THE MANAGEMENT OF UNCERTAINTY IN INTELLIGENCE DATA: A SELF-RECONCILING EVIDENTIAL DATABASE

Marvin S. Cohen, Kathryn B. Laskey, and Jacob W. Ulvila

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, VA 22043
(703) 790-0510

Submitted to:

Russell Rose
U.S. Government
Washington, DC 20505

June 1987

ABSTRACT

Although increasing attention is being given to the problem of representing uncertainty in database systems, very little work has been done on methods for representing the evidential arguments that underlie assessments of uncertainty. Yet understanding and manipulating such arguments are essential to the intelligence analyst's tasks of making sense out of unreliable and inconsistent data and communicating conclusions.

A critical review of current approaches to modeling uncertainty suggests that the following themes have the most bearing on the design of systems to support intelligence analysis: the adequate representation of alternative types of uncertainty, especially the reliability and completeness of arguments (in contrast to other concepts such as chance and "fuzziness"); the role of assumptions and assumption-revision in reasoning and conflict resolution; and the structure, components, and relevant characteristics of arguments based on evidence.

A system design is proposed, called The Self-Reconciling Evidential Database (SED), which addresses these problems: (a) by providing a generic schema for an evidential argument based on qualitative causal models that link conclusions and evidence; (b) by permitting the analyst to investigate different representations of the same argument through adoption and revision of assumptions; and (c) by embedding evidential arguments within a higher-level "metareasoning" process that responds to conflict between different lines of argument by tracing the assumptions involved in the conflict and recommending revisions.

KEY WORDS

Uncertainty; Certainty; Bayesian Probability; Fuzzy Sets; Data; Database; Logic; Evidence; Inference Theory; Reasoning; Elicitation

TABLE OF CONTENTS

1.0 INTRODUCTION

1.1 The Problem

In recent years, efforts to improve the quality of U.S. intelligence assessments have focused less exclusively than before on the development and deployment of technical data collection methods, such as satellites and sensors. There is a growing awareness that the success of intelligence activities is at least equally dependent on the processes of interpretation and inference that take place *after* the data have been collected, which extract their significance for policy and in many cases prompt and guide the collection of new data. A major goal of the work reported here is to explore ways in which support can be provided for the processes by which intelligence data are managed and analyzed.

A pervasive aspect of those processes is the handling of uncertainty. In a recent report (Cohen, Schum, Freeling, and Chinnis, 1985), we began by arguing that, "in a given analytic problem, identifying sources of uncertainty, assessing the amount of uncertainty from each source, and combining uncertainty across sources to make a final judgment are the most crucial, and perhaps the most difficult, components of the analysis." The appropriate handling of uncertainty plays a central role both in the analyst's own understanding of the problem and in his communication of it to others. If support is to be provided for the analytical task, there is ample reason to target such support here.

1.2 Overview

This report is divided into two relatively self-contained major sections. Section 2.0 is a critical and selective review of current approaches to modeling uncertainty. It focuses on a relatively small set of issues which seem to have the most bearing on the design of systems to support intelligence analysis: i.e., the adequate representation of alternative types of uncertainty, especially the reliability and completeness of arguments (in contrast to other concepts such as chance and "fuzziness"); the role of assumptions and assumption-revision in reasoning and conflict resolution; and the structure, components, and relevant characteristics of arguments based on evidence. Section 2.0 lays the intellectual groundwork for the conceptual design of a system which is described in Section 3.0.

Section 3.0 introduces the Self-Reconciling Evidential Database (SED). In essence, the system has three layers: (1) qualitative causal modeling of the relationships between evidence and conclusions; (2) Shafer/Dempster measures of belief to capture uncertainty about the validity of causally based arguments; and (3) an iterative process of conflict resolution in which assumptions pertaining to the causal model are reexamined, tested, and revised. The components fit together into an interdependent whole. Thus, belief function measures are especially appropriate for assessing the validity of evidential arguments and are themselves clarified by the causal nature of the underlying models. At the same time, belief functions provide natural ways to represent assumptions and to measure conflict, essential ingredients in the conflict resolution process.

Section 4.0 summarizes conclusions and attempts to bring together some of the themes of Sections 2.0 and 3.0. Appendix A describes the algorithms utilized by SED in detail. Appendix B provides a derivation of some relevant results in the theory of belief functions. Finally, Appendix C critically reviews some of the recent work by others on the representation of uncertainty in database systems.

2.0 MODELS OF UNCERTAINTY: THEORETICAL FOUNDATIONS FOR A SELF-RECONCILING EVIDENTIAL DATABASE

In this section we identify and evaluate a number of theories regarding the representation and manipulation of uncertain data. The present discussion focuses on issues and ideas that shaped the design of the Self-Reconciling Evidential Database. It thus provides the historial and intellectual context for ideas that will be laid out more fully in Section 3.0. Most of the principle concepts of the Self-Reconciling Evidential Database, along with their motivation, are here touched on and defended in the context of on-going theoretical debate about uncertainty and reasoning. Additional discussion of theories of uncertainty can be found in Cohen et al. (1986) and Cohen et al. (1985).

The following ideas are most central to our discussion:

(1) focusing on the meaning and reliability of arguments which link evidence and hypothesis;

(2) increasing the precision or certainty of an argument by making explicit assumptions;

(3) resolving conflict among competing arguments by tracing and revising the assumptions that led to the conflict;

(4) providing a natural structure for representing the components of an argument; and

(5) identifying features of each component which form the basis for assumptions and which affect the reliability of the argument.

While most of these topics have attracted attention in isolation, we know of no serious attempts to integrate them within a single framework of reasoning. We will argue that the affinities among the topics are deep; each of them addresses, in a different but complementary way, the problem of *justifying* belief rather than merely *quantifying* it. And they support a process of resolving conflict that delves into the reasons for the conflict, and attempts to rectify them, rather than imposing (as in the Bayesian tradition) a mere statistical aggregation. A system which combines these elements may provide a more powerful and at the same time more transparent approach to the representation of uncertainty in intelligence databases.

A core idea of recent theorizing in this field has been completeness of evidence or reliability of an argument. A secondary theme of this section concerns shortcomings of traditional probabilistic attempts to capture that notion. This is not to deny the utility of Bayesian probability theory: indeed, a Bayesian analysis might well be the basis of an argument whose reliability needs to be considered.

A final theme concerns the representation of imprecise or fuzzy beliefs, and the potential contribution of fuzzy set theory.

2.1 Bayesian Probability Theory

Bayesian probability theory is a strong contender for the central role in any system which represents and manipulates uncertainty. Its supporters date back to work on "probabilistic information processing" (see Edwards, 1966) and earlier; more recent contributors have been de Dombal (1973), in the field of medical decision making, and Schum (1980) in the intelligence field. Our focus in this discussion, however, will be on some shortcomings in Bayesian theory that bear on its use in evaluating an evidential argument.

A simple example can illustrate important features of Bayesian probabilistic reasoning. Consider an uncertain hypothesis, e.g., that "Country X has built a nuclear device." We call this hypothesis H. Under Bayesian theory, uncertainty about H would be represented by assigning a number between 0 and 1. This number, Pr(H), reflects one's degree of belief about H: Pr(H)=.9 means we are fairly sure of H; Pr(H)=.1 means H is unlikely.

The essence of Bayesian inference is the representation of a probability of interest, e.g., Pr(H), in terms of the probabilities of other hypotheses which are easier to assess. Thus, Pr(H) can be computed from these other probabilities rather than directly assessed. There are a variety of ways to express a given probability in terms of other probabilities, but the most familiar in Bayesian theory is "updating" by means of Bayes' Theorem. In the above example, suppose we obtain some evidence about H, e.g., a report from Agent Z that materials from a nuclear reactor in country X have been diverted by the government. Let us call this evidence D. The implications of datum D for the relative likelihood of H and its complement, $\bar{H}$ (not H), are given by application of Bayes' Theorem (here shown in odds-likelihood form):

$$\frac{\Pr[H|D]}{\Pr[\bar{H}|D]} = \frac{\Pr[D|H]}{\Pr[D|\bar{H}]} \cdot \frac{\Pr[H]}{\Pr[\bar{H}]}$$

The first ratio to the right of the "=" is the "likelihood ratio", and reflects the impact of D on the probability of H; the second ratio is the "prior odds" of H, before observing D. For example, if our initial belief is that H and not-H are equally likely ($\Pr(H) = \Pr(\bar{H}) = .5$), then our prior odds equal 1. If we judge that D is 8 times as likely if H is true than if not-H is true, then our likelihood ratio is 8. Multiplying 8 by 1 gives a posterior odds of 8, which implies a posterior probability $\Pr(H|D)$ of $8/(8+1) = .89$.

Bayesian updating can be extended in a number of ways. First, the number of hypotheses can be increased to encompass any mutually exclusive set. For example, we might choose to discriminate further among the cases in which H is not true; e.g., we might consider three possibilities: H_1, that country X has built a nuclear device; H_2, country X *will* have the capability of building a device within 5 years; H_3, country X will not have the capability within 5 years. Second, the impact of additional evidence can be quantified, and the probability of the hypothesis further updated. For example, suppose agent W reports overhearing conversations in which scientists from country X discussed the technical difficulty of building a nuclear device. Thirdly, Bayes' Theorem can be applied in cases where the impact of evidence on the hypotheses of interest is mediated by other hypotheses (Pearl, 1986; Barclay et al., 1977; Peterson et al., 1976; Schum, 1980). For example, our first item of evidence, Agent Z's report of diversion of nuclear material, bears on H indirectly, _via_ the intermediate hypothesis that nuclear material was in fact diverted, i.e., that Agent Z is honest and accurate. Fourth, hierarchical inference structures can be built which are able to accommodate a number of ways that different items of evidence can be related to one another with respect to hypotheses (Schum and Martin, 1980): e.g., they may be contradictory (reporting and denying the same intermediate event), corroboratively redundant (reporting the same intermediate event), cumulatively redundant (reporting different events which reduce one another's evidential impact on the ultimate hypotheses), or non-redundant (reporting different events which enhance or do not change one another's evidential impact). In other, more complex cases of interdependence, Bayesian techniques capture the evidential impact of biases or "noise" in an information source or other element in the network, or non-independence of information source reliability with respect to what is being observed.

Not surprisingly, applications of Bayesian probability theory to real inference problems are typically quite complex. In all but the most trivial cases, a proper Bayesian analysis requires the assessment of a great many conditional probabilities. Moreover, the relations between them are difficult to organize, and the coherence of the total set of assessments is often difficult to determine. Although simplifying assumptions can reduce the assessment burden, the judgments of whether such assumptions are justified may themselves be quite subtle. Typically, model-building is an ~~intera~~iterative process, in which simplifying assumptions are made, conclusions are derived and checked for plausibility, and assumptions revised.

A more serious problem concerns the unnaturalness of some of the assessments that are required in a Bayesian analysis. In the example above, judgments are required regarding the probability of observing a piece of evidence, D, given the hypotheses, H and $\overline{H}$; such judgments must be made in a counterfactual frame of mind, *as if* the datum D had not yet in fact been observed. Further, for most types of evidence, this probability is extremely small; and judgments, even of likelihood *ratios*, are likely to be quite unreliable.

An alternative way of representing Pr(H) in terms of other probabilities, which avoids these problems, involves the "law of total probability":

$$\Pr(H) = \Pr(A)\Pr(H|A) + \Pr(\overline{A})\Pr(H|\overline{A}).$$

Here, H is once again a hypothesis of interest (e.g., "Country X has built a nuclear device"), and A is some conditioning event whose truth or falsity affects the probability of H, e.g., A = "Country X has diverted nuclear material"; $\overline{A}$ = "Country X has not diverted nuclear material." In other words, the probability of H equals the probability of A times the probability of H given that A occurs plus the probability of not-A times the probability of H given that not-A occurs. While in some cases these assessments may be more natural, this form of analysis offers no simple measure, like the likelihood ratio, for assessing the separate impact of each new item of evidence (e.g., the impact of Agent Z's report on Pr(H)).

Axiomatic derivations of Bayesian probability theory from certain desirable properties of beliefs (e.g., deFinetti, 1937/1964) have given the theory a pre-eminent claim to validity among current approaches to handling uncertainty. For example, it can be shown that unless one's beliefs obey the probability axioms, one may be subject to a "Dutch book," i.e., a gamble in which one loses regardless of the outcome of an uncertain state of affairs. Such demonstrations, however, do not conclusively rule out alternative formulations. First, they invariably make technical assumptions that are not compelling (e.g., that belief is measured by a single number rather than an interval). Secondly, such foundational arguments do not establish that Bayesian theory is *uniquely justified*, since other theories may possess desirable properties that Bayesian probability theory lacks. We shall consider one such property, the ability to represent ignorance, later in this section.

The thrust of Bayesian analysis is to improve, rather than replicate, ordinary thinking. Bayesians argue that if one's ordinary intuitions are probabilistically incoherent, they ought to be changed. In other words, the plausibility of the axioms should outweigh the initial plausibility of an incoherent set of judgments. The problem here, though, is two-fold: (1) although the theory demands coherence, it provides no guidance as to *which* judgments of an incoherent set should be changed; and (2) it is not clear why the plausibility of a highly abstract and technical set of axioms should *always* outweigh probabilistic judgment about concrete states of affairs.

It is sometimes implied that a Bayesian analysis is simply a matter of selecting a model, eliciting the required inputs, and calculating the answer. On the contrary, we argue that applying the theory is inevitably an iterative, bootstrapping operation. We use the theory, as if it were true, to perform an analysis, and then test the results for consistency with our direct judgment or with the results of other analyses. If we find inconsistency, we may choose to revise our direct judgment, *or* to revise the values in one or more analyses to *make* them consistent. In other words, we *construct* a probability model for our beliefs rather than elicit or confirm a pre-existing one.

Moreover, we would argue, implausible results might in some cases persuade us to adopt a non-Bayesian approach to an inference problem.

Bayesian theory is designed to capture the concept of chance, or uncertainty about facts. Bayesian theory fails to deal adequately with a quite different type of uncertainty: i.e., the amount and reliability of the knowledge underlying an analysis. In our example above, we began with Pr(H) = Pr(not-H) = .5, reflecting simple ignorance prior to the consideration of evidence. But it might happen that after careful integration of a large quantity of (conflicting) information, the probability of H returns to .5. The single number .5, therefore, does not indicate the degree to which knowledge has been brought to bear on the problem. Similarly, in a Bayesian analysis, a quite high probability might be arrived at based on only a very limited sampling of data (e.g., Pr(H) = .89 after consideration of datum D). A high probability, therefore, means that the *available* evidence favors H over not-H; it tells us nothing about how complete or reliable the available evidence is. Probabilities, therefore, whether high or low, tell us nothing about the quality or goodness of the analysis, in the usual sense of those words.

One result of the failure to represent the amount or type of knowledge underlying a probabilistic assessment is the inability, in Bayesian theory, to provide guidance as to *which* probabilistic beliefs to revise in reconciling alternative, conflicting assessments (e.g., the result of using Bayes' Theorem may differ from an analysis based on the Law of Total Probability, and both may differ from direct judgment). According to Bayesian theory, all such analyses are based on *all* the available evidence; hence, they cannot be evaluated in terms of the amount of knowledge they successfully capture.

A second result of this failure is a theoretical incoherence in the notion of prior probabilities. Prior probabilities are assessed as if the data under consideration had not been obtained, and are then updated by means of the rule (Bayes' Theorem) discussed above. Prior probabilities are often taken to represent complete ignorance about the hypotheses in question. The most common device for handling such ignorance is to assign equal probabilities to all the hypotheses (e.g., $Pr(H) = Pr(\overline{H}) = .5$). Unfortunately, there is always more than one way to assign "equal" probabilities, viz., by rescaling a continuous variable or regrouping discrete hypotheses. Recall in our example that not-H can be broken down into H_2 (Country X will have the capability in 5 years) and H_3 (it will not). Thus we can argue by ignorance that $P(H) = P(H_2) = P(H_3) = 1/3$. Yet our previous argument, also based on ignorance, concluded that $P(H) = .5$. "Informationless priors" are thus not unique, and the choice of a particular representation of the hypothesis set can significantly affect the outcome of a Bayesian analysis.

A superficially persuasive reply is to reject the idea of assigning vacuous "equal" prior probabilities, on the grounds that we are never totally ignorant as long as the hypotheses are meaningful, and that priors, therefore, should always reflect *some* knowledge. There are two problems with this reply. (1) Let us grant that priors for *meaningful* hypotheses must reflect at least some knowledge of the language and of the problem area. In other words, if we are concerned about a hypothesis H and we *understand* H, then our prior probability for H must be implicitly conditioned on a body of knowledge K: i.e., $Pr(H|K)$. It does not follow that K reflects knowledge bearing on the *truth* of the hypothesis. Yet only the latter kind of knowledge eliminates the arbitrariness of assessing prior probabilities. (2) Nevertheless, let us grant for the moment that we always have some knowledge bearing on the truth of the hypothesis. Even so, the conclusion (that *priors* always reflect some of this kind of knowledge) does not follow. An "informationless prior" can be extracted from a prior that embodies knowledge by decomposing that prior into elements that do embody relevant knowledge and elements which do not. In other words, it should sometimes be possible, at least in principle, to divide K into a set K_1 of beliefs that we judge does not bear on the truth of H but which is sufficient to make H meaningful (e.g., much if not all of our knowledge of the English language) and a set of beliefs K_2 that (in conjunction with K_1) we judge *does* bear on the truth of H. Then we can (in principle) treat K_2 as evidence in a Bayesian updating scheme. In that case, we get $Pr(H|K_1,K_2) \propto Pr(H|K_1)Pr(K_2|H,K_1)$. $Pr(H|K_1)$ must be regarded as an "informationless prior" since K_1 is sufficient to render H meaningful, but has no impact on our beliefs about the *truth* of H. Assessment of $Pr(H|K_1)$ thus cannot escape the difficulties discussed above regarding arbitrariness and non-uniqueness.

The upshot of this argument is *not* that "informationless priors" must ever be assessed *in practice*. We can, after all, simply decline to decompose $Pr(H|K)$

the way just described. The argument points, rather, to a more fundamental theoretical difficulty: that Bayesian theory implies the intelligibility of certain judgements which are in fact arbitrary and non-unique. This finding weakens the normative basis of the theory.

An interesting recent defense of Bayesian priors is provided by Cheeseman (1985), who argues that while an understanding of the problem may not yield information bearing on the *truth* of hypotheses, it typically yields invariance requirements on the answer, and these imply a definite partition (i.e., grouping or scaling) of the hypotheses. In the absence of further information, equal priors should then be assessed for that partition. Cheeseman appears to recommend an attitude toward prior probabilities that regards them not as beliefs, but as *assumptions*, subject to revision when new information (implying new invariance requirements) dictates a different partitioning. For example, if we are asked to give the probability of finding a ship within a particular square mile in the Atlantic, invariance (with respect to translations, rotations, transformations of scale?) requires assigning equal probability to equal areas. "Since the Atlantic is roughly diamond shaped, this means that the probability of finding the ship at an equatorial latitude is higher than at a polar latitude." Imagine that we now acquire information that the ship has been instructed to move to a particular latitude, but "interference scrambles our reception of *which* latitude. Then after the ship has had time to move, our knowledge is represented by assigning uniform probability to each latitude." Note that neither the original problem formulation nor the subsequent information (about instructions to the ship) tells us anything that bears directly on the truth of where the ship is.

While this approach is ingenious, it raises some questions:

- The justification for equal priors ("maximum entropy") is that they provide a "neutral background against which any systematic (non-random) patterns can be observed." But this formulation depends on a metaphor of physical equilibrium which has dubious applicability to belief revision in general. Thus, assignment of equal probabilities remains unjustified.

- Separate items of evidence bearing on the actual location of the ship may be obtained, some of which are couched in terms of area and some of which are couched in terms of latitude. No single partition can serve as a "neutral background" for both types of evidence.

- Are invariance requirements always clear? Do they always uniquely determine a partition? If the answer to these questions is yes (which we doubt), then it seems wrong to suppose that new information (e.g., about the ship's instructions) will always *supplant* the requirements implied by the original problem formulation. It appears, rather, that in some cases two or more sets of requirements will be imposed, which cannot simultaneously be satisfied.

- If the answer to the two questions above is no, then in some cases at least, partitions are not uniquely determined and assessment of prior probabilities is arbitrary and non-unique. The *theoretical* difficulty for Bayesian Theory remains as long as there is only one such example.

Bayesian theory provides a clear behavioral interpretation of probabilities in terms of preferences among bets. Such behavioral implications are often missing in theories that incorporate more satisfactory representations of ignorance; i.e., in those theories the available evidence may be insufficient to discriminate among the options. We return to this topic at the end of Section 2.2.

In summary, Bayesian theory possesses a uniquely compelling logical foundation in application to the concept of chance and a strong link to decision making. However, it provides no guidance in terms of reliability of knowledge for choosing among alternative coherent analyses, is not well suited for dealing with incompleteness of evidence, and seems to imply the intelligibility of judgments about prior probabilities which in fact appear to be arbitrary and non-unique.

