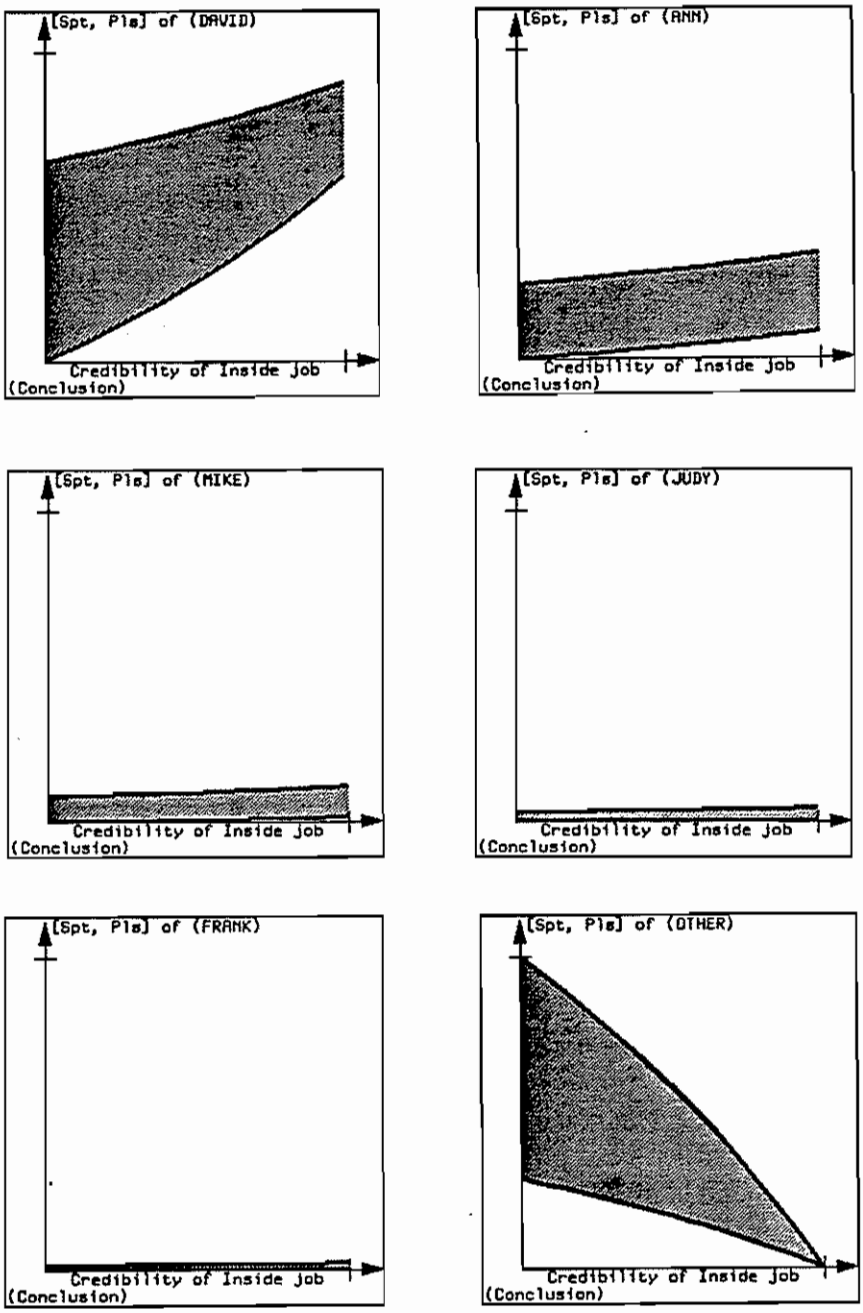
Figure B.7: These graphs show the effect that credibility of a report has on the evidential intervals. The upper boundary of the shaded region shows how the plausibility changes as the credibility of a report is increased. The lower boundary shows the support.