2.2 Belief Functions

In the theory of belief functions introduced by Shafer (1976), Bayesian probabilities are replaced by a concept of evidential support. The contrast, according to Shafer (1981; Shafer and Tversky, 1983) is between the chance that a hypothesis is true, on the one hand, and the chance that the evidence *means* (or proves) that the hypothesis is true, on the other. Thus, we shift focus from truth of a hypothesis (in Bayesian Theory) to the evaluation of an evidential argument (in Shafer's Theory). By stressing the link between evidence and hypothesis, Shafer's system (a) is able to provide an explicit measure of quality of evidence or ignorance (i.e., the chance that the evidence is not linked to the hypothesis by a valid argument); (b) is less prone to require a degree of definiteness in inputs that exceeds the knowledge of the expert, and (c) permits segmentation of reasoning into analyses that depend on independent bodies of evidence. We will find that each of these properties can contribute significantly to the representation of uncertainty in intelligence analysis.

In Shafer's system, the support for a hypothesis and for its complement need not add to unity. For example, if a witness with poor eyesight reports the presence of an enemy antiaircraft installation at a specific location, there is a certain probability that his eyesight was adequate on the relevant occasion and a certain probability that it was not, hence, that the evidence is *irrelevant*. In the first case, the evidence proves the artillery is there. In neither case could the evidence prove the artillery is not there.

To the extent that the sum of support for a hypothesis and its complement falls short of unity, there is "uncommitted" support, i.e., the argument based on the present evidence is unreliable. Evidential support for a hypothesis is a lower bound on the probability of its being true, since the hypothesis *could* be true even though our evidence fails to demonstrate it. The upper bound is given by supposing that all present evidence that is *consistent* with the truth of the hypothesis were in fact to prove it. The interval between lower and upper bounds, i.e., the range of permissible belief, thus reflects the unreliability of current arguments. This concept is closely related to *completeness* of evidence, since the more unreliable a given argument is, the more changeable the resulting beliefs are as new evidence (with associated arguments) is discovered. These concepts are not directly captured by Bayesian probabilities.

In Shafer's calculus, support m(•) is allocated not to hypotheses, but to *sets* of hypotheses. As with probability, however, the total support across these subsets will sum to 1, and each support m(·) will be between 0 and 1. It is natural, then, to say that m(·) gives the probability that what the evidence *means* is that the truth lies somewhere in the indicated subset.

Suppose, for example, that we have three hypotheses of interest: H_1 (Country X has built nuclear device), H_2 (Country X will have the capability within 5 years), and H_3 (Country X will not have the capability in 5 years). If we are ignorant regarding these hypotheses, we simply assign full support to the universal set, i.e., $m(\{H_1, H_2, H_3\}) = 1$; in other words, our available evidence tells us only that *something* is true, but not what. In a Bayesian analysis, arbitrary decisions would have to be made about allocating probability *within* this set, requiring judgments that are unsupported by the evidence.

On the other hand, the evidence may be partly but not wholly imprecise. For example, Agent W's report of a pessimistic technical discussion among scientists in Country X may mean that building a nuclear device is at least five years away (H_3); alternatively, it might only mean that Country X has not *yet* built a device (i.e., $\{H_2$ *or* $H_3\}$); finally (if Agent W is unreliable or if these scientists are not in the know), this evidence could mean nothing at all (i.e., $\{H$, or H_2 or $H_3\}$). We can now assess the chance of each possibility: e.g., $m(H_3) = .3$; $m(\{H_2$ or $H_3\}) = .2$; $m(\{H_1$ or H_2 or $H_3\}) = .5$.

This same device, of allocating support to subsets of hypotheses, enables us to represent the reliability of probability assessments. Suppose, for example, that a certain type of seismograph reading has been associated with natural seismic activity 70% of the time and with nuclear tests 30% of the time based on past frequency data. If we are confident that seismic data now being analyzed are representative of this set, we may have m(earthquake) = .7 and m(nuclear test) = .3. But if there is reason to doubt the relevance of the frequency data to the present problem (e.g., because the frequency data come from U.S. tests and the seismic data from other geological regions), we may *discount* this support function by allocating same percentage of support to the universal set. For example, with a discount rate of 30%, we get m(earthquake) = .49, m(nuclear test) = .21, and m({earthquake, nuclear test}) = .30. The latter reflects the chance that the frequency data are irrelevant.

Shafer's belief function Bel(•) summarizes the implications of the m(·) for a given subset of hypotheses. Bel(A) is defined as the total support for all subsets of hypotheses contained within A; in other words, Bel(A) is the probability that the evidence *implies* that the truth is in A. The plausibility function Pl(·) is the total support for all subsets which overlap with a given subset. Thus, Pl(A) equals 1-Bel($\overline{A}$); i.e., the probability that the evidence does not imply the truth to be in not-A. In the example above (about Agent W), we get:

$$\mathrm{Bel}(H_3) = m(H_3) = .3;$$

$$\mathrm{Pl}(H_3) = 1\text{-}\mathrm{Bel}(\{H_1 \text{ or } H_2\}) = 1$$

$$\mathrm{Bel}(\{H_2 \text{ or } H_3\}) = m(H_3) + m(\{H_2 \text{ or } H_3\}) = .5;$$

$$\mathrm{Pl}(\{H_2 \text{ or } H_3\}) = 1\text{-}\mathrm{Bel}(\{H_1\}) = 1.$$

Thus far, we have focused on the representation of uncertainty in Shafer's system. For it to be a useful calculus, we need a procedure for inferring degrees of belief in hypotheses in the light of more than one piece of evidence. This is accomplished in Shafer's theory by Dempster's rule. The essential intuition is simply that the "meaning" of the combination of two pieces of evidence is the intersection, or common element, of the two subsets constituting their separate meanings. For example, if evidence E_1 proves $\{H_1$ or $H_2\}$, and evidence E_2 proves $\{H_2$ or $H_3\}$, then the combination $E_1 + E_2$ proves H_2. Since the the meanings of the two pieces of evidence are assumed to be independent, the probability of any given combination of meanings is the product of their separate probabilities.

Let X be a set of hypotheses H_1, H_2,...,H_n, and write 2^X for the power set of X, that is, the set of all subsets of X. Thus, a member of 2^X will be a subset of hypotheses, such as $\{H_2, H_5, H_7\}$, H_3, or $\{H_1, H_2, H_3, H_4\}$, etc. Let A refer to any subset in 2^X. Then if $m_1(A)$ is the support given to A by one piece of evidence, and $m_2(A)$ is the support given by a second piece of evidence, Dempster's rule is that the support that should be given to A by the two pieces of evidence is:

$$m_{12}(A) = \frac{\sum\limits_{A_1 \cap A_2 = A} m_1(A_1)m_2(A_2)}{1 - \sum\limits_{A_1 \cap A_2 = \emptyset} m_1(A_1)m_2(A_2)}.$$

The numerator here is the sum of the products of support for all pairs of subsets A_1, A_2 whose intersection is precisely A. The denominator is a normalizing factor which ensures that $m_{12}(\cdot)$ sums to 1, by eliminating support for impossible combinations.

The following table illustrates the application of Dempster's Rule for combining two sources of evidence: Agent Z's report of diversion of material, $m_1(\cdot)$, and Agent W's report of a pessimistic technical discussion, $m_2(\cdot)$.

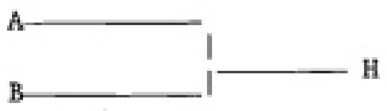
	$m_1(\cdot)$	$m_2(\cdot)$	$m'_{12}(\cdot)$	$m_{12}(\cdot)$
H_1	.9	0	(.9)(.5) = .45	.82
H_2	0	0	0	0
H_3	0	.3	(.1)(.3) = .03	.05
H_1H_2	0	0	0	0
H_1H_3	0	0	0	0
H_2H_3	0	.2	(.1)(.2) = .02	.04
$H_1H_2H_3$	.1	.5	(.1)(.5) = .05	.09
Null set			(.9)(.3) + (.9)(.2) = .45	

where $m'_{12}(\cdot)$ represents combined belief prior to normalization.

In our discussion of Bayesian theory (Section 2.1), we noted that a variety of probabilistic structures were available (a) to represent the propagation of uncertainty in a chain of events from data to hypotheses, and (b) to capture various types of non-independence among items of evidence and hypotheses. By contrast, the application of Dempster's Rule to belief functions appears (a) to apply only to the combination of evidence bearing on the same hypotheses, and (b) to demand that the evidence be independent. A more complete understanding of the Shafer/Dempster framework, however, reveals that both of these appearances are misleading.

Dempster's Rule can be used not only to combine evidence, but also to propagate uncertainty across chains of events (Laskey and Cohen, 1986; Shafer, 1982). Suppose, for example, that A is a conditioning event whose truth or falsity affects the probability of a hypothesis H; e.g., H = "Country X has built a nuclear device"; Agent Z has reported diversion of nuclear material by Country X; A = "Nuclear material has been diverted"; $\overline{A}$ = "Nuclear material has not been diverted" (cf., Section 2.1 above). To see how Dempster's Rule could apply to propagation, let us suppose we have a belief function over $\{A, \overline{A}\}$, and two conditional belief functions over $\{H, \overline{H}\}$, one conditional on A and the other conditional on $\overline{A}$. We are interested in the implications of these beliefs for Country X's building a nuclear device: what is the implied *unconditional* belief function over $\{H, \overline{H}\}$? We begin by extending each of these three belief functions to the space $\{A, \overline{A}\} \times \{H, \overline{H}\}$. For example, support for A tells us nothing about H versus $\overline{H}$; so we translate $m_1(A)$ as $m_1(A \times \{H, \overline{H}\})$. Support for H conditional on A, $m_2(H|A)$, indicates a link between H and A, but says nothing about H versus $\overline{H}$ if $\overline{A}$ is true or about A versus $\overline{A}$; so we represent it as $m_2(\{H \times A\} \cup (\{H, \overline{H}\} \times \overline{A}))$. Assuming independence of the evidence used to assess them, the three extended belief functions are then combined by Dempster's Rule. Finally, the combined belief function over the product space gives rise to a marginal belief function over $\{H, \overline{H}\}$; e.g., $m(H)$ is obtained simply by summing the support for subsets contained in H. Details on the application of Dempster's Rule for propagation can be found in Appendix B.

These same procedures can be applied to chains of events of any length and may also be used to represent interdependencies among evidence and hypotheses of any degree of complexity. For example, suppose we have an additional item of evidence bearing on H: Agent W has reported discussions of nuclear devices among senior scientists in Country X. Let B = "Agent W is honest"; $\overline{B}$ = "Agent W is not honest." The truth of H now depends on both A and B:



Assume independence of the evidence for belief functions on {A, $\overline{A}$}, on {B, $\overline{B}$}, and on {H, $\overline{H}$} conditional on A, $\overline{A}$, B and $\overline{B}$. Then we can derive two marginal belief functions on {H, $\overline{H}$} by utilizing the method in the previous paragraph for {A, $\overline{A}$} and {B, $\overline{B}$} separately; we then combine the results by Dempster's Rule. But what if our assessment of A (which involves Z's honesty) is *not* independent of our assessment of B, W's honesty, e.g., Agents Z and W are business associates? In that case we must assess belief in the product space {A, $\overline{A}$} x {B, $\overline{B}$}, then assess belief in {H, $\overline{H}$} conditioned on the four combinations of {A, $\overline{A}$} and {B, $\overline{B}$}, before applying the method outlined in the previous paragraph.

In an important sense, belief functions can be shown to handle all the interdependencies among events and evidence to which Bayesian Theory has been applied (e.g., Schum, 1980). Dempster's Rule requires independence of the evidence *underlying* the belief functions that are combined; it does not require independence of the events or claims represented explicitly *within* the belief functions. Dependent items of evidence can be dealt with, then, by making questions about the reliability of evidence explicit, to be dealt with as conditioning events or intermediate hypotheses.

One of the main difficulties standing in the way of a Bayesian analysis is its complexity. At first sight Shafer's approach seems simpler, since complicated independence judgments and conditional probability assessments appear not to be required. This appearance is illusory. First, as just shown, conditional non-independence can be represented in a belief function analysis. Secondly, support functions must be assessed over not just the hypothesis set, but over the power set of the hypothesis set. With 10 hypotheses, for example, the support distribution has 1,023 elements. In principle, then, the number of assessments required by a Shaferian model is likely to exceed the number required by a comparable Bayesian model.

A Shaferian response to this, in parallel with the Bayesian response, is that specialized models may be developed that require far fewer assessments (such as consonant belief functions, in which all support goes to a nested series of hypotheses). Here again, however, (as in the Bayesian case) rather subtle judgments must be made to determine that a particular specialized model is applicable before savings in quantity of assessments can be realized. Moreover, combining via Dempster's Rule does not preserve consonance; similarly, complex hierarchical models may become quite baroque.

Shafer's theory offers a mild advantage over Bayesian theory, however, in the ease of organizing judgments into modular *arguments*. Belief functions are assessed separately for independent bodies of evidence, thus providing a measure of belief in the hypotheses based on each separate argument. In Bayesian theory, by contrast, there is no *modular* way to represent degree of belief based on different subsets of the evidence. Likelihood ratios for independent evidence represent the impact of an argument on belief, but do not represent degree of belief itself until prior probabilities are factored in. As a result, the same priors must be included in each argument bearing on the same hypotheses, thus compromising the modularity of the representation.

Theories of belief like Shafer's, which provide intervals rather than single measures of support, may fail to uniquely determine a best decision. For each option, an upper and a lower expected utility may be computed. If the lower utility of option A is higher than the upper utility of option B, then A is preferred; but if the intervals overlap, the evidence fails to discriminate between the two alternatives.

The occasional failure of belief functions to yield non-ambiguous prescriptions for action is, from the Bayesian point of view, a drawback. Within a framework that acknowledges the concept of ignorance, however, it may be regarded as a virtue. If the available evidence (plus preferences) does not uniquely determine a decision, we want to *know* this, rather than sweep it under the rug. Under circumstances of ignorance, the traditional Bayesian system requires the analyst to make arbitrary judgments to fill in the gaps. Such judgments are then treated exactly as if they were determined by the evidence. The belief function framework, by contrast, enables us to see how far the evidence itself takes us toward a decision, and how much of the work must be done by some other, non-evidential process. In fact, two such processes are available which, in conjunction with the evidence, will uniquely determine an action in any belief function analysis: (1) Adoption of a "decision attitude" for selecting a single point within the utility interval, e.g., best case, worst case, or a weighted average: (2) Narrowing the intervals by introducing additional substantive assumptions. Shafer himself has investigated a variant of (1) (Shafer, 1976); the system to be described in Section 3.0 provides a mechanism for (2).

It would be appealing to regard Shafer's theory as simply a special-case Bayesian model, in which probabilities are assessed over the meaning of the evidence rather than over the truth of hypotheses. In that case, inference by Dempster's Rule would share the axiomatic derivation from formally desirable properties of belief that is a major feature of Bayesian theory. Such a derivation, is, however, unsuccessful.

The basic support assignment $m_1(H)$ represents the "chance" that evidence collection 1 means that H is true. This chance cannot be interpreted as a Bayesian probability, however, because it is to be assessed as if the *content* of evidence collection 1 were unknown. In other words, there are two phases (at least in a conceptual sense) in the assessment process (see Shafer, 1985): First, we assess the *general* reliability of the source of this evidence, e.g., the accuracy and honesty of a witness or sensor, the *general* validity of the experimental method, type of argument, or analysis being employed. Only then, secondarily, do we take into account the content of the evidence, e.g., what the source said on *this* occasion, the result of the experiment, the conclusion of *this* argument or analysis. In the first phase, one is considering the capabilities of the relevant evidential process in general to establish a link between evidence (whatever it turns out to be) and hypotheses. In our example, the chance that a report from Agent W about the tone of a discussion among scientists in Country X could discriminate H_2 and H_3 from both H_1 and one another, is assessed as .3; the chance that such a report could discriminate H_3 or H_2 from H_1 but not from one another is .2; the chance that it could make no discriminations at all among $\{H_1, H_2, H_3\}$ is .5. In the second phase, we use our knowledge of the content of the evidence (i.e., the tone of the discussion was pessimistic) to map these chances onto possible meanings of the evidence (e.g., we assign support of .3 to H_3; we assign support of .2 to $\{H_2, H_3\}$; and support of .5 to $\{H_1, H_2, H_3\}$).

The upshot of this restriction is that in a Shaferian analysis, we cannot use the content of the evidence (Phase 2) in our assessment of the reliability of the evidential process (Phase 1). Thus we cannot reassess the reliability of an evidential argument based on its agreement or conflict with other lines of argument or with our independent judgment of the plausibility of its conclusion. Modularity thus comes at a certain cost.

While a Bayesian analysis can, in principle, accommodate such considerations, it would require a very large set of conditional assessments, linking all components of one evidential argument with all components of every other argument that has the potential for corroborating or disconfirming it. Indeed, such considerations could also be incorporated into a belief function analysis, if we treat the reliability of a source with respect to a particular type of evidential content as an explicit conditioning event or intermediate hypothesis. (Such an approach, of course, pushes back the constraint against using the content of the evidence, to the assessment of reliability of the evidential processes underlying our explicit reasoning about reliability.) Few Bayesian (or Shaferian) analyses in fact attempt such completeness. Seen in this light, the gain in naturalness, modularity, and expressive power in the belief function approach may offset the largely *pro forma* loss in validity.

How then should conflict or corroboration among different lines of reasoning be handled? Seldom, if ever, do human reasoners laboriously spell out all possible combinations of conflict and corroboration--as required by a Bayesian approach. Human reasoners typically use conflict among witnesses, sensors, or arguments as a symptom of the existence of problems in one or more of the relevant evidential processes and as a prompt for corrective action, such as re-examining the credibility of sources, reconsidering basic assumptions of the analysis, or searching for new information. Typically, an iterative conflict resolution process is adopted, in which an argument or set of arguments is constructed, conflict prompts revisions in beliefs, the revised arguments are again checked for consistency, and the process repeats until a satisfactory result is achieved.

In this light, the theory of belief functions contains the seeds of several useful tools for the resolution of conflict:

(1) *Diagnosis of conflict*. To the extent that two arguments support incompatible hypotheses, the application of Dempster's Rule

results in allocation of support to the null set (support for non-null sets is then increased, via the normalizing constant, so that total support to non-null sets sums to unity). This null set support (equal to 1 minus the normalizing constant) provides a natural measure of the degree of conflict in the evidence. No equally natural measure is available in Bayesian Theory. (e.g., Spiegehalter and Knill-Jones, 1984).

(2) *Assumptions.* To the degree that current evidence is uncommitted in regard to the truth or falsity of a hypothesis, there is room for assumptions. An assumption, therefore, could be naturally represented in Shafer's framework as a decision regarding the allocation of uncommitted belief. Such a decision, by definition, goes beyond the evidence, but remains within the constraints of the evidence. There is no comparable concept in Bayesian theory.

(3) *Discrediting arguments.* The outcome of a process of conflict resolution is typically the discrediting of one or more lines of reasoning that led to the conflict, by rejecting assumptions involved in those arguments. Shafer's concept of discounting is a natural means of representing the outcome of such a discrediting process. In discounting, belief in specific hypotheses is decreased, and proportionately greater support is assigned to the universal set (i.e., the chance that the evidence tells us nothing). Once again, there is no measure of the reliability of an argument in Bayesian theory.

Shafer himself does not address the notion of an assumption, as just outlined. Nor, therefore, does he link discounting to the rejection of assumptions. More fundamentally, as noted above, actions in response to conflict, such as re-examining source credibility, must occur outside the theoretical structure of belief functions. Later, in Section 3.0, we propose a system (based on but going beyond Shafer's theory) which embeds a belief function model within an iterative conflict resolution process, and which utilizes the tools implicit within Shafer's calculus to formalize and direct that process.

In sum, a major strength of Shafer's theory is the naturalness of the input format it imposes: Assessments need go no further than the evidence justifies. "Ignorance" is naturally represented by assigning support to a subset of hypotheses, with no further commitment to an allocation within the subset. A Bayesian must decide among quite definite and distinct, but equally arbitrary, allocations of probability. Weight or reliability of evidence is quite intuitively represented as the degree to which the sum of belief for a hypothesis and its complement falls short of unity. Shafer's theory does not permit reassessment of the quality of an information source in terms of what that source says; the credibility of one witness cannot be increased by corroboration of a second witness or decreased by contradiction. We argue, however, that these processes are best implemented, *in any case*, by procedures of qualitative reasoning that re-examine sources of evidence as an analysis proceeds, and recalibrate them in the light of corroboration or conflict. As we shall see, Shafer's calculus provides a framework within which useful tools for that purpose can be developed.

2.3 Non-monotonic Reasoning

Non-monotonic logic is a direct effort, within the non-numeric tradition of artificial intelligence, to address the problem of assumptions and completeness of evidence. The first application of the ideas of non-monotonic reasoning was by Stallman and Sussman (1977), and since that time the theory has generated intense interest in the artificial intelligence and expert systems communities (e.g., Doyle, 1979; McDermott and Doyle, 1980; McDermott, 1982; Reiter, 1980; Moore, 1985; deKleer, 1986).

Traditional logical frameworks fail to capture the non-monotonicity of natural human reasoning. Specifically, traditional formal logics are monotonic, in that the number of provable statements in the system increases monotonically in time as new axioms or premises are added to the system. In contrast, in a non-monotonic system a theorem may be retracted when new information (axioms) is introduced.