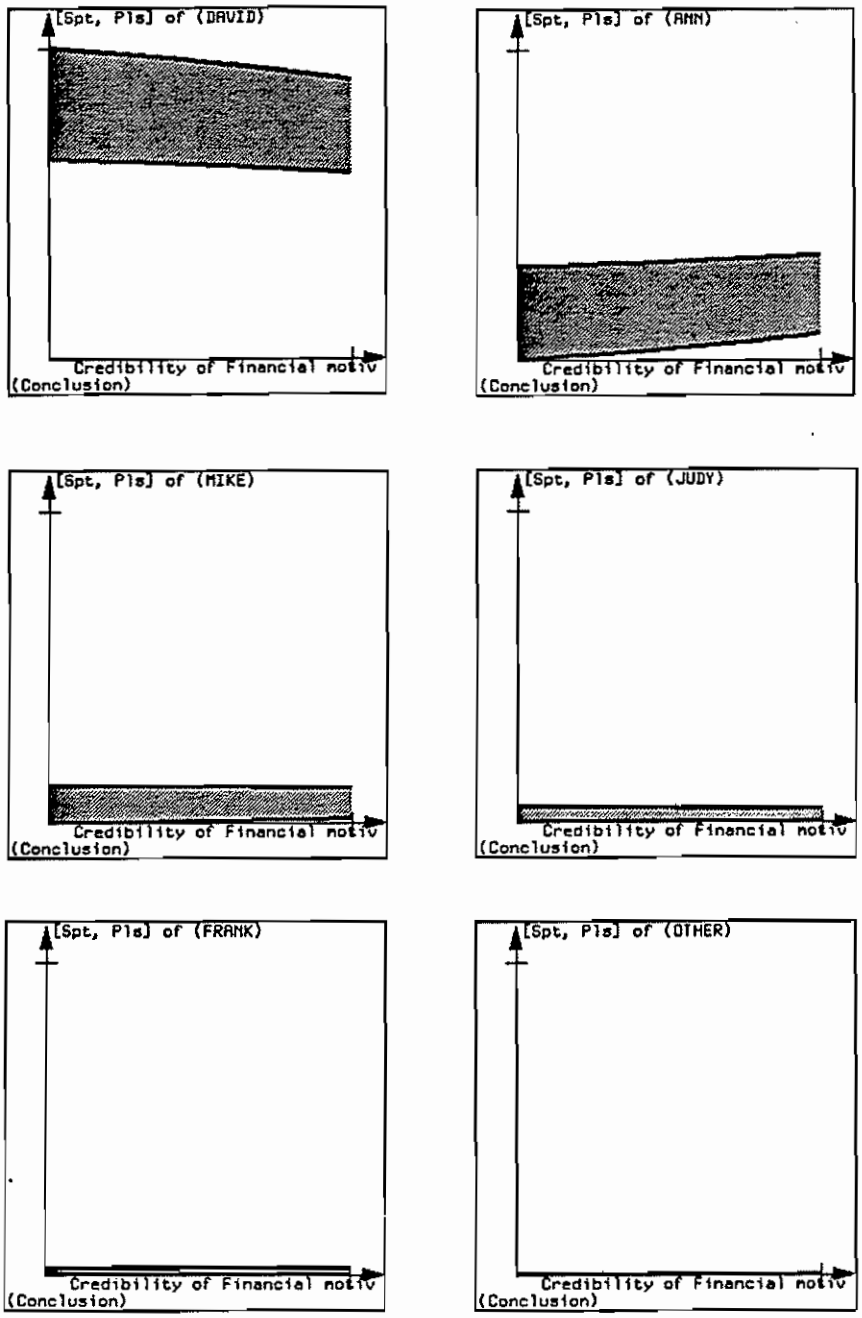
(a) The witness' report raises the evidential interval for the male employees, David, Mike, and Frank, while lowering it for the two female employees. Since "other" is completely eliminated by the Inside Job evidence, the witness' report (like all other bodies of evidence) has no impact on "other."