Human reasoning is commonly non-monotonic because conclusions must often be arrived at on the basis of incomplete information. We cannot afford to wait until all possible relevant evidence is obtained. In the face of incomplete evidence, people adopt assumptions, acting as if they are true until evidence arises to the contrary. For example, if a building in Beirut is bombed, and the PLO publicly claims responsibility, we might adopt a provisional assumption that a faction of the PLO is in fact responsible. If later, however, an Iraqi government double agent provides believable testimony that the Iraqi government was behind the attack, we will drop our initial assumption. (We might later re-introduce the PLO assumption if we receive further information which casts doubt on the credibility of the Iraqi agent). Non-monotonic reasoning systems attempt to model this process of revising systems of belief to accommodate conflicting information.

At any point in time, a non-monotonic system has a list of currently believed statements, together with a record of how these beliefs are justified. For example, in the Truth Maintenance System (TMS) developed by McDermott and Doyle, one basic form of representation is:

Statement A   SL<*in*list; *out*list>

In this representation, statement A is associated with a "support list" (SL) justification containing two different sets of statements, the *in*list and the *out*list. A is accepted if all statements in its *in*list are accepted and no members of its *out*list have been accepted. A statement with a non-empty *out*list is an *assumption*: it is accepted provisionally (if its *in*list is accepted), unless or until some member of its *out*list is shown to be the case. In our example above, the initial attribution of responsibility to the PLO was an assumption, adopted on the basis of the public claim (*in*list), but subsequently dropped when evidence was obtained to the contrary from the Iraqi double agent (*out*list). Similarly, the credibility of the Iraqi agent was itself an assumption, adopted until evidence to the contrary is obtained.

As long as new information is consistent with current beliefs, the system incorporates the new information by combining it with the currently believed statements, using its inference rules to derive new beliefs. It is possible, however, for new information to lead to an inference that contradicts a currently held belief. When this happens, a process of dependency-directed backtracking is initiated. The system traces back through the network of justifications to find the assumptions upon which the contradictory inferences depend, and makes revisions to achieve consistency. One assumption is selected as the "culprit," and is retracted by *assuming* that some member of its *outlist* is true. This causes any statements depending on the culprit assumption to be disbelieved. The process of retracting assumptions and changing truth values continues until a consistent set of beliefs is obtained.

Unfortunately, the procedure for selecting among alternate belief revisions is arbitrary in two ways: typically, more than one assumption is implicated in a contradictory inference, and there is more than one way to reject each of them, by assuming a member of its *outlist*. Moreover, the theory lacks a measure of the *degree* of support for beliefs. Such a measure could provide the basis for selecting among possible belief revisions and would provide users an index of confidence in the conclusions of the argument.

Non-monotonic reasoning is typically viewed as an alternative to systems based on the probability tradition, which employ numerical measures of uncertainty. In the terrorist example, a Bayesian or Shaferian system would assign a numerical degree of support to the different hypotheses concerning who was responsible for the bombing. Uncertainty is expressed by assigning degrees of support of less than unity to each of the hypotheses. When further information is received, degrees of support are updated to incorporate the new information. In the sense that probabilities may go down as well as up, Bayesian and Shaferian systems are "non-monotonic," and belief functions in particular have been recommended as a numerical version of non-monotonic reasoning.

We think this is wrong. Current numerical systems are monotonic in at least three important senses: (a) Bayesian and Shaferian theory both lack a mechanism for provisionally accepting an uncertain hypothesis; (b) once a conclusion is declared certain (i.e., degree of support = 1), its support cannot be reduced; and (c) in Shafer's system, new evidence can only decrease the range of possible belief (i.e., the difference between Pl(H) and Bel(H)), but can never increase this measure of ignorance (e.g., by prompting the rejection of assumptions). As they stand, neither Bayesian nor Shaferian theory is an adequate substitute for the assumption-based reasoning in non-monotonic logic.
¶ A deeper look at the distinction between numerical theories and non-monotonic logic reveals a fundamental difference in their attitudes toward conflict. In both Bayesian and Shaferian theories, divergence among uncertain lines of

reasoning can be loosely characterized as "stochastic": except when probability or belief equals 1, conflict is *expected* to occur some small percentage of the time, even when both of two conflicting lines of reasoning are normatively correct, due to the imperfect correlation or causal link between cues and hypotheses, or the chance accumulation of small errors in measurement. From the point of view of non-monotonic logic, however, divergence can be characterized as "epistemic": it is a result of faulty beliefs or methods. Conflicting results are taken as evidence that one or more premises or forms of argument that led to the conflict are mistaken.

These two conceptions of conflict lead to different rationales for the process of combining evidence or lines of reasoning. From the first point of view, the object is to reduce variance by a blind process of statistically aggregating evidence, akin to that in which chance errors tend to cancel one another out across repeated measurements. Both Bayesian updating and Dempster's Rule fall into this category, as does the theory of uncertainty embodied in MYCIN. From the other point of view, however, the object is to improve the overall truth of a system of beliefs--to explicitly identify potentially erroneous steps in the argument and to change them.

We argue that these views of conflict are *complementary* rather than *competitive*. Probabilistic arguments, although they lead to conclusions in the form of probabilities or degrees of belief, nevertheless depend on assumptions in the same way that deterministic arguments do. Assumptions, whether explicit or implicit in a probabilistic or belief function analysis, include some that pertain to modeling (e.g., normality, independence, linearity) and some that pertain to substance (e.g., the credibility of a source, proper functioning of a technical collection system, continued accuracy of a dated observation, absence of a conspiracy to deceive among apparently unrelated sources). Not all of these assumptions, of course, are made all the time; but unless *some* such assumptions are made, analytic arguments (whether deterministic or probabilistic) are condemned to perpetual inconclusiveness. Put another way, it is never possible to rule out, on the basis of positive evidence, *all* the factors that could undermine the validity of a particular argument. Therefore, in order to construct a Bayesian or Shaferian argument that discriminates adequately among the possible hypotheses, it is nearly always necessary to *assume* the absence of such factors.

We will argue in Section 3.0 that non-monotonic logic has its most useful application as a control process for the application of an uncertainty calculus. Its role is to keep track of assumptions and direct the process of belief revision when those assumptions lead to anomalous results.

2.4 Toulmin's Model of Argument

In the preface to *Uses of Argument*, Toulmin (1958) rejects as confused the "conception of 'deductive' inference which many recent philosophers [and, we may add, AI researchers] have accepted without hesitation as impeccable." Toulmin's motivation in that book is to turn away from the highly abstract character of traditional logic; to examine actual methods of reasoning in different substantive areas, such as law and medicine; and to develop a theory of logic capable of capturing the rich variety of methods in everyday use. In so doing, he developed a framework for argument which may provide a useful linkage between numeric and non-numeric approaches to incompleteness of evidence.

The basic framework of an argument, according to Toulmin, is as follows (Toulmin, et al., 1978):

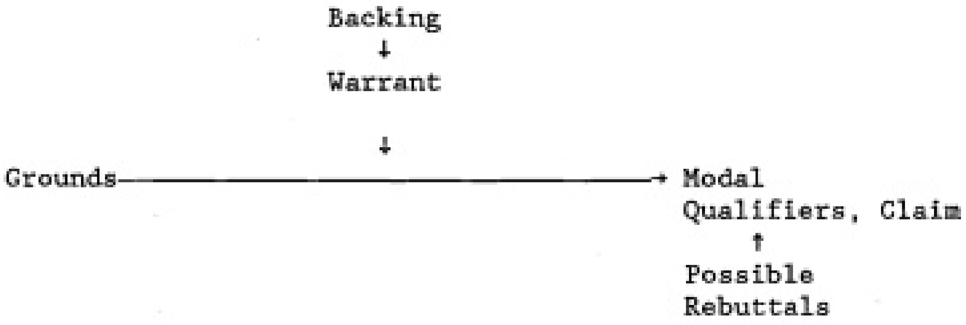


Figure 2-1. Toulmin: Strucuture of an Evidential Argument

A claim, or conclusion whose merits we are seeking to establish, is supported by grounds, or evidence. The basis of this support is the existence of a warrant that states the general connection between grounds and conclusion: e.g., a rule of the form, if this type of ground, then this type of conclusion. The backing provides an explanation of the warrant, i.e., it provides a basis (theoretical or empirical) for the existence of a connection between ground and claim. Modal qualifiers (e.g., "probably," "possibly," "almost certainly") weaken or strengthen the validity of the claim. Possible rebuttals are factors capable of deactivating the link between grounds and claim, by asserting conditions under which the warrant would be invalid. A way of reading this structure is: Grounds, so Qualified Claim, unless Rebuttal, since Warrant, on account of Backing.

Toulmin rejects the subjectivist's concept of probability as degree of belief, on the basis that this is incompatible with the natural requirement that estimates of probability be reliable. If such estimates were purely subjective, there would be no sense in asking whether or not they were reliable. On the other hand, Toulmin also rejects the objectivist's definition of probability in terms of frequencies, on the basis that such a definition confuses the meaning of probability (i.e., as a qualification of a conclusion) with the reasons for regarding the event as probable (i.e., the observed frequencies). In fact, he contends that, "the attempt to find some 'thing', in terms of which we can analyze the solitary word 'probability' and which all probability-statements whatever can be thought of as really being about, turns out to be a mistake" (p. 70). He defines probability as a modal qualifier whose function is to qualify the strength of the link between grounds (evidence) and conclusion. This notion is closely akin to Shafer's concept of belief as an evaluation of the validity or reliability of an evidential argument.

Toulmin's framework also bears some important resemblances to non-monotonic logic. Both depart from traditional logic by providing for a process in which conclusions are accepted *unless* other propositions (members of the *outlist*; rebuttals) turn out to be true. There are two important differences: (1) Toulmin proposes a highly differentiated knowledge structure, in which the roles of grounds, warrant, backing, conclusion, and rebuttals are distinguished within an argument, while non-monotonic logic proposes a much more homogeneous, undifferentiated knowledge structure; (2) Toulmin provides for graded or qualified acceptance of conclusions.

Toulmin's model may find its most significant application as the basis for a data structure representing the uncertainty in a single argument, to be combined with other arguments via some numerical uncertainty calculus. Because of the focus in Shafer's theory on evidential links, a synthesis of Toulmin's theory with the theory of belief functions seems a promising avenue of research, and will be explored further in Section 4.3.0 X

Toulmin fails to explore a further important dimension which relates numeric and non-numeric approaches: i.e., the linkage between probabilistic modal qualifiers (which, as noted above, assess the quality of an evidential argument) and the assumptions upon which the argument depends (the rebuttals). In one sense, of course, an argument is *weakened* to the degree that it depends on assumptions; in such cases, says Toulmin, we add the modal qualifier, "presumably," to the conclusion. In another sense, however, we become *more* certain of a conclusion, the more we assume. When the conclusion of an argument is insufficiently definitive for the purposes at hand, the argument can be strengthened (i.e., probabilistic qualifiers express a higher degree of confidence) by making additional assumptions, so long as it is understood that the validity of the strengthened argument depends on these new assumptions. Conversely, when new evidence forces the rejection of an assumption (i.e., a possible rebuttal turns out to be true), the assessed strength of the argument must be further qualified or reduced. In Section 3.0, Toulmin's framework is expanded to include these tradeoffs.

2.5 Theory of Endorsements

Paul Cohen's (1985) theory of endorsements, like non-monotonic logic, is a descendant of the AI-based logic tradition. Cohen shares with Toulmin, however, a concern to replicate the natural, qualitative structure of ordinary processes of reasoning. It is of interest, therefore, that both Cohen and Toulmin focus on an evaluation of arguments that purport to establish a link between a conclusion and evidence. We have already noted that Shafer's numerical theory had the same concern. For Shafer, however, belief about the validity of such an argument can be adequately summarized in a numerical measure, the belief function, i.e., the likelihood that the evidence proves the hypothesis. To Cohen, by contrast (as for Toulmin), it seems unnatural to assess the strength of an argument without actually saying what the argument is. The problem with numerical approaches, he argues, is that they allow us to state beliefs without providing justifications. The theory of endorsements attempts to provide a consistent format for representing and evaluating arguments.

In Cohen's theory, each hypothesis is associated with a "ledger" of confirming and disconfirming "endorsements." Endorsements state reasons for believing or disbelieving an argument that purports to establish a hypothesis. (Since the system implemented by Cohen has a rule-based architecture, such arguments take the form of data and rules.) Finally, Cohen's theory contains procedures for ranking different types of endorsements in terms of importance, for determining when a set of endorsements qualifies a hypothesis for acceptance, and for resolving conflicts among hypotheses in terms of the relative strengths of their endorsements.

Cohen's classification of endorsements by type is of interest to us, since it sheds light on qualitative differences among the considerations that influence the reliability of an argument. Cohen argues that procedures for ranking hypotheses and resolving conflicts are affected differently by different types of endorsements. Five categories of endorsements are distinguished:

(1) **Rule-condition** endorsements characterize the relationship between a condition clause in the antecedent of the rule and the consequent of the rule. For example, endorsing a clause as *flexible* or *inflexible* specifies whether data must exactly match the rule antecedent for the conclusion to be acceptable, or if some degree of approximation is sufficient. Other such sets of endorsements include *maybe-too-general*, *maybe-too-specific*, *exact;* and *supportive*, *necessary*.

An additional set of endorsements in effect permits condition clauses to be treated as assumptions. A clause not-x that is endorsed by *ostrich* is accepted as long as x is *not* acceptable (i.e., adequately endorsed); a clause not-x endorsed by *closed-world-assumption* is accepted if an attempt to prove x has failed.

(2) **Rule-inference** endorsements characterize the basis for deriving a conclusion from its conditions: i.e., *model-based*, *causal*, or *correlational*.

(3) **Data** endorsements are characterizations of the data by the person who supplies it; they include *source*, *type-of-data*, and *accuracy*.

(4) **Task** endorsements involve a comparison of the potential conclusion of an inference with the conclusions of other inferences. They include *corroborate*, *conflict*, *redundant*, and are used in determining the sequence with which rules are applied to the database.

(5) **Conclusion endorsements** are relevant *after* a rule has been applied and conclusions actually drawn. They include *corroborate*, *conflict*, and *redundant*, as in (4). Additional endorsements of this type (*unlikely*, *modal*) involve a comparison of the conclusion to prior beliefs about its "usualness." A final endorsement (*unwarranted*) indicates that a conclusion is assumed.

Cohen's approach captures a significant aspect of actual reasoning: the dependence of belief on qualitative characteristics of an argument. In a sense, therefore, it goes beyond Toulmin's argument framework by specifying relevant *features* associated with each component of the argument. Figure 2-2 shows how Cohen's taxonomy of endorsement types might be mapped into Toulmin's framework.

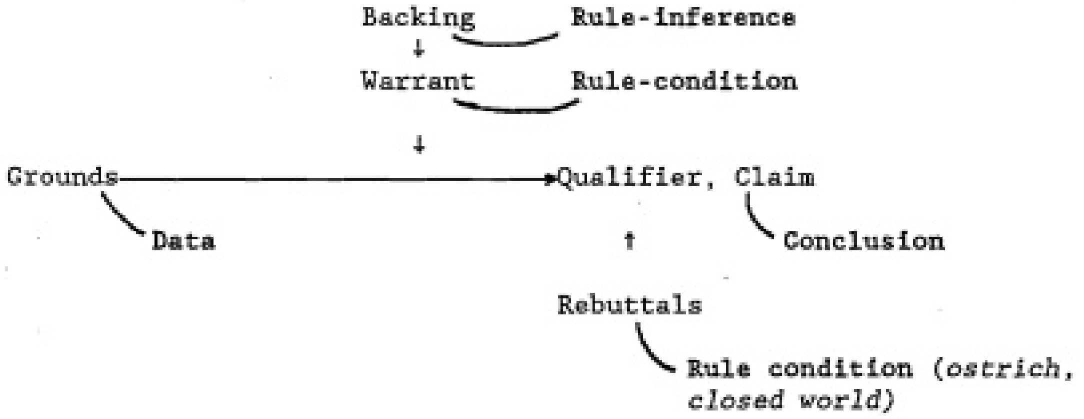


Figure 2-2. Endorsement Types Associated with Toulmin's Argument Structures

The chief difficulty with Cohen's theory is the *ad hoc* nature of the qualitative mechanism for ranking endorsements. For example, Cohen provides a scheme which divides endorsements very coarsely into two categories: "preferred" versus "not so good or downright bad". However, when hypotheses that have *both* "good" and "bad" endorsements must be ranked, additional stipulations are required. Further, even if all endorsements could be ranked by preference, the problem of appraising *groups* of endorsements would be unsolved. Thus, the decision of whether one proposition is better endorsed than another is quite arbitrary, and available procedures can often be insufficiently powerful to resolve conflicts.

Cohen himself expresses the view that a hybrid approach might ultimately be desirable, in which the "strength" of an endorsement is numerically assessed. However, without some guidance as to the nature or meaning of such numbers, a "hybrid" approach might ultimately prove to be equally as ad hoc as Cohen's qualitative system. In Section 3.0, we propose a method for quantifying the impact of endorsements that is firmly based on an extension of belief function theory.

2.6 Alternative Probabilistic Approaches to Argument

The basic intellectual building blocks of the system to be described in Section 3.0 have now been laid out:

- assessing the reliability of arguments (Shafer, but also Cohen and Toulmin);
- making assumptions explicit (non-monotonic logic, Toulmin);
- resolving conflict by revising assumptions (non-monotonic logic);
- representing arguments in a natural structure (Toulmin); and
- identifying features of arguments that underlie assumptions and affect reliability (Cohen).

We have argued that the belief function calculus may be a more appropriate starting point than Bayesian theory for an assumption-based conflict resolution process: because it provides a modular representation of evidential arguments and a definition of conflict, and because it already contains the basic elements of mechanisms for introducing assumptions and discrediting arguments. Although Bayesian theory has a strong axiomatic foundation, it lacks the expressive power to quantify the reliability of an argument or to measure ignorance.

The latter claim would in fact be contested from a variety of directions by Bayesians who have developed variants of the standard approach (as discussed in Section 2.1). Here we can only very briefly identify some of these variants and indicate the reasons for our current dissatisfication.

2.6.1 Probabilifying logic. Applications of probability theory usually assume, implicitly, that the only relevant structure among propositions is reflected in the probability assessments which we make regarding them. For example, if

$$\frac{\Pr[D|H]}{\Pr[D|\bar{H}]} > 1,$$

then D is linked to (i.e., is evidence for) H. There is no way to determine this linkage by simply examining D and H. Yet consider the following argument:

P1 (Warrant): If witness X says he saw the installation, it is there.
P2 (Datum): Witness X says he saw the installation.
P3 (Claim): The installation is there.

Here, P3 is a *logical* consequence of the conjunction of P1 and P2. Thus, if we know the probability of P1&P2, we can derive an upper and a lower bound on the probability of P3. Logical relationships among propositions thus place constraints on the probabilities, and perhaps hold out some hope for using probability theory to assess the reliability of arguments.

Several recent theories (Nilsson, 1984; Lagomasino and Sage, 1985) address the logical relationships among the sentences whose probabilities are being assessed.

Lagomasino and Sage (1985). These authors present a framework for imprecise inference that purports to combine Toulmin's logic of reasoning and the calculus of probability. In fact, we would argue that their use of Toulmin is quite incidental to their basic approach. A better characterization is that Lagomasino and Sage attempt to probabilify traditional logical relationships, and this turns out to be inconsistent with evaluation of an evidential argument.

Lagomasino and Sage use Toulmin's model of argumentation to frame the relations among events, and to structure an inference model. In particular, the relationship between two events, grounds D and claim C, are represented as:

$$\begin{tikzcd}
 & W = (D \to C) \arrow[d] & \\
D \arrow[r, dash] & \otimes \arrow[r] & C \\
 & R = (\bar{D} \to C,\ \bar{D},\ \bar{C}\ \ldots). \arrow[u] &
\end{tikzcd}$$

where W refers to warrant, and R refers to rebuttal. [Their use of the term rebuttal to include the negation of grounds or claim appears at odds with Toulmin's (1958) definition of rebuttal as "indicating circumstances in which the general authority of the warrant would have to be set aside" (p. 101).]

The basic methodology is, first, to set up constraints on the probabilities (e.g., Pr(D) or Pr(D → C)) based on logical relationships and probability theory. Information about the problem domain is then encoded as additional constraints (e.g., Pr(D) > .5). Finally, upper and lower bounds for any probability can be obtained by solving the appropriate linear programs (with objective function min Pr(·) or max Pr(·)). In this framework, constraints on the probability of the warrant, Pr(D → C), combine with constraints on the probability of the data, Pr(D)--as well as constraints on probabilities of various rebuttals--to yield constraints on the probability of the conclusion, Pr(C).

At first glance the probability of the warrant Pr(D → C), appears to give a probabilistic assessment of the reliability of the argument from D to C. If D → C is true, the argument is reliable and C is true (given D). Thus, equating Pr(D → C) with belief in C when D is known to be true, appears to parallel Shafer's concept of support (see Section 2.2 above). Unfortunately, the parallel with Shafer is quite spurious.

Pr(D → C) *cannot* plausibly by construed as the probability (or strength) of an evidential link between D and C as long as "D → C" is interpreted within traditional logic (as the authors clearly intend). Within traditional logic "D → C" is true unless D is true *and* C is false. Thus, $P(D \to C) = P(\bar{D} \text{ or } C)$. By the axioms of probability theory, this equals $P(\bar{D}) + P(C) - P(\bar{D} \text{ and } C)$. Therefore, in assessing the probability of D → C, we are focusing on the chance the antecedent is false and the chance the conclusion is true, and are not addressing the existence of any physically or logically real connection between the antecedent and the consequent. For example, "If the moon is made of green cheese, then the PLO is responsible for the attack" would be true in traditional logic, since the antecedent is false; yet clearly there is no evidential connection. Similarly, "If Albany is the capital of New York, then Reagan was President in 1986" is also true, since the consequent is true. A warrant construed in this way is quite trivial. It in no way addresses the *impact* of belief in D on belief in C, i.e., the strength of an argument.

There is another respect in which Lagomasino and Sage's approach violates our intuitions about an evidential argument. Since $D \rightarrow C$ is true whenever C is true, but is also true if D is false, $Pr(D \rightarrow C) \geq P(C)$; i.e., $Pr(D \rightarrow C)$ is an *upper bound* on the probability of the conclusion, C. In probabilified logic, if $Pr(D \rightarrow C) = 0$, then $P(C)$ must also equal 0. But if $Pr(D \rightarrow C)$ did in fact capture the strength of an evidential argument, it would provide a *lower bound* for belief in C (analogous to Shafer's $Bel(\cdot)$). The reason is that C might be supported by other arguments in addition to $D \rightarrow C$. (By the Law of Total Probability, the probability of C is simply the sum of the probability that the argument based on D is valid plus the probability that that argument is invalid and some other argument is correct.) If the argument based on D is unreliable, then it should tell us *nothing* about the chance of C (recall our discussion of ignorance in Shafer's system). Thus, $Pr(D \rightarrow C)$ does not adequately quantify our intuitions regarding the reliability of the argument based on D.

An alternative interpretation of "$D \rightarrow C$" is as an implicit universal generalization, $\forall x(Dx \rightarrow Cx)$, i.e., all instances of D are also instances of C. The probability of "All D's are C's" is not an upper bound on the probability that a particular D is C, since the latter is not sufficient to establish the former. On the face of it, therefore, this appears to assert a more general *relation* between D and C, which could be used non-trivially to support an argument for C based on D in a particular case. Unfortunately, this will not work either. Such a generalization may be true because the antecedent is false for all instances or because the consequent is true for all instances. X
Thus, "For all x, if x is a 20 foot tall person, then x is a spy" would be true, since no one is 20 feet tall; similarly, "For all x, if x is Russian, then x cannot run faster than the speed of light" would be true, since nothing

can exceed the speed of light. In neither case is there a true evidential link between the predicate in the antecedent and the predicate in the consequent. Finally, note that a single counter-example (i.e., a case of D and not-C) is sufficient to establish falsity; i.e., Pr[∀x(Dx → Cx] would be zero. Yet we often assert the existence of evidential relations (e.g., "the public claim of responsibility suggests the PLO is to blame") even when the relationship is subject to exceptions (some public claims of responsibility are spurious).

The source for all these confusions, we believe, is in the well-known inability of traditional logic to capture what is essentially a *subjunctive* or counterfactual relationship. According to a recent discussion by Nozick (1981), datum D is in an evidential relationship to conclusion C when D "tracks" C, in the sense that (a) if C were *not* true, D would not have been true; and (b) if C had been true in a different but reasonably similar circumstance, D would still be true. Nozick argues persuasively that *both* conditions are required to exclude cases of accidental or lucky true belief where a person would not be said genuinely to *know* what he believes. When we assess the reliability of an argument, we are assessing the chance that some mechanism (causal or logical) exists which produces tracking of this sort. For example, when we assess the warrant P1 in the example above ("If witness X says he saw the installation, it is there"), we are concerned with witness X's *eyesight* and *truthfulness* on this occasion: these are the mechanisms which might cause X's testimony (D) to track the claim (C). Formulations in traditional logic do not capture this.

A more promising alternative would be to drop the truth-functional interpretation of D → C as not-D or C, and to probabilify some version of modal, rather than traditional, logic. Thus, we might be asked to assess constraints on the probability of "N(D → C)", where "N" indicates an "evidentially necessary" link. There is considerable discussion regarding the construal of subjunctive statements in terms of a possible would semantics (e.g., Lewis, 1976). But "N(D → C)" might mean that C is true in all possible worlds where D is true and which are similar to some requisite degree to the actual world. It is sufficient here to note (a) that no plausible metric of similarity has as yet been specified, and (b) that no attempts have been made, to our knowledge, to apply probabilistic constraints to sentences in a modal logic.

In sum, although Lagomasino and Sage's framework claims to be based on Toulmin's structure of argument, the similarity is superficial. The classical interpretation of logical connectives is inconsistent with using probabilified logic to assess evidential arguments.

Nilsson (1984). Nilsson presents an approach that in essential respects is similar to that of Lagomasino and Sage. His method is presented as a generalization of classical first-order logic that is "appropriate for representing and reasoning with uncertain knowledge."

Nilsson starts by specifying a logical sentence whose truth values are of interest. These could be any conjunction of sentences of first-order logic; for example:

$$S = \{D,\ D \rightarrow C,\ C\}$$

The truth-value of any one of the three components of this sentence is bounded by logical consistency relationships. For example, all three components could be true; this is logically consistent. However, the three components could not all be false; this is inconsistent, since D → C is true if D is false. Note that this bounding is based on the combination of truth-values for all components of the sentence, not any individual component. Indeed, in the example any component could be true or false (value of 0 or 1); it is only combinations that are prohibited.

Each permissible combination of truth-values represents a "possible world," that is, a possible combination of true and false components. If the truth or falsity of each component is represented by the number 1 or 0 respectively, then a possible world can be represented as a three-dimensional vector of zeros and ones for a permissible state. In the example above, the following four vectors represent all possible worlds:

$$\begin{array}{c}[1,1,1]\\ [1,0,0]\\ [0,1,1]\\ [0,1,0].\end{array}$$

If each component of the sentence is thought of as a dimension in three-space, then possible worlds are represented as four points in that space.

Nilsson next generalizes the interpretation of the vector by allowing probabilistic "smearing" over worlds. This is done by allowing probability distributions over different worlds and by defining the probability of a component as "the sum of probabilities of all possible worlds in which it is true." Under this definition, probabilities of components will be logically "permissible," in the sense that they must fall within the convex region bounded by the set of possible worlds.

"Probalistic entailment" is a process in which we are given the probabilities of some sentences and then compute bounds on the probabilities for other sentences by using the logical constraints described above. Since logical consistency by itself rarely determines probability uniquely, Nilsson investigates supplementary techniques. He both solves for "maximum entropy" probabilities and those produced by geometric projection, although neither method is provided with a basis or defended. In addition, one could presumably assess probabilities of "possible worlds" directly to derive the desired probabilities, although the assessment problem here seems immense.

The principal difficulty of Nilsson's approach, from the present viewpoint, however, is that (like Lagomasino and Sage) it fails to capture true evidential relationships. As noted above, these are not adequately represented in the first-order predicate calculus. Nor is it clear how effectively Nilsson's method could be extended to handle consistency constraints among sentences in a modal logic. In any case, it is likely that the assessment task would be enormously complicated (e.g., by the introduction of possible worlds containing sets of possible worlds).

Shafer and Logic. It is illuminating to contrast Shafer's approach with those discussed in this section. $m_1(C)$ can be interpreted as the chance that evidence collection 1 means that C is true; it depends on an assessment of the mechanisms that underlie the reliability of the evidential argument, e.g., the chance that the witness had adequate eyesight and is truthful. As we noted in Section 2.2, in Shafer's system evidential mechanisms are assessed in a general way, independently of the actual content of the evidence; thus, whether the witness had testified to C or to not-C, honesty and good eyesight would mean he should be believed; lack of honesty or poor eyesight mean his testimony should be disregarded. Shafer's measure of support, $m(\cdot)$, thus captures the notion of an evidential link directly. For example, given three hypotheses $\{H_1, H_2, H_3\}$, $m(H_2 \text{ or } H_3) = .3$ means there is a .3 chance that the relevant evidential process (e.g., reports from a particular witness; statistical arguments making a certain type of assumption; etc.) can discriminate H_2 and H_3 from H_1, but not from one another. Outputs from such an evidential process would vary depending on whether the truth were in H_1 or in $\{H_2, H_3\}$. In short, the evidential process "tracks" H_1 versus $\{H_2, H_3\}$ with probability .3.

In passing, we note that Nozick's (1981) concept of tracking may shed some further unexpected light on Shafer's approach. Evidence D "tracks" conclusion C in the sense that (a) if C were not true, D would not be true, and (b) if C were true (in somewhat different circumstances) D would still be true. But this formulation turns out to imply (by substituting not-D for D and not-C for C) that D tracks C if and only if not-D tracks not-C. Thus, the assessment of whether "tracking" exists (like assessments of $m(\cdot)$) is independent of the content of the evidence, i.e., whether D or not-D in fact occurs. Failure to conditionalize on the evidence, while a shortcoming from the Bayesian point of view, thus has a justification of sorts in its correspondence to a well-founded concept of knowledge.

Shafer does not postulate a *logical* relationship between data and warrant, on the one hand, and conclusion, on the other. Thus, unlike other authors in this section, he does not seek to exploit such a relationship to provide a bound on belief in the conclusion. He directly *stipulates* that such a bound is provided by the assessment of the reliability of the evidential link between D and C.

Nevertheless, it is possible to apply a belief function analysis to logically related sentences. Suppose independent belief functions have been assessed on two sets of hypotheses $\{\ldots A_i \ldots\}$ and $\{\ldots B_j \ldots\}$ which, in combination, may have logical implications for a third set of hypotheses, $\{\ldots H_k \ldots\}$. Then it can be shown by methods discussed in Yager (#MIT-508) that

$$m(H_k) = \sum m_1(A_i) * m_2 (B_j)$$

where the summation is over all i, j such that H_k is logically derivable from A_i & B_j. Lower and upper bounds on belief in subsets of $\{\ldots H_k \ldots\}$, e.g., $Bel(H_k)$ and $Pl(H_k)$, may then be computed in the usual way from $m(\cdot)$. In our simple example of truth-functional implication, $m(C) = m(D) * m(D \rightarrow C)$. This approach is easily generalized to the combination of more than two sets of hypotheses.

To what extent can incorporation of logic within a belief function framework improve on probalified logic? In particular, does $m(D \rightarrow C)$ capture an evidential link between datam D and conclusion C, or does it merely depend, like $Pr(D \rightarrow C)$, on belief in not-D or belief in C? Recall that the truth of $D \rightarrow C$ means simply that it is not the case that D is true and C false; therefore, $D \rightarrow C$ is true if D is false, whether or not C is true, and if both D and C are true. Thus, in a Shaferian analysis, support for $D \rightarrow C$ can be represented as support for a subset of the hypotheses in the Cartesian product $\{D, \overline{D}\} \times \{C, \overline{C}\}$, in particular, $m(D \rightarrow C) = m(C \text{ or } \overline{D}) = m(\{C \times D\} \cup (\{C, \overline{C}\} \times \overline{D}))$. (Note that this is precisely the expression used in Section 2.2 to represent the support for a hypothesis conditional on evidence.) The key point is that $m(C \text{ or } \overline{D})$ is not a function of $m(C)$ and $m(\overline{D})$, as in Bayesian probability theory. Thus, $m(D \rightarrow C)$ is not affected by evidence that bears directly on the falsity of D; such evidence is summarized by $m(\overline{D} \times \{C, \overline{C}\})$. Nor is it affected by evidence that bears directly on the truth of C; that evidence is summarized by $m(C \times \{D, \overline{D}\})$. Rather, $m(D \rightarrow C)$ summarizes evidence that bears directly on the falsity of D&not-C, i.e., on the *link* between D and C. Thus, $D \rightarrow C$ may keep its proper truth-functional interpretation (true if D is false or C is true); there is no need to apply belief to sentences in a modal logic. Incorporation of $D \rightarrow C$ within the context of $m(\cdot)$ removes the triviality and focuses concern on the existence of a real evidential connection.

note: (in modal semantics (in S-1

2.6.2 <u>Higher-order probabilities</u>. A quite different Bayesian approach is to assess higher-order probability distributions reflecting confidence in the first-order probabilities (e.g., Lindley, Tversky, Brown, 1979; Tani, 1978). The second-order assessments capture the amount of knowledge (or ignorance) underlying an assessment of the first-order probabilities. Thus, a very sharply peaked second-order distribution centered on .5 reflects a high degree of confidence that .5 is the "true" first-order probability (e.g., the conclusion of a long process of sifting evidence). But a flat second-order distribution whose mean is .5 reflects a high degree of ignorance (e.g., prior to consideration of any evidence). Second-order distributions of this sort thus appear to increase the expressive power of Bayesian theory, by capturing completeness of knowledge or reliability of an argument. And they may be used in higher-order inferences (e.g., regarding hypotheses about the "true" first-order probability), to resolve conflicts among inconsistent probabilistic analyses.

This approach, however, suffers from a variety of drawbacks. In the first place, it threatens an infinite regress of high-order judgments; secondly, the assessment burden can be massive (especially if higher-order non-independence were to be taken into account); and thirdly, the formal justifications associated with Bayesian theory, e.g., avoiding a "Dutch book," are not compelling at the level of second-order bets (or bets about bets).

Most importantly, however, the meaning of second-order probabilities is itself unclear. One approach is to regard them as capturing "measurement error" in assessing first-order probabilities. But how can we be uncertain about our own current subjective probabilities (especially if we construe them as equivalent to choices among gambles)? More to the point, why would such uncertainty (if it exists) be of interest in resolving conflicts among lines of reasoning?

To make the latter question vivid, consider two probabilistic analyses regarding the same hypothesis. The first analysis yields a probability of .8; the second yields a probability of .6. According to the measurement error approach (Lindley, Tversky, and Brown, 1979), reconciliation of the two analyses gives us an estimate of our "true" probability that typically lies between .6 and .8 - e.g., .7. What this ignores is the possibility that the two analyses tapped independent sources of evidence. Suppose that is the case, and that the two collections of evidence both favor the hypothesis. Then the second analyses should cause us to increase our probability (e.g., from .8 to .9), not decrease it (from .8 to .7). In other words, the measurement analogy (applied in this way) precludes use of second-order probabilities to evaluate distinct evidential arguments . It treats the problem as an inference about someone's inner state rather than about the world, and fails to combine the two arguments properly in terms of the amount of independent evidence in each.

Suppose instead that we assess second-order probabilities regarding the components of a single probabilistic analysis (e.g., likelihood ratios and prior odds in Bayesian updating), and then compute the second-order distribution for the output probability (e.g., the posterior odds). Unfortunately, the spread of the second-order output distribution is *larger* the more components we introduce to the argument (since measurement error propagates to the conclusion in a standard way). Yet we would expect the *reliability* of the output probability to be greater as the number of items of evidence we consider increases. Again, it seems clear that measurement error does not capture reliability of the argument.

An alternative interpretation of second-order probabilities refers to our *future* subjective probabilities; i.e., we assess the chances that our first-order probability will assume various possible values after we obtain further information. To produce such an assessment, however, we must make rather arbitrary decisions to delimit in terms of time or effort the future evidence we consider relevant (since most if not all probabilities will become either zero or one after *all* future evidence is obtained); we must then make extremely speculative and hypothetical judgments about what we are likely to learn within the delimited sphere.

Even then, the usefulness of the resulting assessments is in doubt. Suppose our task is to decide whether or not to *revise* a probability assessment derived by method A, because of conflict with another assessment derived by method B. It seems pointless to demand a *direct* assessment of the shiftability of the probability derived by A in the face of future evidence (e.g., B): it is shiftability that we are trying to *determine* in resolving the conflict between A and B. In other words, the second-order probabilities don't help much in this problem, since they require that the problem be solved before they can be assessed.

It seems more useful to obtain assessments of the *reliability* of argument A and of the *reliability* of argument B. The essential difference is that in assessing Bayesian second-order probabilities, one must somehow consider possible future evidence whether it pertains to the present argument or to other (possibly unanticipated) types of argument. In assessing reliability, however, we focus on the argument at hand: how thoroughly have the assumptions been checked, where could it go wrong and how? To be sure, shiftability is correlated with reliability: The more reliable we think an argument is in its own right, the less likely it is that some future argument will be devised which impugns it--and we will assess a correspondingly narrow range of future probability values. Nevertheless, reliability seems to be the more fundamental, and easier, judgment.

2.6.3 Evidentiary value. A final Bayesian approach, proposed by a group of Swedish researchers (Hallden, 1973 Edman, 1973; Gardenfors, Hansson, and Sahlin, 1983) is closer in some respects to Shafer's work. As for Shafer, the focus of attention has turned from the truth of a hypothesis to the probability that the evidence proves the hypothesis. The essential difference between the Swedish work and Shafer's is that the former conditionalizes the assessment of an argument's credibility on what that argument (and other arguments) have actually concluded. This work however, lacks most, if not all, of the virtues of the belief function representation (see Shafer, 1984). Formulations which conditionalize on the evidence become extremely complex even for the simplest examples. Little progress has been made in deriving rules for the combination of evidence involving the full range of cases to which Dempster's rule applies. Finally, this work sacrifices a significant virtue of Shafer's system, the ability to segment evidence into independent arguments.

2.6.4 Conclusions. In sum, a variety of extensions to traditional Bayesian Theory have been proposed, with the explicit object of capturing such notions as causal or evidential relationships and completeness of evidence. These extensions include (a) applying probabilities to logically interrelated sentences, (b) applying probabilities to probabilities, and (c) assessing the probability that an evidential relationship exists, conditional on the evidence. The first approach fails because of a lack of expressive power in the first-order predicate calculus to represent causally related sentences. The second approach fails because of the lack of an appropriate, evidentially relevant interpretation of higher-order probabilities. The third approach does capture the relevant concepts, but sacrifices simplicity and modularity.

The model presented in Section 3.0 addresses this problem in an altogether different way: not by elaborating on a first-order uncertainty calculus, or re-applying it at a higher level, but by attention to *the reasoning process by which the calculus is applied*. It thus retains both the simplicity and the modularity of Shafer's representation (and its direct approach to the concept of an evidential relationship), but captures interdependencies by embedding a Shaferian argument within a corrective process of heuristic reasoning.

Let us stress, however, that a belief function approach is not necessarily the only, or even the best, model of uncertainty for incorporation in such a heuristic conflict resolution process. We are impressed by the naturalness with which it can be made to serve in such a process (Section 2.6.2 above), and have found other approaches less satisfactory in this respect. However, we would argue that the application of *any* uncertainty calculus is based on assumptions, hence, is guided by a heuristic, iterative process. Other approaches which in the future way prove more satisfactory, but which at present seem insufficiently well developed, are Kyburg's notion of convex Bayesian sets (1985, 1986) and Nau's axiomatic extension of Bayesian probability to allow nested sets of probability intervals with associated measures of confidence (1986).

2.7 Fuzzy Set Theory

In this section, we turn to a quite different type of uncertainty: the fuzziness or vagueness that characterizes ordinary language in virtually all fields of reasoning. Since L.A. Zadeh advanced fuzzy set theory in 1965, an enormous amount of interest, and a very large literature, has been generated. Most of this interest has been theoretical, concerned with the mathematical implications of the theory, but there have been a number of attempts to apply the theory to practical problems. Zadeh argued that most highly complex problems could not be understood or modeled at the level of precision demanded by traditional analytical tools. The appropriate (and intuitively natural) approach to such problems involves imprecise concepts and methods of reasoning. Attempts to eliminate such imprecision inevitably involve arbitrary, *ad hoc* decisions and, in the end, distort one's understanding of the central facts. Nevertheless, analysis (especially with computers) requires precision. To resolve this paradox, Zadeh introduced the now well-known concept of the fuzzy set--a set with imprecise boundaries. While ordinary sets have all-or-nothing membership functions, a fuzzy set has a membership function $\mu_A(x)$ which represents the degree to which an element x belongs to some set A. If $\mu_A(x) = 1$ then x fully belongs to A, while if $\mu_A(x) = 0$, x does not belong to A. An intermediate value, such as $\mu_A(x) = 0.6$, indicates that x belongs to the set to some degree. Fuzzy sets are thus a precise tool for representing and manipulating imprecise notions.

Application of fuzzy set theory involves: (1) the translation of an imprecise natural language problem into a representation involving fuzzy sets; (2) the use of a calculus to transform the initial fuzzy sets into other fuzzy sets representing the required output variables of the analysis; and (3) reinterpretation of the results in imprecise natural language (see L.A. Zadeh, 1975). The first and last steps are crucial if the flavor of fuzzy set theory is to be fully captured. The core idea is to construct a calculus for the formal (i.e., *precise*) manipulation of imprecise concepts, which takes in imprecise inputs and puts out imprecise outputs.

Fuzzy set theory provides a powerful range of tools for (a) interpreting the meaning of natural language utterances, and (b) fuzzifying any existing logical or mathematical structure. The issues of concern to us in this discussion are, first, whether these tools can be used to construct a viable approach to evaluating evidential arguments, and second, aspects of reasoning relevant to intelligence analysis that are captured by fuzzy logic but not by other approaches.

Consider the following exchange:

Q: How large is the enemy installation?
A: It's pretty large. 300 personnel are assigned there.