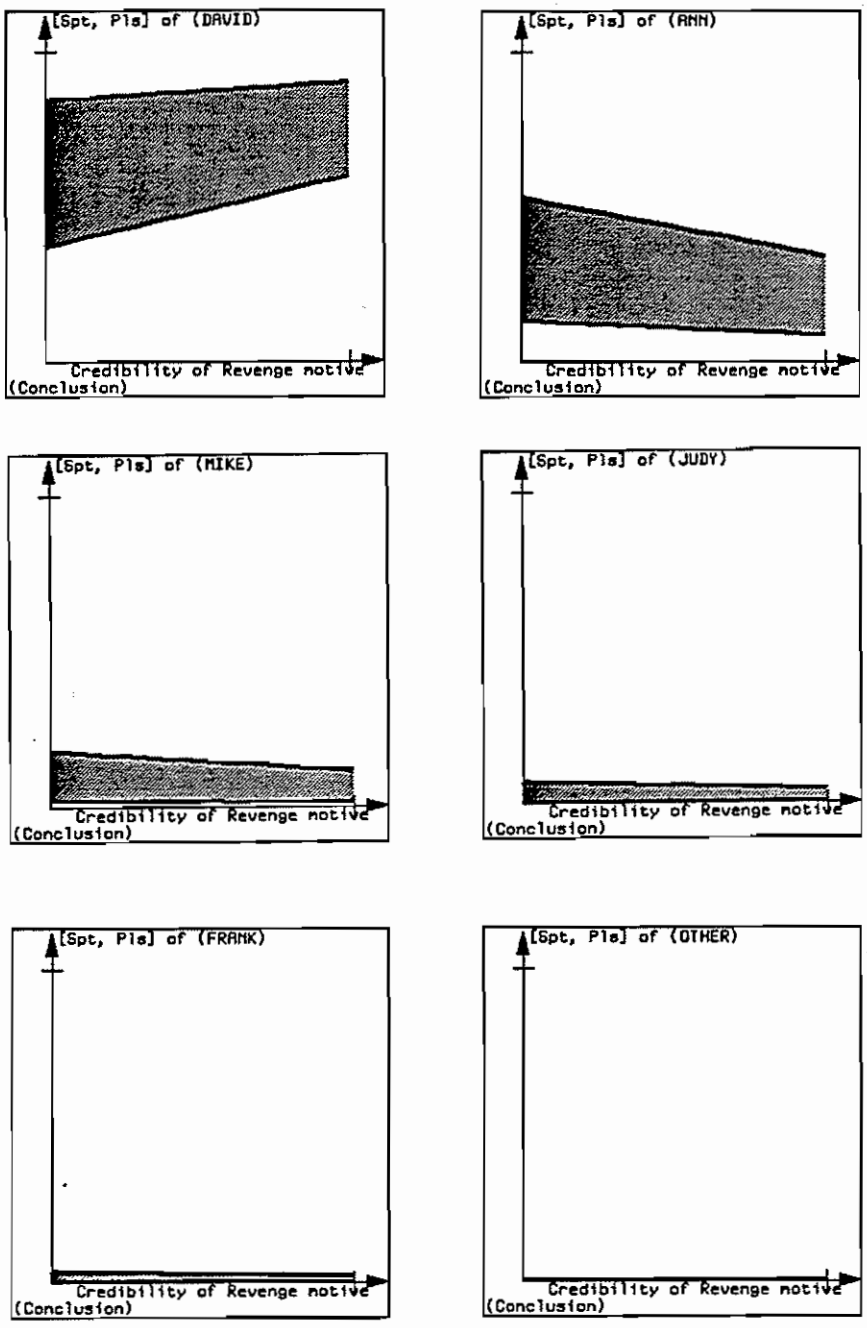
(b) The fingerprints incriminate the left-handed employees, David and Ann, while supporting the innocence of the others.



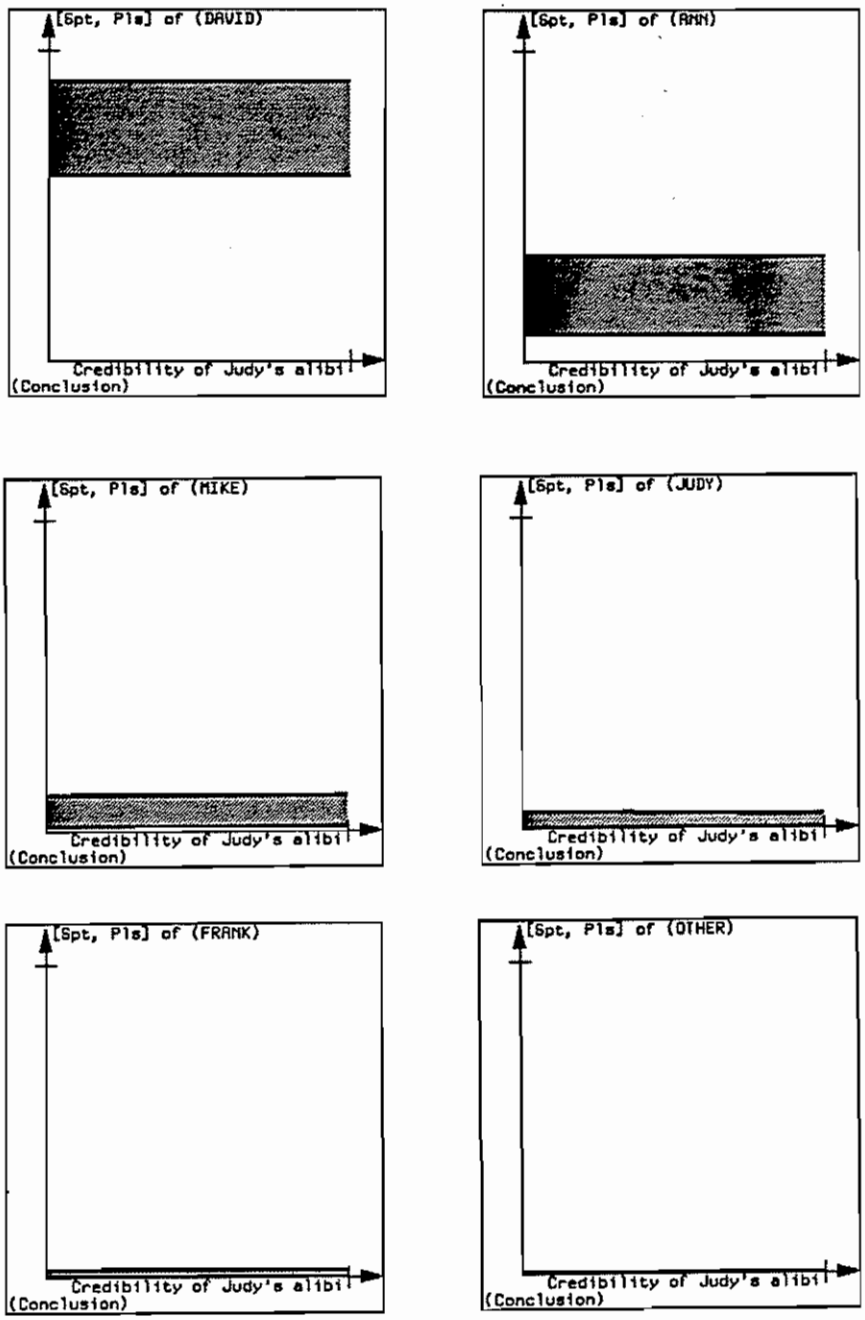
(c) Inside job contributes to the guilt of all the employees. The plausibility of "other," which was 1.0 without this information, is completely eliminated as a possibility with its inclusion.