In fuzzy set theory, the denotation of "large installation" could be represented by a fuzzy set membership function. Such a function gives the degrees of membership for different numbers of personnel in the fuzzy set "large installation," e.g., $\mu_{\text{large-installation}}$ (300 personnel) = .4; $\mu_{\text{large-installation}}$ (800 personnel) = .9; etc. The impact of the qualifier "pretty"

$$\begin{tikzcd} \tau(A)=1.0 \arrow[d] & \mu_A(x)=r \arrow[dl] \\ \pi(x)=r & \end{tikzcd}$$

$$\begin{tikzcd} \tau(A)=r & \mu_A(x)=r \arrow[l] \\ \left[\begin{matrix}\pi(x)=1.0\\ \pi(y\neq x)=0\end{matrix}\right] \arrow[u] & \end{tikzcd}$$

might be represented by a simple transformation of this function:

$$\mu_{\text{pretty-large-installation}}(\cdot)=[\mu_{\text{large-installation}}(\cdot)]^{1/2}.$$

So,

$$\mu_{\text{pretty-large-installation}}(300 \text{ personnel}) = .4^{1/2} = .63.$$

Two concepts, degree of truth and possibility, are at the core of Zadeh's theory of fuzzy reasoning. Both are defined in terms of fuzzy sets. If we know the installation has 300 personnel, then the degree of truth in the claim that it is large, τ(installation is large), equals $\mu_{\text{large-installation}}$(300 personnel)=.4. Conversely, suppose we do not know the exact number of personnel assigned to the installation, but we do know that it is large (e.g., through frequent interception of signals or sightings of supply activity). This proposition ("the installation is large") induces a possibility distribution on related, perhaps implicit, variables (in this case, the number of personnel). In other words, given that the installation is large, we can assign a degree of possibility to any given size. The degree of possibility is just the degree of membership of that size in the fuzzy set "large installation".

Fuzzy membership functions can be combined via operations which represent "fuzzifications" of the usual set-theoretic and logical operations (union, intersection, implication, etc.). For example, according to the most commonly used definitions of intersection and union,

$$\mu_{A\cup B}(x) = \max(\mu_A(x),\mu_B(x))$$

$$\mu_{A\cap B}(x) = \min(\mu_A(x),\mu_B(x))$$

In other words, the degree of membership of an object or value x in the union (intersection) of two fuzzy sets, A and B, is equal to the maximum (minimum) of x's degrees of membership in A and B, respectively.

In the evaluation of an evidentiary argument, implication may play a key role. Recall that in classical logic, a statement of the form $p \rightarrow q$ (if P, then q) is true as long as it is not the case that p is true and q is false (see Section 2.6.1 above). There are a variety of ways to fuzzify this relation. The essential intuition behind one popular approach (Gaines, 1977) is a direct generalization of the traditional definition: The degree of truth of the conclusion (q) must be *at least as great* as the degree of the truth of the antecedent (p), assuming the rule itself (p→q) is completely true. If the degree of truth of the rule is less than one, then to that degree the conclusion may be less true than the antecedent. If the degree of truth of the rule and the antecedent of the rule are characterized fuzzily rather than precisely, linear programs may be used to find the degree of truth of the conclusion.

In fuzzy set theory, propositions may involve a variety of different types of fuzzy hedges: truth-qualifications ("very true," "pretty true," etc.), possibility-qualifications ("possible," "almost impossible," "quite possible"; etc.), and probability-qualifications. The latter includes fuzzy probabilities like "quite probable" and "not very likely," in addition to fuzzy quantities such as "most", "usually," "several," "few" and "more than half". Hedges are themselves interpreted within the theory as fuzzy sets. Thus, truth-qualifications are fuzzy sets containing numbers between 0 and 1 to varying degrees. For example a degree of truth of .9 might belong to the set "very true" to the degree .8, while degree of truth .4 might belong only to degree .1. Thus, "it's very true that this is a large installation" induces a possibility distribution over the degrees of truth of "this is a large installation."

With these tools, a large variety of fuzzy reasoning processes [redacted] can be explicated: e.g.,

P1 It is very true that if an installation is large, it is dangerous.

P2 This installation is moderate in size.

P3 It is only slightly true that this installation is dangerous.

In this argument, both the datum, P2, and the warrant, P1, are imprecise. What is of special interest is that the datum (a moderately sized installation) and the antecedent of the rule (a large installation) do not exactly match. In fuzzy set theory, the applicability of the rule is not all-or-none; the strength of the conclusion is *diminished*, but not destroyed, as a function of this dissimilarity. The degree of truth of the antecedent is computed as the intersection of the two fuzzy sets, "large" and "moderately sized":

$$\mu_{\text{true}} \text{ (installation is large)} = \sup_{n} [\mu_{\text{large}} (n) \wedge \mu_{\text{moderately-large}}(n)]$$

where n ranges over number of personnel and "∧" is the min operation.

To what extent can fuzzy hedges and fuzzy implication be utilized to assess the strength of an evidentiary argument? Consider the following warrant (cf., Section 2.6.1):

> It is very true that if witness X says he saw the installation, it is there.

The hedge "very true" induces a possibility distribution over the truth value of the implication. But for any of these possible truth values to be correct, all that is required is that the antecedent not exceed the conclusion in truth value by more than some specified amount. Thus, "if the moon is made out of cheese, then the installation is there" would have degree of truth one, since the degree of truth of the antecedent is zero. Fuzzification of truth-functional implication thus provides no escape from the problems identified earlier with traditional logic. Assessments of degree of truth in this context do not capture an actual evidential link, and (to the degree that the antecedent is true) are reducible to direct assessments of the degree of truth of the conclusion.

An alternative approach within fuzzy logic is to use probability-qualifications rather than truth-qualifications. Probability hedges and fuzzy quantifiers (like "usually") are fuzzy sets whose members are proportions or relative frequencies, rather than degrees of truth. Thus, consider an alternative version of our warrant:

> Usually, when witness X says he saw an installation, it is there.

The hedge "usually" induces a possibility distribution over the proportion of cases in which, when X reported he saw an installation, the installation was there. Such hedges generalize traditional universal quantification ("all") in several ways: (1) fuzzy quantifiers make many more discriminations than the traditional "all" and "some" (e.g., "nearly all," "most," "a few," etc.); (2) quantifiers do not correspond to exact proportions, but include different proportions to different degrees; and (3) the proportions themselves may be fuzzy (i.e., cases in which witness X said he saw an airplane and it was or was not there, might count toward the proportion--through to a lesser degree than cases where an installation was reported). Thus, it might be possible to gauge the strength of an evidential link by selection of a fuzzy qualifier from the rough continuum ranging from "none" to "all."

This approach avoids some of the problems addressed in Section 2.6.1 in connection with universal quantification. While "Pr (for all x, if x is F then x is G)" is zero if there is a single x which is F and not G, "usually (F's are G's)" appears to describe a strong evidential link where most, though not all, F's are G's; "few (F's are G's)" describes a weaker link; etc. Nevertheless, we argue that this approach fails to capture adequately the notion of an evidential link. First, since probability-hedges are defined in terms of proportions or relative frequencies, there is no provision for propagating such measures to assess the strength of a particular, unique conclusion (is *this* reported installation actually there?). More fundamentally, such hedges (like universal quantification) miss the subjunctive (counterfactual) nature of evidential connections (Section 2.6.1 above). The point of the warrant is that if the installation were *not* there, witness X would not have reported it: there is a mechanism (good eyesight, adequate visibility, honesty, etc.) which ensures that the evidence "tracks" the hypothesis on this occasion. That reports from this witness have been true in the past is *evidence* for the existence of tracking now, but is not the same thing (imagine the witness has just gone blind).

A third problem with this approach is that fuzzy probabilities do not change in an appropriate way in the course of an evidential argument. An argument based on Bayesian updating (Section 2.1) can be fuzzified by providing fuzzy rather than exact likelihoods for the evidence given the hypotheses (e.g., this observation is "pretty likely" if H is true, but "very unlikely" if H is false). As we add items of evidence in this way, however, the degree of imprecision in the conclusion increases rather than decreases. Fuzzy probabilities, therefore, appear to capture imprecision in the *assessment* of probabilities, e.g., "measurement error" (Section 2.6.2 above), rather than unreliability or incompleteness of the evidential argument. The former increases as more sources of error are added; the latter decreases as new items of evidence are considered.

Still another approach to evidential evaluation within fuzzy logic involves the use of possibility-qualifications, rather than truth-qualifications or probability-qualifications. Possibility hedges (like "impossible" or "very possible") are fuzzy sets that take degrees of possibility as members. Thus, "it is very possible that the installation is there" induces a higher-order possibility distribution over the possibility that the installation is there. First-order possibility measures of this sort intuitively do capture a notion quite close to the strength of an evidential link: i.e., the degree to which the evidence fails to exclude a given hypothesis. (A correlative notion of the "necessity" of H can be defined as 1-Poss($\bar{H}$).) Moreover, it turns out that Shafer's plausibility, Pl(•) (Section 2.2 above), is a possibility measure in Zadeh's sense when the subsets to which belief is assigned are nested. (In fuzzy logic, there is a simple rule for deriving the possibility of the union of two sets from the possibilities of the sets:

$$\text{Poss}\ (A \cup B) \;=\; \text{Max}\ (\text{Pos}(A),\ \text{Pos}(B)).$$

This rule also applies to Shafer's Pl(•) for consonant, i.e., nested, belief functions.) However, consonant belief functions are only one special case of a belief function model. Nor is there any guarantee that the combination of consonant belief functions *via* Dempster's Rule will result in a new consonant belief function. Zadeh's Rule for combining evidence (i.e., for combining different possibility measures on the same hypotheses, based on independent evidence) is:

$$\text{Poss}_{1,2}(H) \;=\; \text{Min}\ (\text{Pos}_1(H),\ \text{Pos}_2(H)).$$

Shafer (1981) has pointed out this rule may be regarded as the strongest combination rule for consonant belief functions that guarantees a consonant belief function as the conclusion. Thus, although Zadeh's possibility measure can be seen in one sense as a generalization of belief functions (permitting fuzzy as well as precise specifications of Pl(•)) in another sense it is a restriction (applying only to consonant elements).

In sum, fuzzy logic is a highly flexible and versatile tool for handling imprecision, a concept of uncertainty not modeled well by the other theories we have discussed thus far. Unfortunately, the meaning of fuzzy measures is not always clear. In particular, there are neither behavioral specifications (as in the betting interpretations of Bayesian theory) nor canonical examples (as Shafer believes are important). The procedures for combining membership functions are not unique, and the justification for the ones Zadeh recommends is not clear. Moreover, no consistent method appears to have been developed for translating output membership functions back into linguistic expressions. Finally we have shown that efforts to use fuzzy logic to define an evaluation measure for the reliability of an evidential argument are inadequate, except for an approach (possibility measures) that is in critical respects a special case of Shaferian belief functions.

The major contribution of fuzzy set theory, we feel, may be the identification of a type of uncertainty not addressed at all in other frameworks. The core idea is that *similarity* plays a role in reasoning that cannot be equated with concepts like chance, reliability of an argument, or completeness of evidence. Consider the example above, involving the partial match between a rule antecedent ("large installation") and an observation (a "moderately sized installation"). In determining whether or not, or to what extent, the rule should be applied, the issue is *not* uncertainty about the facts (i.e., chance). We are not wondering about the *chance* that a moderately sized installation is, or is not, large. The problem is that there is no well-defined situation which would determine the outcome of a bet on this matter. Fuzzy set theory gives a plausible account of this problem: the denotations of "moderately sized" and "large" are neither identical nor disjoint, but overlap. Nor are we wondering about the exact size of the installation. Suppose we *know* its size: e.g., 200 personnel; there is *still* the question of the *degree* to which such an installation is "large." Efforts to translate such concerns into the language of probability theory will inevitably involve (a) arbitrary and *ad hoc* specifications of cutoffs (e.g., when an installation has more than 300 personnel, it abruptly becomes "large"); and/or (b) unnatural shifting of the problem (e.g., the chance than an English-speaker would *say* that an installation with x number of personnel was large).

Partial matching of current observations and stored knowledge or previous experience is, we are convinced, an inescapable aspect of any intelligent reasoning process. We would argue, however, that effective implementation of such a concept requires a viable notion of the reliability of an evidential argument. In short, the most common effect of an imperfect match between evidence and knowledge is to reduce the impact of the argument based on that evidence. The system to be described in Section 3.0 thus links dissimilarity of rule antecedent and evidence to discounting of an argument in a belief function framework.

3.0 A SELF-RECONCILING EVIDENTIAL DATABASE

3.1 Introduction to SED

In communicating conclusions based on incomplete, unreliable, and inconsistent data, the intelligence analyst faces a dilemma. If he provides an explicit account of divergent possible interpretations, confused policy-makers may object that he has hedged too much. If he reports only the uncontroversial elements of divergent views, he may be accused of being too bland. Nevertheless, if he takes a definite position, which turns out to be wrong, the consequences may be even worse.

Although increasing attention is being given to the problem of representing uncertainty in database systems, very little work has been done on methods for representing the evidential arguments that underlie assessments of uncertainty. Yet understanding and manipulating such arguments are essential to the intelligence analyst's tasks of making sense out of unreliable and inconsistent data and communicating conclusions. A key element in all these tasks, we contend, is the utilization by the analyst of qualitative and approximate causal models of how the available data could be linked to an as yet unknown "ground truth".

In this section we turn to the description of a system, the Self-Reconciling Evidential Database, which addresses these problems on three levels: (a) by providing a generic schema for an evidential argument based on the underlying causal chains that link conclusions and evidence; (b) by permitting the analyst to investigate different representations of the same argument through adoption and revision of assumptions; and (c) by embedding evidential arguments within a higher-level "metareasoning" process that responds to conflict between different lines of argument by tracing the assumptions involved in the conflict and recommending revisions.

ARGUMENT SCHEMAS

SED represents evidential arguments within a framework that combines elements from numerical models of uncertainty (Shafer/Dempster belief functions) and aspects of more qualitative approaches to reasoning (based on Toulmin and Paul Cohen). The latter provide a natural structure for representing the components of an argument and factors which affect its validity. We stress, however, the preeminent role of causal relationships among these factors - in agreement, incidentally, with recent philosophical analyses of knowledge in terms of explanatory connections linking evidence and conclusions (e.g., Harman, 1973; Nozick, 1981). Shafer's theory of belief provides an effective calculus for capturing these causal relationships. In addition, it provides a quantitative measure of the strength of an argument (in terms of uncommitted belief), methods for combining arguments, and, in an extension of the theory to be described here, tools for representing assumptions (as decisions regarding the allocation of uncommitted belief).

ASSUMPTION-BASED REASONING

The design of SED rests on the premise that there is no uniquely "true" problem representation at the level of probabilities or degrees of belief. A variety of valid models may exist which differ in the precision and/or divergence of their conclusions, as well as in the number and magnitude of the assumptions which they require. Assumptions may pertain to modeling (e.g., normality, independence, linearity) or they may pertain to substance (e.g., the credibility of a source, proper functioning of a data collection system, continued accuracy of a dated observation, absence of a conspiracy to deceive among apparently unrelated sources). Not all of these assumptions are made all the time; but unless some such assumptions are made, analytic arguments are condemned to perpetual inconclusiveness. SED permits the analyst to explore a space of problem representations by imposing or rejecting assumptions regarding the credibility and direction of an argument. Investigation of such a space may be utilized to develop the analyst's understanding of the problem, test the sensitivity of conclusions, and select a representation that suits the information requirements of intelligence consumers.

CONFLICT RESOLUTION

Standard numerical approaches to uncertainty regard conflict of evidence as stochastic; it is bound to occur by chance some portion of the time when sources of evidence are imperfectly correlated with hypotheses. Thus, methods for combining evidence (e.g., Bayes' Rule or Dempster's Rule) arrive at conclusions by a process akin to averaging, in which different bits of evidence are aggregated. SED encourages an alternative point of view: that conflict of evidence can often be regarded as a symptom of erroneous assumptions in the arguments that contributed to the conflict. SED, therefore, uses conflict as an opportunity to learn, i.e., to diagnose and remedy errors in reasoning, and thus to develop improved arguments.

The principal goal of this section is to explicate the concepts incorporated in SED. However, to keep the discussion from becoming too abstract, we shall do so in the context of a concrete application to an illustrative intelligence problem. Moreover, we will refer throughout to displays in a hypothetical computer implementation. It will be assumed that the user, an intelligence analyst, interacts with the implementation *via* a keyboard and a mouse and mouse button. The mathematical basis of the system is described in more detail in Appendices A and B.

3.2 A Hypothetical Intelligence Problem

For historical and geographical reasons, the Soviet Union is extremely concerned about the military capabilities of West Germany. In particular, a major concern of the Soviets is that West Germany should remain a non-nuclear state. The Soviets have recently expressed concern through diplomatic channels that the West Germans may be developing a nuclear weapon. The Soviet ambassador has cited reports, believed to have originated in East Germany, of "peculiar activity" around a spent fuel pond at a West German reactor. The President's advisors have informed him of the Soviets' concerns, and have recommended that he gather information on the matter and formulate a reply to the Soviets. As a consequence, the President has called in the director of the Central Intelligence Agency and asked him to prepare a briefing on the question of whether West Germany possesses a nuclear weapon.

The following items of evidence have been gathered for inclusion in the report.

1. West Germany has signed the Treaty on the Non-Proliferation of Nuclear Weapons (NPT) as a non-nuclear weapons state.

2. West German nuclear facilities are inspected by IAEA inspectors and West German nuclear material is under Agency safeguards. The International Atomic Energy Agency's (IAEA) Safeguards Implementation Report (SIR) stated that "materials under Agency safeguards in the year remained in peaceful nuclear activities."

3. The West German government has recently and repeatedly made public statements against nuclear proliferation.

4. The United States is committed under its NATO obligations to defend West Germany against attack.

5. The United States has not received any reports suggesting the existence of clandestine nuclear facilities or nuclear weapons tests in West Germany.

6. A former inspector at a West German nuclear facility has testified that his reports of anomalous material account imbalances at a West German facility were covered up by the IAEA.

7. The USSR, as noted above, has reported "unusual activity" in regard to a West German reactor.

8. West Germany has lodged complaints about the frequency of IAEA inspections.

9. West Germany has lodged complaints about the frequency of IAEA inspections.

3.3 The Structure of a Simple Evidential Argument

In SED, items of evidence support conclusions only *via* arguments. The analyst is thus encouraged to state the *reasons* why a given conclusion might follow from a particular piece of evidence--not simply a number measuring the degree to which the conclusion is associated with that evidence. Arguments in SED will typically have a natural causal structure. Reasons for (or against) a conclusion reflect the analyst's understanding of the sequence of events and processes that go between what the analyst wants to know and the available evidence. Analysts naturally reason in terms of causal chains of this sort in order to evaluate the reliability and significance of a piece of evidence.

The first step in an analysis is to structure the problem. In SED, this means (a) identifying the hypotheses about which the analyst is concerned, (b) identifying the available evidence and (c) specifying the *arguments* which link the evidence to the hypotheses. Figure 1 illustrates the result of this process for evidence item 2, the IAEA report certifying that West German nuclear material has remained in peaceful nuclear activities.

Figure 1: SED Display for IAEA Argument in Non–Numerical Mode

TO PIC – nuclear facility &
DIVERSION 3 AND [illegible]

GROUNDS Arg. #2: IAEA Report report name	ASSUME Y	ASSUME N	BELIEVE Y	BELIEVE N		HYPOTHESES No Diversion	HYPOTHESES Diversion for West German Bomb	HYPOTHESES Diversion to Other Country
CONCLUSION			v.d.					
BACKING & ASSUMPTIONS								
1. Concealment from inspectors						0		
		1.0	.3	.2		+		+
					I		+	+
2. Diversion opportunities during inspector absence						0		
		1.0				+		+
3. Falsification by inspectors						0		
		1.0				+	+	+
					I		+	+
4. Inspector incompetence						0		
		1.0				+	+	+
5. Hampering of inspectors						0		
		1.0				+		+
					I		+	+
6. Statistical inadequacy in analysis						0		
		1.0				+	+	+
7. Bias in interpretation of results						0		+
		1.0					+	0
8. Fixed response of "no diversion"						0		
		1.0				+	+	+

As shown in Figure 1, the analyst has identified three relevant hypotheses: nuclear material has not been diverted, nuclear material has been diverted for the purpose of building a West German bomb, and nuclear material has been diverted to another country which intends to build a bomb. A given item of evidence may support any *subset* of these elementary hypotheses. For example, the IAEA report indicating no anomalous activity supports the conclusion {no diversion}. However, the former inspector's report of a cover-up (evidence item 7) provides support for the subset of hypotheses {diversion for West German bomb, diversion to another country}, since it points to diversion, but fails to discriminate between the two purposes for which the diversion might have taken place. By contrast, the commitment of the U.S. to West Germany's defense (evidence item 4) supports the subset {no diversion, diversion to another country} since it disconfirms the claim that West Germany would build a bomb for itself, but does not exclude the possibility that West Germany is assisting some other country. The conclusions of an argument are thus subsets of hypotheses; and these may be large or small depending on the power of the evidence and the argument based upon it to discriminate among the possibilities. A totally inconclusive "argument" is maximally imprecise; it has as a "conclusion" the universal set, consisting of all three elementary hypotheses.