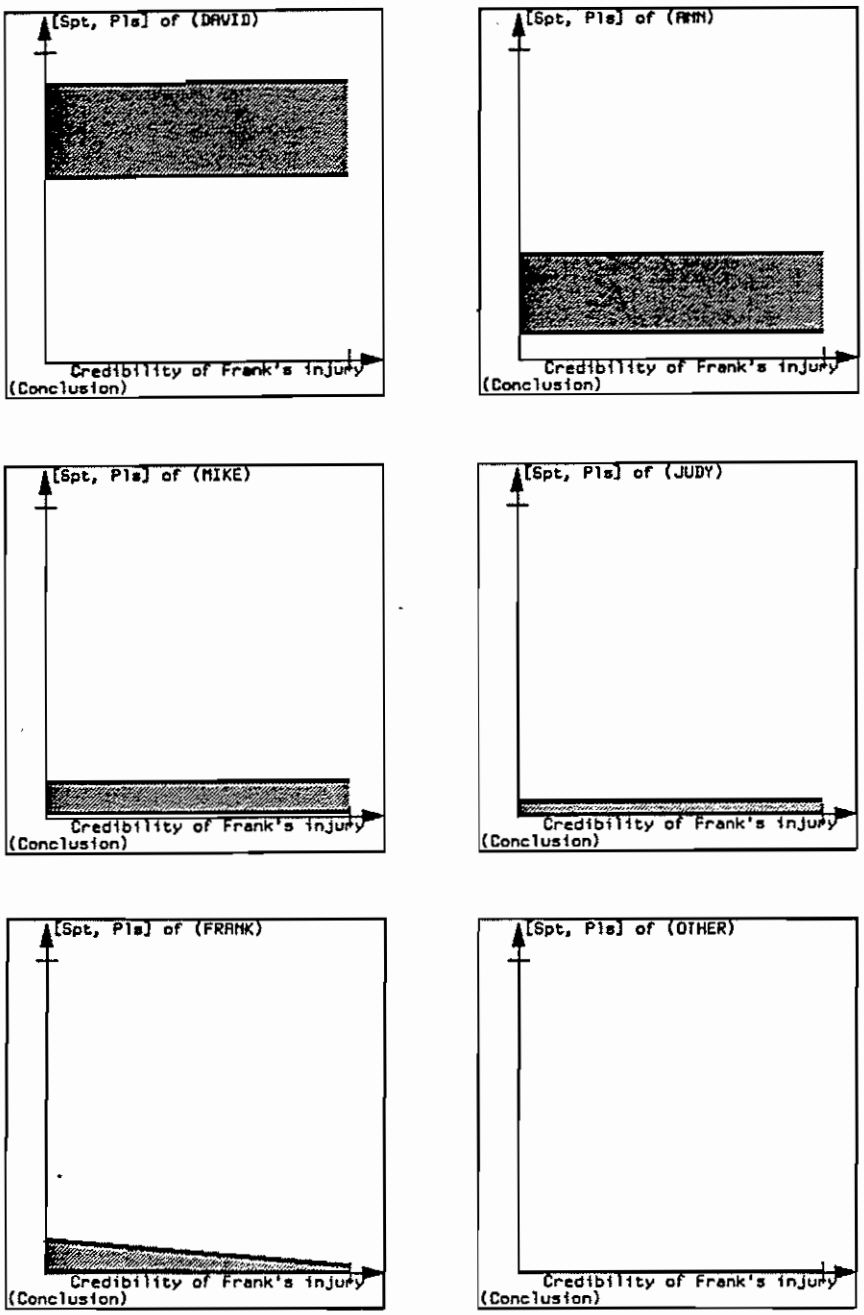
(d) The financial motives tend to absolve David while incriminating Ann. It has little effect on the other employees since they are not suspected by the other evidence.



(e) Revenge motive points toward David and away from all others.



(f) Judy's alibi has no effect on anyone but Judy.



(g) Frank's injury decreases the plausibility of Frank, while remaining noncommittal toward everyone else.