The star in Figure 1 under the hypothesis of no diversion indicates that the analyst intends to construct an argument, based on the IAEA inspection report, with the conclusion that no diversion has taken place. For the IAEA report to *prove* this conclusion, a variety of premises must be asserted. The analyst has enumerated these premises in the area labeled "Backing and Assumptions." These are the conditions which must be satisfied for the evidential link between the IAEA report and the conclusion of no diversion to be valid. They are here stated in the form of "rebuttals" to that linkage; i.e., *unless* any of these conditions holds, the IAEA report proves no diversion has taken place. The analyst's argument is this: *if* there has been no successful concealment from IAEA inspectors, no opportunities for diversion of material during the absence of an inspector, no intentional falsification of data by the inspectors, no errors due to inspector incompetence, no hampering of inspectors by nuclear plant personnel, no statistical error in the analysis of the data, and no bias in the interpretation of results, *then* the IAEA report proves that no diversion has taken place.

The analyst constructs such a list by asking himself what factors could disrupt the causal relationships that ordinarily link the truth of different hypotheses with different types of evidence. For example, a typical causal chain would be:

No Diversion → Inspectors Perform Observations → Inspectors Make Reports → Data Analyzed → Conclusions Formulated → Report No Diversion

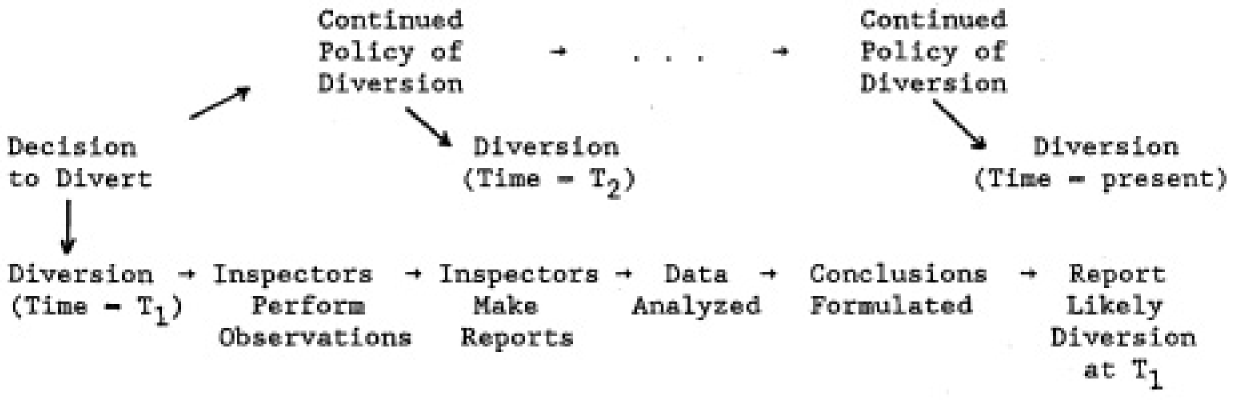
Similarly,

Diversion → Inspectors Perform Observations → Inspectors Make Reports → Data Analyzed → Conclusions Formulated → Report Likely Diversion

It is by virtue of these causal relationships that the evidence (the IAEA report) covaries with the truth about the hypotheses regarding diversion (cf., Nozick, 1981). Each of the premises in the core argument is concerned with a different way that one of the links in these chains could break.

Evidential arguments may be classified in terms of the *type* of causal connection by means of which they link conclusions and evidence. Such linkages may, in principle, involve arbitrarily large and complex networks of causes and effects. In any case, they will not always be as direct or as simple as in the IAEA example; yet, they appear to fall into a small set of characteristic, recognizable patterns.

The IAEA example involves evidence linked by a direct causal chain to a past set of events (i.e., incidences of diversion). A slightly more complex pattern occurs when the evidence and the event of interest are linked only by virtue of sharing a common cause. For example, suppose the question before the analyst is whether or not nuclear material is currently being diverted, and the most recent IAEA report pertains to inspections conducted a year in the past. A causal chain of the following sort links the IAEA report of likely diversion at time T_1 to the hypothesis of diversion in the present:


Continued Policy of Diversion
. . .
Continued Policy of Diversion
Decision to Divert
Diversion (Time = T_2)
Diversion (Time = present)
Diversion (Time = T_1)
Inspectors Perform Observations
Inspectors Make Reports
Data Analyzed
Conclusions Formulated
Report Likely Diversion at T_1

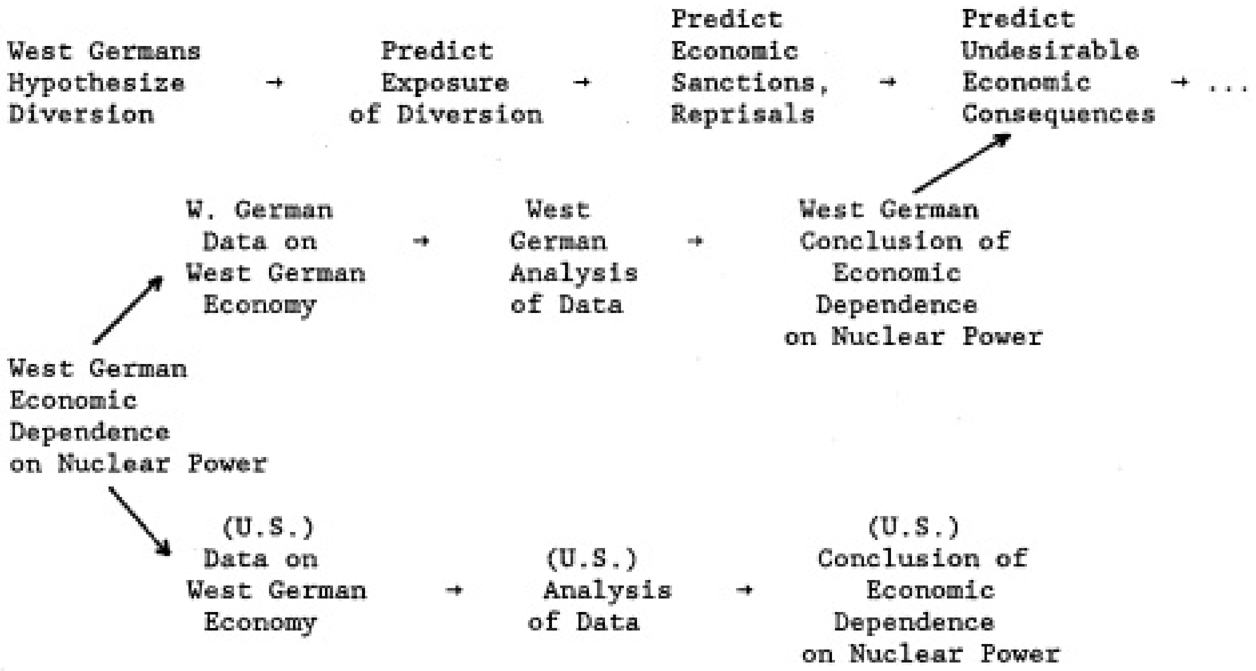
According to this argument, two causal sequences are operative, with a common origin in an earlier West German decision to divert. One is identical to the sequence discussed above, in which the significance of the IAEA report for past incidences of diversion is evaluated; the other addresses the causal continuity of diversion policy and practice from that time to the present. An argument based on this causal structure would thus require additional premises to ensue that the second causal sequence is unbroken, i.e., to exclude factors or events during the past year that could have caused West German policy toward diversion to change (e.g., change in political leadership, change in perceived defense requirements).

Another pattern of causal influence arises when the significance of an item of evidence depends in part on an analysis of the motivations and beliefs of other people. In such cases, the causal model that the analyst works with may include a "second-order" model that depicts a chain of reasoning *about* causal relationships in the heads of those other people. Making such second-order models explicit and judging the evidence or lack of evidence for their components may be an effective antidote to the danger of "mirror-imaging" (i.e., supposing that others act on the same beliefs and values that we do). For example, West Germany's economic dependence on nuclear power (evidence item 6) leads to a conclusion of {no diversion} via a chain of "mental events" (attributed conjectually to the West Germans):

West Germans Hypothesize Diversion → Predict Exposure of Diversion → Predict Economic Sanctions, Reprisals → Predict Undesirable Economic Consequences → Conclude Benefits of Diversion Do Not Outweigh Costs

The validity of the argument based on evidence item 6 depends on premises that ensure that these links remain unbroken: that the West Germans believe exposure of a diversion would be likely, that they believe economic sanctions or reprisals would be adopted by other countries, that they believe undesirable economic consequences would in fact follow, and that they find these consequences unacceptable even in light of possible benefits from diversion.

This chain, however, does not yet form a link between the analyst's belief that West Germany is economically dependent on nuclear power and the conclusion of (no diversion). Thus, it does not yet capture all the factors that bear on the validity of this inference. To do so, it must be elaborated in two ways. First, it must provide a more detailed account of the West German decision making process, linking it to West Germany's level of economic dependence on nuclear power. Note that (if this argument is valid) the *analyst's* beliefs in regard to West Germany's economic dependence and the *West Germans'* beliefs regarding their own economic dependence share a common cause: the ground truth economic situation; and the West Germans' beliefs on this matter will influence their prediction of the economic impact of sanctions and reprisals. Thus, we get:



West Germans Hypothesize Diversion → Predict Exposure of Diversion → Predict Economic Sanctions, Reprisals → Predict Undesirable Economic Consequences → ...

W. German Data on West German Economy → West German Analysis of Data → West German Conclusion of Economic Dependence on Nuclear Power

West German Economic Dependence on Nuclear Power

(U.S.) Data on West German Economy → (U.S.) Analysis of Data → (U.S.) Conclusion of Economic Dependence on Nuclear Power

The bottom branch of this network reflects the analyst's own reasoning (or reasoning which he accepts) regarding West Germany's economic dependence on nuclear power; the middle branch reflects hypothesized reasoning by West Germans on the same subject. Adding these two branches makes clear that the validity of the argument based on evidence item 6 depends on the correctness of the analyst's inference that West Germany is dependent on nuclear power, and on the awareness of the West Germans themselves that this is true.

Finally, these mental events can be embedded in a first-order causal model:

West Germans Propose Diversion → Initiate Policy Debate About Diversion → [...reasoning as above...] → Implement Results of Policy Debate → No Diversion

bringing out the dependence of the argument on a crucial additional set of premises: that the decision on diversion would in fact be preceded by and based on some sort of high-level policy debate (and not made, for example, by lower level, independently acting personnel).

It is worth pointing out that, in interesting problems at least, no conclusion is really "proven", and any list of premises must always be regarded as provisional and incomplete. The analyst can choose the level of detail to which he carries the examination of any particular argument, and important contingencies may initially be overlooked or dismissed as improbable (e.g., could the IAEA report have been mistranslated?). Nevertheless, it is the effort to construct casually based arguments of this sort which clarifies the basis for belief, leads to the discovery of hidden premises or assumptions, and provides a framework for revision of beliefs and assumptions when they lead to conflict.

3.4 Beliefs and Assumptions

The analyst may have evidence or knowledge regarding some of the premises of an argument, but it is unlikely that he will have evidence or knowledge regarding all of them. Premises about which the analyst is ignorant may nevertheless play a crucial role both in his understanding of and reasoning about the problem, and in decisions regarding the collection of further information. To the degree that knowledge is lacking, therefore, SED permits the analyst to make assumptions and to explore the implications of those assumptions for the conclusions of the argument.

SED may operate either in a non-numerical mode or in a numerical mode. In the first, belief in a premise is all or none. If neither a premise nor its negation is believed, however, the premise (or its negation) may be *assumed*.
In the numerical mode, on the other hand, belief is graded, i.e., belief in a premise may fall anywhere in the interval between 0 and 1. To the extent that belief in a premise and its negation fail to sum to 1, the analyst may make a decision regarding the allocation of the remaining uncommitted belief. Such a decision constitutes an *assumption*. Note that the non-numerical mode is a special case of the numerical mode, in which all belief assignments and assumptions are either 0 or 1. However, use of the non-numerical mode (a) may be a source of valuable insight into the logical structure of an argument, and (b) may be preferred by analysts in problems where there is a high degree of confidence, or when, for whatever reason, they do not desire to use numerical

measures.

Our exposition will begin with the non-numerical mode, for reasons akin to (a). Exploration of the basic logical forms that arguments take will clarify the functions that numbers perform and their meaning. Numbers will be introduced quite naturally as quantifications of the validity of the logical structures created in the non-numerical mode. Numbers represent the *chance* that (non-numerical) arguments are valid.

An analyst may input both his beliefs and his choice of assumptions in the context of the display shown in Figure 1. In the non-numerical mode, the analyst indicates whether he believes a proposition to be false or true by pointing with the mouse at "Believe No" or "Believe Yes" respectively and clicking. If neither is believed with confidence, he may indicate his assumption that the proposition is false or true by pointing with the mouse at "Assume No" or "Assume Yes" respectively and clicking.

Note that the premises in an argument can be expressed in either a positive or a negative manner, depending on the preference of the analyst. Thus, the following two formulations are equivalent to one another:

Statistical inadequacy

Believe No	0	Believe Yes	0
Assume No	1	Assume Yes	0

No statistical inadequacy

Believe No	0	Believe Yes	0
Assume No	0	Assume Yes	1

Both formulations (in the context of the argument in Figure 1) represent the premise that statistical techniques are adequate; in both cases, the premise is assumed to be true. We referred to "Statistical inadequacy" (in the first formulation) as a "rebuttal," since acceptance of statistical inadequacy would invalidate the argument.

3.5 Many Arguments From One

In constructing a core argument, the analyst specifies a conclusion together with a set of premises, i.e., a set of propositions (such as "Statistical inadequacy") with associated credibility status (Believe Yes, Believe No, Assume Yes, or Assume No). These premises are then regarded by SED as a *core argument* which *proves* the specified conclusion.

In addition to the core argument, SED requests that the analyst specify the impact on the conclusion of *changing* the credibility status of each premise. This provides an economical, implicit specification of a large number of *alternative* arguments, each corresponding to a different change (or combination of changes) in the premises. In other words, given the core argument and the impact of changing any premise, SED can automatically construct a variety of other arguments based on the same item of evidence, but leading to different conclusions. To do so, SED utilizes presuppositions about the function of each premise in the causal chain linking conclusions and evidence. This capability enables the analyst to explore freely the implications of changes in his beliefs and assumptions; in addition, it permits SED to make recommendations regarding revision of assumptions and beliefs, and the collection of additional information, to resolve conflict.

3.5.1 The impact of negating a premise. SED provides a simple method of encoding the potential impact of changing a belief or assumption. The impact of rejecting a premise (in the non-numerical mode) is always: (i) to eliminate support for one subset of hypotheses, and (ii) to shift that support onto some other subset of hypotheses. Such impact may be of two kinds, depending on whether the effect is on the precision or the direction of the core argument.

Impact on precision. Rejection of a premise may remove, dilute, or enhance the bearing of the evidence on the hypotheses. Instances of the first type occur, for example, if IAEA inspectors are incompetent observers, if the sample size or statistical techniques applied to the data are inadequate to detect diversions, or if the IAEA always reports "no diversion" even if there is evidence to the contrary. In none of these cases does it follow that a diversion of nuclear material *did* take place. Rather, the IAEA report (evidence item 2) is deprived of evidential significance; the subset of hypotheses which it now proves is the universal set, i.e., the set consisting of all three elementary hypotheses. In short, the new argument tells us nothing regarding whether or not a diversion took place. This type of impact is indicated in the following way:

No Diversion	Diversion for West German Bomb	Diversion to Other Country
0		
+	+	+

The first row represents the subset of hypotheses which loses support if the premise turns out to be false. The "0" means that support for the subset (no diversion) would be invalidated (reduced to 0% of its former value) if this rebuttal turned out to be true. The second row indicates the subset to which support is shifted. In this case, the belief deducted from (no diversion) is to be added to the subset consisting of all three hypotheses.

Secondly, rejection of a premise may decrease the precision of an argument without totally invalidating it. For simplicity of exposition and display, we have recognized only two possibilities with regard to each premise: acceptance or rejection. It may be desirable, however, to provide a more flexible representation of the negation of a premise. For example, we may wish to distinguish several levels of competence of IAEA inspectors: ability to detect small amounts of diverted material, ability to detect large amounts of diverted material, and inability to detect even quite large quantities of diverted material. The core argument assumes the first (high inspector competence); negation of that premise, however, may lead to adoption of either of the other possibilities (moderate competence, low competence). Suppose further that building a West German bomb would entail diversion of large quantities of material, while aiding another country might involve diversion of significantly smaller amounts (since that country might have other sources of material in addition to West Germany). If then, the analyst were to reject the premise that IAEA inspectors are highly competent observers, and adopt instead the idea that they are moderately competent, support would be shifted from {no diversion} to {no diversion, diversion to other country}. This impact is represented as follows:

No Diversion	Diversion for West German Bomb	Diversion to Other Country
0		
+		+

The new argument (based on moderate competence of the inspectors) still has some force: it excludes the possibility of diversion for building a West German bomb, although it no longer excludes diversion to another country.

Finally, rejection of a precision premise can enhance the force of an argument. In the Figure 1 core argument, the analyst chose to make assumptions that maximized the precision of the conclusion (viz., he assumed that IAEA inspectors were competent observers, that statistical methods were adequate, and that the IAEA reported results appropriately). Assumptions of this sort are made constantly and are often necessary if anything at all is to be concluded from the data. Nevertheless, on other occasions the analyst may wish to adopt worst case assumptions, at least initially. For example, information from a new, and as yet untested, source may be treated with suspicion until the source has proven to be reliable; or the analyst may wish to see how far he can get without utilizing certain data sources at all. Subsequent rejection of these assumptions will increase, rather than decrease, the precision of the conclusion.

For example, the analyst might construct a core argument with the following premise: IAEA statistical methods can only discriminate large amounts of diverted material. The conclusion of the new core argument, is {no diversion, diversion to other country}. The impact of rejecting this premise, and adopting the view that IAEA statistical methods are highly discriminating, increases the precision of the conclusion by shifting support to {no diversion}. The analyst may represent such impact as follows:

No Diversion	Diversion for West German Bomb	Diversion to Other Country
0		0
+		

In sum, SED permits core arguments to be constructed at any initial level of precision. Rejection of a premise may then change the precision of a conclusion by shifting support to any "nested" subset, i.e., a subset which contains, or is contained by, the originally supported subset.

Impact on direction. The second type of impact which negation of a premise may have is to change the direction of an argument. In this case, support is shifted from one subset of hypotheses to a subset which is, at least partly, inconsistent with the original subset. Such a change in direction might occur when the negation of a premise means that an observer, sensor, or a mode of analysis has a built-in bias.

For example, the validity of the argument based on the IAEA report does not depend merely on the ability of the inspection and analysis processes to detect diversions of material. It also depends on the correct interpretation and non-deceptive reporting of what is detected. Thus, the analyst includes as a premise in the core argument, that IAEA personnel are not systematically biasing results in the direction of no diversion. If this premise were known or assumed to be false, then the IAEA report might say "no diversion" when the relevant data indicated that small amounts of diversion had taken place; and it might say that small quantities of material had been diverted when the data indicated that diversion had been substantial. This impact of negating the premise can be represented as follows:

No Diversion	Diversion for West German Bomb	Diversion to Other Country
0		+
	+	0

The negation of this premise does not constitute an *independent* argument in favor of (diversion for West German bomb) or (diversion to other country); rather, it shifts the conclusion in a way which is *conditional* on the initial conclusion. Built-in biases thus depend on the core argument for their effect: they "deflect" it in the direction of a new conclusion. Note that the newly supported subset neither contains nor is contained by the originally supported subset. The impact of negating the premise is thus on the direction of the argument, not its precision. Automated sensors and analytical or statistical techniques, as well as human sources, may produce systematically erroneous conclusions, hence, are susceptible to biases of this type.

Independent biases. In other cases, by contrast, the negation of a premise - in addition to its effect on the precision or direction of the core argument - is in itself an argument for a different conclusion. This new conclusion is independent of the conclusion that otherwise would have followed from the core argument. For example (Figure 1), suppose it is learned that there was deliberate concealment of certain nuclear plant activities from the IAEA inspectors, providing opportunity for diversion of small quantities of material. In that case, the core argument based on the IAEA report is rendered less precise, i.e., support is shifted from {no diversion} to the subset {no diversion, diversion to other country}. An additional result, however, is a new argument which independently supports the claim that diversion did take place, since diversion is a highly plausible explanation of the efforts at concealment. The double impact of rejecting the premise of no concealment is represented in the following way:

	No Diversion	Diversion for West German Bomb	Diversion to Other Country
	0		
	+		+
I		+	+

The revised core argument now points to {no diversion, diversion to other country}; and a separate argument (with a single premise: that concealment occurred) is created which points to {diversion for West German bomb, diversion to other country}. The "I" associated with the third row means that the new argument for diversion based on concealment is independent, hence, does not depend on the conclusion of the core argument.

The negation of a premise in an argument may sometimes provide independent support for an alternative conclusion only in conjunction with some additional preconditions. For example, deliberate concealment of nuclear plant activities proves diversion only if there is no other motive for concealment, e.g., protecting secret technology, or reducing disruption of plant routine caused by the presence of inspectors. If the analyst wishes to make these preconditions explicit, he can do so in SED by creating an entire new argument, instead of using the single line abbreviation (with "I") as just illustrated. The conclusion of this new argument would be (diversion for West German bomb, diversion to other country); its premises are that deliberate concealment has taken place, that there is no motive to protect secret technology, and so forth.

3.5.2 Combination of information within an argument. What if more than one premise is rejected, and the impact of negating each premise is support for a different subset of hypotheses? For example, the analyst may obtain evidence *both* that there was statistical error *and* that activities were concealed from IAEA inspectors. SED's ability to construct alternative arguments efficiently and automatically depends on its ability to exploit insights about the structure of reasoning in problems of this kind. In principle, the number of alternative arguments can grow exponentially with the number of premises (since there are 2^n combinations of acceptance and rejection of n premises). The burden on the analyst of assessing a separate conclusion for all those combinations would be inordinate. Typically, however, there is an implicit pattern of independence and dependence among premises and their negations which considerably simplifies the assessment task. These patterns are based on plausible assumptions about the causal chains that link evidence and hypotheses, and the manner in which they are replaced or modified through the negation of different types of premises.

If a core argument has n premises, then the analyst using SED need only specify n + 1 links between evidence and conclusions. The first link is the core argument, which associates the truth of all the premises with the conclusion of the core argument:

$$\text{If } p_1\ p_2\ \&\ \ldots\ p_n, \text{ then } C_o$$

where the premises p_i are propositions of any logical form (including negations) and C_o is a subset of hypotheses; the truth of all the premises implies that the truth about the hypotheses of interest is contained in the subset C_o. (Later, we shall generalize this formulation to include degrees of belief in different subsets ~~within C_o.)~~ The other links associate the negation of each premise with an operation which substitutes a specified subset S_i in the core conclusion by a specified alternative conclusion C_i:

$$\text{If } \sim p_1 \text{ then } S_1 \rightarrow C_1$$

$$\text{If } \sim p_2 \text{ then } S_2 \rightarrow C_2$$

$$\vdots$$

$$\text{If } \sim p_n \text{ then } S_n \rightarrow C_n$$

Since any combination of premises could be negated simultaneously, $2^n - 1$ possible alternative arguments are implied. The conclusion of any particular argument is computed automatically by SED according to a relatively simple set of rules. The negation of a premise has a different effect depending on the type of the premise, i.e., on whether it concerns precision, built-in bias, or independent bias.

Precision premises. The simplest case involves an argument in which only precision premises have been negated and all premises of the operate at same stage in the causal chain. A causal chain linking different kinds of evidence and different subsets of hypotheses will not produce "tracking" of hypotheses by evidence if any link in that chain is broken. Negating a precision premise means that the discrimination addressed by the negated premise cannot be made regardless of the strength of the rest of the argument. When more than one precision premise is negated, the new conclusion is the enlarged subset of hypotheses that includes all the conclusions of the different negated premises, i.e., their set-theoretic union.

For example, suppose there is evidence that inspectors are moderately incompetent (i.e., are unable to discriminate small quantities of diverted material that might be associated with aiding another country to build a bomb); and also evidence that statistical techniques are seriously flawed (i.e., unable to discriminate even large quantities of diverted material). Negation of the first premise means the argument can no longer discriminate (no diversion) from (diversion to another country); negation of the second premise means the argument can no longer discriminate (no diversion) from either (diversion to another country) or (diversion for West German bomb). Negation of *both* premises results in support for the universal set, which includes all these possibilities.

More formally, if p_2, p_4, and p_5 are precision premises that are negated, with associated substitutions $S_2 \rightarrow C_2$, $S_4 \rightarrow C_4$, and $S_5 \rightarrow C_5$, the new conclusion is the result of applying each substitution to the original core conclusion C_0 and taking the set-theoretic union:

$$\text{If } \sim p_2 \,\&\, \sim p_4 \,\&\, \sim p_5 \text{ and } S_2, S_4, S_5 \subset C_0, \text{ then } C_0 \rightarrow C_0 \cup C_2 \cup C_4 \cup C_5.$$

Built-In Bias Premises. Matters are not quite so simple if built-in bias premises are also negated, resulting in changes of direction in the argument. In that case, it is necessary for the analyst to give somewhat more careful consideration to the causal model which underlies the argument.

The simplest case of a causal chain linking conclusion and evidence consists, at least conceptually, of two "stages": a first stage in which data inputs are processed, discriminated, and categorized, and a second stage in which interpretations are associated with the results of the first stage and explicit outputs generated. More complex chains can be built up by repeated sequences of this sort: e.g., IAEA inspectors (1) make observations at nuclear facilities and (2) generate reports; then IAEA analysts (1') process and analyze the inspector's reports and (2') generate conclusions. This partitioning of events is, of course, somewhat arbitrary. Nevertheless, it provides a simple framework for understanding the relative roles of precision and built-in bias premises and how they should be combined.

Precision premises concern stage 1, and built-in bias premises concern stage 2 of an evidential causal chain. SED deals with multiple negated premises by working *backward* within a chain, taking each stage 1 - stage 2 segment in turn and first removing the effects of any built-in biases (stage 2), then computing the impact of prior failures to discriminate (stage 1). The result is a

O: "to" subset
F = member of "from" subset

O

F

simple process of search for the possible ground truth situations that could have given rise to the observed evidence.

For example, suppose that three premises have been negated in an expanded variant of the IAEA example: a precision premise regarding statistical error and two bias premises regarding the process of reporting conclusions from the statistical analysis. Suppose the impact of negating one bias premise b_1, is to shift support from {no diversion} to {diversion to other country}:

No Diversion	Diversion for West German Bomb	Diversion to other Country
0		+

b_1: Diversion to other country → No diversion

Looking at this in causal terms, the analyst knows or wishes to assume that IAEA personnel, confronted with results that in fact prove {diversion to other country}, will report {no diversion}. Similarly, the impact of negating the other bias premise b_2 is to shift support from {no diversion} to {diversion for West German bomb}:

No Diversion	Diversion for West German Bomb	Diversion to other Country
0	+	

b_2: Diversion for WG bomb → No diversion

Here, the analyst knows or assumes that if IAEA personnel are confronted with data that in fact prove {diversion for West German bomb} they will report {no diversion}. Finally, suppose negation of the precision premise p_1 shifts support from {diversion to other country} to {no diversion, diversion to other country}, and from {diversion for West German bomb} to {diversion for West German bomb, diversion to other country}:

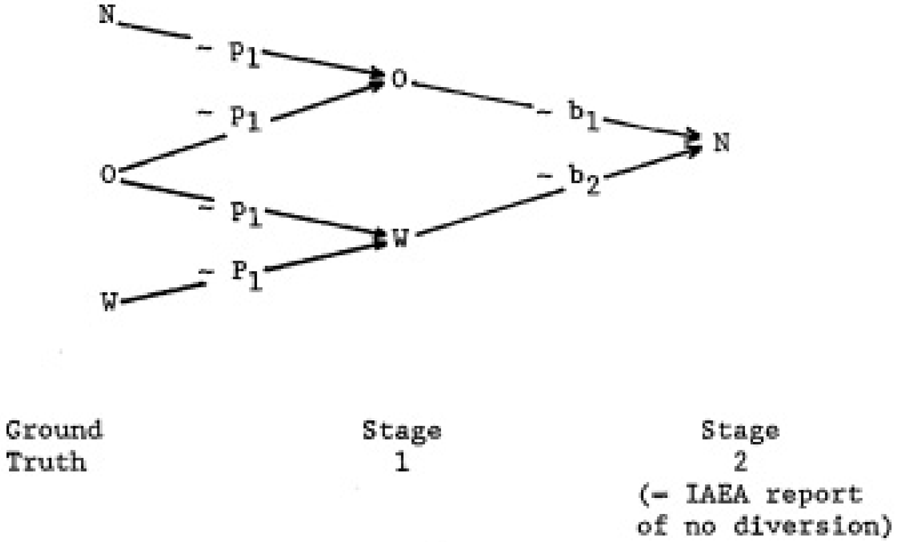
No Diversion	Diversion for West German Bomb	Diversion to other Country	
+		0 +	1
	0 +	+ 1	2

2

p_1: diversion for WG bomb → div for WG bomb
No diversion → diversion to other country
diversion to other country → div for WG bomb
diversion to other country → diversion to other country

In causal terms, the analyst is supposing that data which in fact prove {no diversion}, as well as data which in fact prove {diversion to other country}, could lead to a stage 1 IAEA conclusion of {diversion to other country}; this conclusion at stage 1, therefore, fails to discriminate between the two situations that could produce it. Similarly, data which in fact prove {diversion to other country}, as well as data that prove {diversion for West German bomb}, could lead to a stage 1 IAEA conclusion of {diversion for West German bomb}.

What is the appropriate conclusion from the new argument resulting from the negation of these three premises? Another way of asking this question is, what are the possible "ground truth" situations that could have led to the IAEA report of no diversion? The possible causal sequences implied by the above negated premises are outlined as follows, where N = {no diversion}, O = {diversion to other country}, and W = {diversion for West German bomb}:



Note that the arrows represent the direction of causality, while the inferencing process works in the reverse direction, moving from the "N" on the far right (i.e., the IAEA report of {no diversion}) to the possible original ground truth situations on the far left.

The negation of b_1 tells us that the report of no diversion could have arisen from results that prove {diversion to other country}; the negation of b_2 tells us that the report could also have arisen from results that prove {diversion for West German bomb). Correction for these two biases thus produces the set-theoretic union, {diversion for West German bomb, diversion to other country}. Finally, negation of p_1 tells us that {diversion to other country} might have arisen from data that prove either {diversion to other country} or {no diversion}, and {diversion for West German bomb} might have arisen from data that prove either {diversion for West German bomb} or {diversion to other country}. Taking the set-theoretic union of these possibilities, we get the universal set, {no diversion, diversion for West German bomb, diversion to other country}. The negation of these particular premises leaves an argument that tells us nothing about ground truth.

Arbitrarily complex causal chains can be handled by precisely the same principles. For example, the evidence and an event to be inferred or predicted are often related to one another only by virtue of being causally related to some common third event. In assessing the possibility of current diversion based on a dated IAEA report of diversion in the past, both the evidence and the current diversion have their causal origin in a hypothetical past decision to divert material (Section 3.3). In that case, the search process begins at the evidence and works backward to the cause (e.g., a past decision to divert), just as in the IAEA example. It then works forward from that cause along the other causal chain to the event to be inferred (e.g., the present diversion). At each step, it transforms the current conclusion in accordance with any negated premises pertaining to that step, and takes the union of the result. The final conclusion of the argument is the set of all possible ground truth situations (in regard to the event of interest) that are causally consistent with the evidence.

Causal chains underlying evidential arguments may also involve parallel causal streams which meet. For example, in the argument based on West Germany's economic dependence (Section 3.3), The West German prediction of undesirable economic consequences depends simultaneously on (a) the causal chain in which the West Germans infer that they are economically dependent on nuclear power, and (b) the causal chain in which they initiate a decision-making process, predict exposure to diversion, and predict economic sanctions in response to such exposure. Rejection of a premise in *either* chain may lead to a final conclusion other than {no diversion}. Here the inference process operates on the two causal steams in parallel, and takes the union of possible conclusions at the point where they meet.

Built-in bias premises impose a somewhat greater assessment task on the analyst than an argument involving only precision premises. First, if multiple stage 1-stage 2 segments are included in a causal chain, the analyst must provide a crude temporal ordering of the segments. Secondly, if more complex causal structures are involved, the analyst must indicate the location of each premise within such a structure. Finally, the analyst must assess a larger number of potential impacts of negating a given premise. Since correction of built-in biases changes the direction of the argument, the analyst must address the impact of negating premises on potential new conclusions of this sort, as well as on the original core conclusion.

Independent bias premises. The negation of independent bias premises is a relatively simple case, since each negated independent bias premise represents a new causal chain linking a specified conclusion to evidence (in addition to any effects it has on the precision or direction of the core argument). If more than one independent bias premise is negated, each of the associated independent causal chains is valid. Since the origin of each chain must be the same ground truth situation (albeit processed and interpreted in very different ways), the new conclusion is the set-theoretic intersection or common element (if any) of the conclusions associated with the different negated premises.

For example, suppose there is evidence that activities at nuclear facilities were concealed from some inspectors and also evidence that other inspectors falsified their reports. The first finding provides an independent argument for {diversion for West German bomb, diversion to other country}. Suppose the evidence suggests that only inspectors from country X falsified reports. Then the second finding is an independent argument for {diversion to other country}, i.e., diversion to country X. The only way that both new arguments can be valid, and both conclusions true, is for the truth to lie in their common element, {diversion to other country}.

More formally, if three independent bias premises were negated, p_1, p_3, and

p_7, with C_1, C_3, and C_7 as their associated conclusions, the new conclusion is their intersection:

$$\text{If } \sim p_1 \ \& \sim p_3 \ \& \sim p_7 \text{ then } C_1 \cap C_3 \cap C_7.$$

3.6 Combining Evidence From Different Arguments

The display shown in Figure 2 helps the analyst appraise the significance of the *combination* of all 9 arguments which bear on the hypotheses of interest. Each argument represents an independent causal claim linking its evidence to a subset of hypotheses, since each argument is rooted in the same ground truth situation. The conclusion of the combined argument is simply the intersection, or the common component (if any), of the subsets of hypotheses that are supported by the individual items of evidence. For example, if argument 1 means that the truth is in the subset {no diversion}, and argument 4 means the truth is in the subset {no diversion, diversion to other country}, then if both arguments are valid, the truth must be in the intersection, {no diversion}.

Under "Conclusions," SED displays all the conclusions that logically *might* be supported by the available items of evidence. These possible conclusions include all the intersections of all the possible argument conclusions - under all combinations of beliefs or assumptions regarding the premises of the arguments. Thus, in this example, the possible conclusions from the given collection of evidence are: {no diversion}, {no diversion, diversion to other country} = not building a bomb, {diversion for West German bomb}, {diversion to other country}, and {diversion for West German bomb, diversion to other country} = diversion.

Any conclusion that was *actually* supported by this combination of arguments would be shown with belief = 1.0. In this example, of course, the intersection of the supported subsets of hypotheses is empty, since the items of evidence support conflicting, or non-overlapping, subsets. This is shown by the 0's corresponding to belief in the possible conclusions of the combined argument.

Figure 2: SED Display for Combining Multiple Arguments in Non—Numerical Mode

VIEW			BELIEVE	HYPOTHESES		
				No Diversion	Diversion for West German Bomb	Diversion to Other Country
	CONCLUSIONS					
	PRO	CON				
— — — —	{1,2,3,6}	7,8,9	0	*		
— — — —	8	1,2,3,4,5,6	0		*	
— — — —	{4,5} {7,9}	1,2,3,6	0			*
— — — —	{4,5}	7,8,9	0	*		*
— — — —	{7,9}	1,2,3,6	0		*	*
	ARGUMENTS					
— — — —	1. Signed treaty		1.0	*		
— — — —	2. IAEA report		1.0	*		
— — — —	3. Public statements		1.0	*		
— — — —	4. US/NATO commitment		1.0	*		*
— — — —	5. No detection of facilities		1.0	*		*
— — — —	6. Economic dependence		1.0	*		
— — — —	7. Coverup report		1.0		*	*
— — — —	8. USSR statement		1.0		*	
— — — —	9. West German complaints		1.0		*	*
	CONFLICTS					
	●	○				
— — — —	{1,2,3,6}	{7,8,9}		●	○	○
— — — —	{4,5}	8		●	○	●
— — — —	{1,2,3,6}	8		●	○	

Figure 2 in conjunction with the detailed displays for each separate argument (such as Figure 1) functions as an "argument spreadsheet": changes in beliefs or assumptions in any of the arguments are propagated through the conclusions of those arguments to the conclusion of the combined argument. The analyst can retrieve any of the individual argument displays by selecting and clicking with the mouse on the appropriate line in the "VIEW" column. Alternatively, he can make changes directly to the conclusions of an argument in Figure 2, to examine the consequences of those changes for the combined conclusion.

3.7 Conflict of Evidence

Conflict in the non-numerical mode means that the user has contradictory beliefs or assumptions. SED helps him resolve the conflict by examining the assumptions and beliefs that underlie the conflict, revising them where appropriate, or collecting further information. Support for this process of assumption revision and testing is provided in two ways, corresponding to two rather different cognitive strategies:

Global Conflict Resolution. The analyst may hypothesize a candidate solution and assess its overall acceptability, then go on, if he wishes, to hypothesize another candidate solution, etc. In this case, SED helps the analyst work backward from the hypothesized conclusion, asking what total consistent environment of beliefs and assumptions would make that conclusion true, and what changes in his or her beliefs would be required to establish such an environment.

Local Conflict Resolution. On the other hand, SED may help the analyst decompose the conflict into components, look directly at each source of conflict, ask what changes in his or her assumptions or beliefs would eliminate that particular problem, and then go on to the next source of conflict.

3.7.1 Construction of consistent plausible environments. In order to establish a particular conclusion C, one or both of two actions may be necessary: (1) if no arguments currently support conclusion C, at least one argument will have to be revised by changing assumptions or beliefs so that it supports C;

(2) if one or more arguments currently have conclusions that are inconsistent with conclusion C, each of them will have to be revised so that they no longer conflict with C.

Associated with each possible conclusion in Figure 2, SED displays a "ledger" of reasons pro and con. The numbers refer to currently valid *arguments*; brackets represent an optional selection (or disjunction) of such arguments. In order to establish a given conclusion with no conflict, there must be at least one valid argument in the PRO column and no valid arguments in the CON column. Thus, the conclusion {no diversion} would be justified if any one of the arguments in the PRO column remain valid, i.e., argument 1 *or* argument 2 *or* argument 3 *or* argument 6; and if *all* the arguments in the CON column (i.e., 7, 8, and 9) become invalid. The conclusion that a diversion to another country has taken place would be justified if *either* argument 4 *or* argument 5 remains valid, *and either* argument 7 *or* argument 8 *or* argument 9 remains valid, *and* all of the CON arguments 1, 2, 3, and 6 become invalid.

By choosing to VIEW the arguments referred to in this ledger, the analyst can examine and evaluate the environment of beliefs and/or assumptions that would justify any of these conclusions. As noted above, to derive the conclusion {no diversion} from the available evidence, the analyst would have to continue to accept argument 1 or 2 or 3 or 6; to do that, in turn, requires accepting all the current premises in at least one of these arguments. He would also have to reject arguments 7, 8, and 9; to do that, he would have to reject at least one current premise in each of those arguments.

The analyst will receive more detailed support if he points at and selects the VIEW line corresponding to a particular conclusion. The resulting new display will pinpoint the current assumptions that stand in the way of accepting the specified conclusion. In the present example, if the analyst indicates {no diversion}, the display will show all the premises which are currently assumed true in arguments 7, 8, and 9, and indicate that at least one such premise in each argument must be rejected. In some cases, no change in assumptions will be sufficient to establish a conclusion consistently; in such cases SED will highlight the *beliefs* whose change would be required to justify the conclusion.

The analyst can thus gauge the plausibility of {no diversion} by reviewing the ways it could be derived; by breaking those possible derivations down into changes he would have to make in his current assumptions or beliefs; and evaluating the plausibility or acceptability of the total set of changes. Similarly, he can then examine the set of consistent environments that would justify any of the other possible conclusions, and judge the plausibility of the changes in assumptions or beliefs which those environments demand. SED thus helps the analyst resolve conflict by arriving at a full, coherent story rather than simply performing isolated assessments of individual premises.

3.7.2 Attacking the causes of conflict directly. A second strategy is precisely the opposite: it attempts to diagnose and cure various components of the conflict separately. The conflict situation is broken down in three stages: first, what conclusions are in conflict; second, what arguments are in support of the conflicting conclusions; and third, what beliefs and/or assumptions specifically cause the conflict among arguments. SED thus permits the analyst to identify "culprits" one by one for revision, in order to resolve the conflict.

For example, the "Conflicts" portion of the display in Figure 2 shows that the following pairs of mutually exclusive conclusions are supported: ({no diversion} versus diversion), ({diversion for a West German bomb} versus {no diversion, diversion to other country}), ({no diversion} versus {diversion for a West German bomb}). It also shows the arguments which are pitted against one another in each particular conflict; e.g., acceptance of any one of arguments 1, 2, 3, and 6 in conjunction with any one of arguments 7, 8, and 9 would lead to conflict between ({no diversion}, diversion).

By choosing to VIEW one of these categories of conflict, the analyst receives a display that pinpoints its underlying causes, i.e., the beliefs and assumptions whose change would eliminate that conflict. For example, to resolve the ({no diversion} versus diversion) conflict, the analyst must either (a) reject at least one premise from each of arguments 1, 2, 3, and 6, or (b) reject at least one premise from each of arguments 7, 8, and 9.

Suppose the analyst discredits arguments 1, 2, 3 and 6. The third category of conflict ({no diversion} versus {diversion for a West German bomb}) has now also been resolved, and will vanish from the Figure 2 display. The only remaining conflict is ({diversion for a West German bomb} versus {no diversion, diversion to other country}). The analyst may now choose to VIEW the sources of this conflict, and evaluate potential changes in his assumptions or beliefs to resolve the remaining contradictions in his analysis.

3.7.3 Testing assumptions. When an analyst is ignorant regarding the truth or falsity of a premise, he may nevertheless be able to identify methods of collecting information that would shed light on its status. Such methods can vary widely, and might include, for example, more detailed analysis of data from technical sensors, utilizing more powerful statistical techniques, querying human sources, searching a database, or scanning the open literature. Although in principle "more information is always good," in the real world the decision to collect more information by any of these means is virtually always associated with costs: e.g., a delayed product and, possibly, loss of opportunities for action or policy. As a result, decisions may often have to be made on the basis of incomplete information and assumptions.

Conflict among beliefs and assumptions is an important indication that the collection of more information may be worthwhile, since conflict means that some of those beliefs and assumptions must be mistaken. SED helps the analyst keep track of his or her information collection and analysis options and makes recommendations when such options could significantly contribute to the reduction of conflict.

In SED the analyst may associate with each premise a test or set of tests, i.e., information collection or analysis options that can shed light on the truth or falsity of that premise. Each test is associated in turn with a measure of its cost (in time, money, or some appropriate subjective measure) and with a set of possible outcomes of the test. Finally, each outcome is associated with support either for the premise or for its negation. For example, in order to learn whether or not IAEA inspectors had been hampered in carrying out their duties, the inspectors might be interviewed. The outcome of these interviews would either involve testimony indicating interference or testimony indicating no interference. The cost of the interviews could be assessed as a weighted average of the time and money required to conduct them.

SED's recommendations regarding potential tests work differently depending on the conflict resolution context. In the global mode, SED will indicate tests that could lead to a consistent environment of beliefs for the user-specified conclusion C. Such tests have outcomes that, if obtained, would support a PRO ~~equipment~~ argument for C (if no PRO argument is currently valid) and/or invalidate a currently valid CON argument. In the local mode, SED recommends tests that could eliminate the category of conflict under consideration. Such tests have outcomes that, if obtained, would discredit a currently accepted argument that contradicts some other currently accepted argument.

In all cases, SED recommends tests in accordance with a utility measure based on both benefit and cost (see Appendix A). Benefit is defined in terms of potential reduction in conflict. In the numerical mode, however, where conflict is all-or-none, this measure has little meaning: a test may be well worth performing even though it cannot eliminate conflict entirely. Prioritization among such tests becomes more meaningful when numerical degrees of conflict are recognized.

If collection or analysis of new information is not a feasible alternative, the analyst can still reduce conflict by revising assumptions. SED can make recommendations to the analyst in this process of assumption management; moreover, if desired, SED can automate part of that process. These functions are based on two special types of test, called "query analyst" and "conflict," respectively. The analyst is free to associate either of these tests with any premise. In doing so, he indicates a desire that any assumptions made in regard to that premise should be sensitive to consistency with other beliefs and assumptions.

If the "query analyst" test has been associated with a premise, SED will let the analyst know when a change in assumption regarding that premise would help establish a desired conclusion (in the global mode) or help eliminate a specific conflict (in the local mode). The "outcomes" of this test are, in affect, the decisions of the analyst. If the "conflict" test has been associated with a premise, SED carries out the revision of assumptions automatically.

Here, as in the case of testing more generally, meaningful prioritization of recommendations and automatic procedures requires numerical measures, to gauge the potential impact of changing assumptions on the degree of conflict.

3.8 Adding Numbers to SED

The non-numerical mode of SED, as discussed above, can serve an analyst as a source of valuable insight into the structure of a problem. It may also be of use in constructing an analysis, when the available evidence (or the willingness to make assumptions) is sufficient to warrant all-or-nothing conclusions. More often, however, evidence and the arguments constructed on their basis are inconclusive, and belief is graded rather than all-or-none. Thus, the numerical mode of SED introduces the capability of assessing degrees of belief in the premises of an argument (and their denials), degrees of belief in conclusions, and degrees of impact of denying premises. As a result, a single argument may simultaneously support multiple hypotheses or subsets of hypotheses to various degrees. A second result is that conflict among arguments also becomes a matter of degree.

Numerical measures may be added quite directly to the basic logical structure outlined above. A natural choice for that purpose are Shafer-Dempster belief functions, since (a) they focus on the validity of the link between evidence and conclusion (and permit quantification of the *chance* that an argument proves a conclusion), (b) conclusions are represented as subsets of hypotheses, (c) multiple arguments are combined (*via* Dempster's rule) by computing the chances for common components, or intersections, of their conclusions, and (d) a natural measure of conflict is provided by the chance that the intersection of conclusions is the empty-set.

While belief function measures will significantly enrich the inference structures presented thus far, there is, we believe, a reciprocal benefit. The Shafer-Dempster calculus has been controversial in part, at least, because of the difficulty of assessing and understanding the required measures. More generally, there is a sense that *any* quantitative approach will fail to match people's intuitive ways of reasoning about uncertainty. Numerical measures, we contend, take on a greater degree of clarity and naturalness when they are associated with causally-based arguments of the sort we have described.

3.8.1 Beliefs and assumptions. A belief function is a measure of evidential support that assigns belief to *subsets* of hypotheses rather than (as in Bayesian probability theory) to the hypotheses themselves. As in probability theory, however, each support m(·) is between 0 and 1, and the sum of support for all the subsets must equal 1. The following table is an example of a belief function for the hypotheses in our illustrative intelligence problem:

m(·)	Bel(·)	Pl(·)	No Diversion	Diversion for West German ~~[illegible]~~ Bombs	Diversion to Other Country
.3	.3	.8	*		
.1	.1	.6		*	
.1	.1	.6			*
.0	.4	.9	*	*	
.2	.4	.9	*		*
.2	.2	.7		*	*
.1	1.0	1.0	*	*	*

Since there are three elementary hypotheses, there are 2^3-1=7 subsets of hypotheses for which support must be assessed (the empty set receives no support). m(·) represents the "basic probability assignment" or support measure. Shafer interprets it as the chance that what the evidence *means* is that the truth lies somewhere in the specified subset. In terms of our causal model of inference, m(A) is the chance that A is the set of ground truth situations causally linked to the evidence.

The function Bel(·) summarizes the implications of the m(·) for a given subset of hypotheses. Bel(A) is defined as the total support m(·) for all subsets of hypotheses contained within A:

$$\mathrm{Bel}(A) = \sum_{B \subset A} m(B)$$

In terms of our model, Bel(A) is the chance that A *contains* the set of ground truth situations causally linked to the evidence.

The plausibility function Pl(·) measures the degree to which the evidence fails to exclude a given subset. Pl(A) is the total support for all subsets that overlap with A:

$$\mathrm{Pl}(A) = \sum_{A \cap B \neq \emptyset} m(B) = 1 - \mathrm{Bel}(\text{not-}A).$$

Thus, Pl(A) is the chance that A has at least some element in common with the set of ground truth situations causally linked to the evidence.

Shafer's calculus provides a natural basis for defining the notion of *assumption*. To the extent that support is assigned to subsets containing more than one hypothesis, the evidence fails to discriminate among those hypotheses. Tracing a causal chain from evidence to conclusions (in the manner of Section 3.5), we find branches where the correct causal path cannot be determined, and multiple possible causal origins must be recognized. These multiple possibilities define the scope for assumptions. In particular, an assumption is the replacement, in part or whole, of a less precise conclusion by a more exact conclusion that is contained by it. Put another way, an assumption is a decision to reallocate support assigned to a set of hypotheses (i.e., support that is uncommitted in the respect to the members of the set) to one or more of its subsets.

Suppose that the analyst has reasonably strong evidence (e.g., based on a recent study of procedures at the facility) that no opportunities existed for diverting materials during inspector absences. The analyst therefore assesses a belief of .7 in "Believe No" for the rebuttal, "diversion opportunities during inspector absence." Suppose in addition that there is no evidence suggesting the existence of such opportunities, so "Believe Yes" is assessed as 0. The remaining belief (.3) automatically goes to the universal set, {diversion opportunities, no diversion opportunities}, i.e., it represents a 30% chance that the available evidence proves nothing at all regarding such opportunities.

Diversion Opportunities During Inspector Absences

Believe No .7 Believe Yes 0

Assume No .5 Assume Yes ____

When evidence is insufficient to establish belief or disbelief in a premise, the analyst may make assumptions. In this regard, the numerical mode of SED is a natural extension of the non-numerical mode. To the degree that the sum of belief in a premise and its negation falls short of 1, the analyst is free to make assumptions about the allocation of any portion of the remaining, uncommitted belief. For example, the analyst has a degree of belief of .7 that no opportunities occurred for diversion during inspector absences. By setting "Assume No" at .5 for this premise, he can assign half of the remaining .3 belief to the proposition that no opportunities occurred.

The advantage of specifying assumptions in terms of proportions, rather than absolute amounts, of uncommitted belief is that belief in premises may change as new evidence is acquired, while the assumption *policy* of the analyst remains unchanged. The analyst can set "Assume No" at any proportion p, and "Assume Yes" at any proportion q, as long as the sum of p and q is less than or equal to 1. Then proportion p of the uncommitted belief will be assigned by assumption to the proposition that no diversion opportunities occurred, and proportion q of the uncommitted belief will be assigned by assumption to the proposition that opportunities did occur; any remaining proportion, 1-p-q, stays uncommitted.

3.8.2 A causal-model-based strategy for constructing numerical representation. The belief function concerning diversion of materials was quite complex (with 7 different subsets potentially receiving support), and would undoubtedly be difficult even for an expert to assess directly. In SED, such belief functions do not in fact have to be assessed. Instead, they may be constructed automatically by SED from the user's qualitative, non-numerical judgments regarding causal relationships, together with some far simpler numerical judgments of belief in premises. In this strategy, the analyst *starts* with what is, in effect, a non-numerical argument, i.e., a set of fully accepted premises and a definite conclusion. Each premise receives support totaling 1.0 either by belief or by assumption or by a combination of both. The conclusion of the core argument is a single subset of hypotheses, with support equal to 1. Similarly, the impact of rejecting each premise is specified as the all-or-none replacement of one specified subset by another. After the core argument has been specified, however, beliefs and assumptions may be allocated in any manner between the premise and its negations (or simply remain uncommitted). A graded conclusion, in which multiple subsets of hypotheses are assigned different numerical degrees of support, is then *derived* automatically by SED.

For example, suppose the analyst begins by constructing the causally based core argument in Figure 1. Its conclusion assigns belief of 1.0 to {no diversion}. Then he assigns belief of .2 to the claim that there is statistical inadequacy in the analysis (premise 6). The result will be a new conclusion in which belief of .8 is assigned to {no diversion}, and belief of .2 is assigned to the universal set.

Now suppose in addition that the analyst assigns belief of .3 to the claim that there is bias in the interpretation of results (premise 7). Four different situations are now possible: all premises remain true, only premise 6 is false, only premise 7 is false, and both premise 6 and 7 are false. The techniques of Section 3.5 permit SED to determine what subset of hypotheses would be causally linked to the evidence under each of these conditions. The degree of support for each of those subsets is just the total support for all combinations of premises that have that subset as a conclusion.

The following table summarizes the example:

	Degree of Support	No Diversion	Diversion for West German Bomb	Diversion to Other Country
All premises true	.56	*		
Premise 6 false	.14	*	*	*
Premise 7 false	.24			*
Both premises false	.06	*	*	*

If all premises are true, the truth must be {no diversion}. The chance of this is (1-.2)(1-.3) = .56. If only premise 6 is false, support goes to the universal set. The chance of this is (.2)(1-.3) = .14. Similarly, if only premise 7 is false, support goes to {diversion to other country}, with chance equal to (1-.2)(.3) = .24. Finally, if both premises are false, support goes to the universal set. The chance of this is (.2)(.3) = .06.

The result of adjusting belief in these two premises, therefore, is an alternative argument that assigns .56 belief to {no diversion}, .24 belief to {diversion to other country}, and .14 + .06 = .20 belief to the universal set.

The attractive feature of this assessment strategy is that it provides an *explanation* of the numerical support measures in the conclusion, in terms of the chances that various premises in the core argument are false. These numerical measures are *constructed*, rather than directly assessed, from a perspicuous logical structure. The analyst delineates the basic causal structure of the argument linking evidence and conclusion *before* he assesses any numbers.

3.8.3 A more general numeric framework. Nevertheless, SED also permits a more flexible and general numerical representation of an argument. In this second approach, the conclusion of the core argument may involve assignment of degrees of support to more than one subset of hypotheses; and the conclusion associated with rejecting a premise may also involve assignment of degrees of support to more than one subset of hypotheses. In addition, the impact of rejecting a premise may itself be graded, i.e., it may result in only partial establishment of the conclusion associated with it. In this strategy, like the first, however, the analyst specifies a core argument with a set of fully accepted premises; the system will then automatically compute the effect of altering those beliefs or assumptions on the degrees of support in the conclusion.

This strategy is appropriate under two conditions:

- where a complete causal model, is impracticable or undesirable, and

- where a statistical model is itself part of the causal chain linking evidence and conclusions.

To illustrate the first case, we consider argument 4, based on the U.S. and NATO commitment to defend West Germany in case of attack by Warsaw Pact nations. Suppose that the conclusion of that argument is allocation of .8 support to the subset (no diversion, diversion to other country), and allocation of the remaining .2 support to the universal set. This conclusion rests on an argument to the effect that there is no perceived need for West Germany to build a bomb for itself. The argument rests on several premises: that the West Germans believe the U.S. would honor its commitments, that the West Germans believe that the U.S. response would be sufficient for their defense, and that the West Germans are not expecting a cutoff in U.S. aid due to West German political factors (e.g., rise of the Green Party in West Germany).

This argument does not provide all-or-none support for the conclusion that West Germany is not building a bomb (even if all the premises were true). Nevertheless, this core argument may still be thought of in causal terms, as reflecting premises that have not been made explicit. Such an implicit premise, for example, might be that West Germany does not have some reason to build a bomb *other* than self-defense or deterrence against the Soviet bloc, e.g., for prestige or for its effect on a conflict where U.S. and NATO commitments do not apply. In this example, of course, such premises could be made explicit and added to the core argument, thus permitting an all-or-none conclusion. Nevertheless, it is not always desirable or even possible to enumerate explicitly all the preconditions of an argument; an argument can go wrong in numerous ways that either cannot be anticipated or which seem, individually, too improbable to isolate for separate consideration. The ability to assess direct support for multiple subsets (including the universal set) in the core argument permits the analyst to summarize judgmentally the effects of such implicit preconditions.

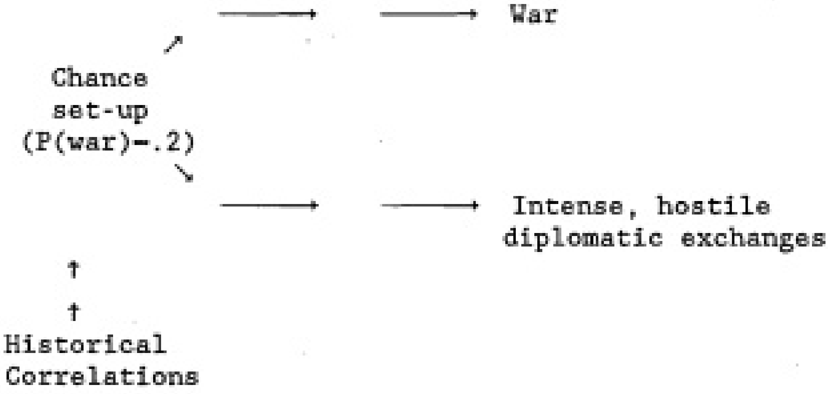
Rejection of a premise in this argument may also have a graded impact, for similar reasons. For example, if the West Germans do not have confidence that the United States will honor its commitments, then they might perceive a need to develop their own bomb; evidence to that effect would support {diversion for West German bomb}. It would not, however, establish that conclusion with certainty. The impact of rejecting this premise (that the U.S. will respond) could be represented by the analyst as follows:

No Diversion	Diversion for West German bomb	Diversion to other country
.6	+	

Rejecting the premise that the U.S. would respond reduces support for {no diversion} to 60% of its former value, and shifts the remaining support to {diversion for West German bomb}. This graded impact can be understood as standing in for causal factors that, for whatever reason, the analyst has chosen not to make explicit. For example, to achieve 100% impact, it would be necessary to know that West Germany did not regard the threat of attack by Warsaw Pact nations as negligible, and that West Germany judged that the deterrent value of possessing its own nuclear capability was significant.

We have illustrated the use of the general numerical approach as an abbreviation for a more complete causal model. Its second function, however, is to represent arguments that appear to be inherently statistical. For example, historical data may suggest that the incidence of war is 20% when diplomatic exchanges of a certain type and frequency occur between two countries. Suppose then that exchanges of this sort are subsequently observed in a particular case. The analyst may infer that the chance of war is 20%.

In effect, this strategy involves adopting the fiction, or metaphor, that a chance set-up is part of the causal chain linking evidence and conclusions. In this example, the chance set-up is a common cause both of the exchanges between the nations and (potentially) of the war between them:



Premises in this argument will concern all three of the branches in this network: i.e., are these errors in the sampling or analysis of the historical data? could the observed diplomatic exchanges be accounted for in some other way? and are there special conditions that would facilitate or hinder the initiation of a war? The latter two questions concern, more broadly, the issue of whether the present situation is *representative* of the historical data.

To the degree that all premises are accepted, the conclusion will be a "Bayesian belief function," in which .20 support is assigned to {war} and .80 support to {no war}. Typically, however, careful evaluation will lead to the rejection of at least some premises, e.g., circumstances will be identified suggesting that the present situation is non-representative in certain ways. In such cases, the result will be a "discounted Bayesian belief function," in which support is assigned to {war} and {no war} in the ratio .2/.8, but some support is also assigned to the universal set, signifying the chance that the present argument is irrelevant. Approaching statistical inference in this way (a) permits the analyst to go beyond a literal acceptance of frequency data, and (b) allows him to integrate the results of statistical analysts with non-statistical processes of reasoning.

3.8.4 Exploring alternative representations. A concern in communicating intelligence results is to render them in as precise and unambiguous a fashion as possible. A key feature of SED is that it permits analysts to explore a set of alternative representations of the same argument, which differ among themselves in degree of precision and in degree of convergence on a single conclusion. Greater precision or convergence can be achieved at the cost of making assumptions. For example, by assuming that no opportunities existed for diversion of nuclear materials during inspector absences, the analyst purchases increased precision: i.e., more support for {no diversion}, as opposed to support for the set {no diversion, diversion to other country}. By assuming that biases did not occur in data analysis, he purchases increased convergence: i.e., exclusive support for {no divergence}, as opposed to sharing support between the mutually exclusive conclusions {no diversion} and {diversion to other country}. In assuming that inspectors were not hampered, he obtains both more precision and more convergence.

We would contend that assumptions of this sort are inevitable in any analysis--that is, in any analysis of a reasonably complex issue that arrives at an intelligible conclusion. The important thing is not to avoid assumptions, but to keep careful, explicit account of the assumptions that are made, to track their impact on the conclusions both of the particular argument in which they arise and the combined argument based on all the evidence, to alert the user in case of trouble, and to support careful revision of assumptions when appropriate. SED is designed to support the analyst in this type of problem solving.

An illuminating representation can be provided of the space of problem representations which SED can explore.1 The two dimensions of the space are precision and convergence. Yager (#MII-504) has proposed measures of these dimensions for Dempster-Shafer belief functions.

His measure of precision is:

$$P = \sum m_i/n_i$$

where m_i is the support assigned to subset i and n_i is the number of hypotheses which belong to i. P is at a maximum (equal to 1) when all support is assigned to elementary hypotheses, e.g.,

Belief	No Diversion	Diversion for West German Bomb	Diversion to Other Country
.33	*		
.33		*	
.33			*

P is at a minimum (equal to 1/n) where n is the total number of elementary hypotheses, when all support is assigned to the universal set.

Yager's measure of convergence is:

$$C = \Pi \, Pl_i^{m_i}$$

where m_i is the support assigned to subset i and Pl_i is the plausibility of subset i. The key to high convergence is that every subset that has any support also have high plausibility, i.e., that there be no support going to subsets that are incompatible with it. Thus C is at a maximum (equal to 1) when all support is assigned to nested subsets of hypotheses (hence, $Pl_j = 1$ for subsets j in the nested series, and $Pl_i = 0$ for all other subsets i); e.g.,

Belief	No Diversion	Diversion for West German Bomb	Diversion to Other Country
.3		*	
.4	*	*	
.3	*	*	*

Convergence is at a minimum (equal to 1/n) when support is divided evenly among the elementary hypotheses. In "Bayesian belief functions," when all support is allocated to elementary hypotheses, C is a simple monotonic function of Shannon's entropy measure.

Figure 3 combines these two dimensions in a space of problem representations, and shows how assumptions can help the analyst arrive at a desirable location anywhere within that space. Bias assumptions typically move the representation toward greater convergence; precision assumptions typically move the representation to greater precision. The *rejection* of these assumptions, conversely, will typically produce more imprecision and more divergence in the conclusion of an argument. Note that maximal imprecision and maximal divergence are mutually exclusive, since assigning support to *large* subsets makes it hard to assign support to many *inconsistent* subsets.

An important side benefit of this framework is that it accommodates different uncertainty calculi as special cases. Figure 4 shows that Bayesian, fuzzy, and deterministic models of uncertainty are all located in this space. Bayesian models are maximally precise belief functions. Fuzzy models are maximally convergent (i.e., nested) belief functions. Deterministic models are both maximally precise and maximally convergent. If he wishes, the analyst can construct a model of any of these types by making suitable assumptions. He may then investigate the success of that representation in terms of the conflict it produces with other arguments and assumptions.

3.8.5 The impact of numbers. As we have seen, incorporation of numerical measures into SED permits the analyst to represent gradations of belief and to manipulate statistical hypotheses. Even more importantly, however, it provides the possibility of a more flexible process of control over reasoning, in which the relative importance of different arguments can be weighed and the

Figure 3: A Space of Problem Representations

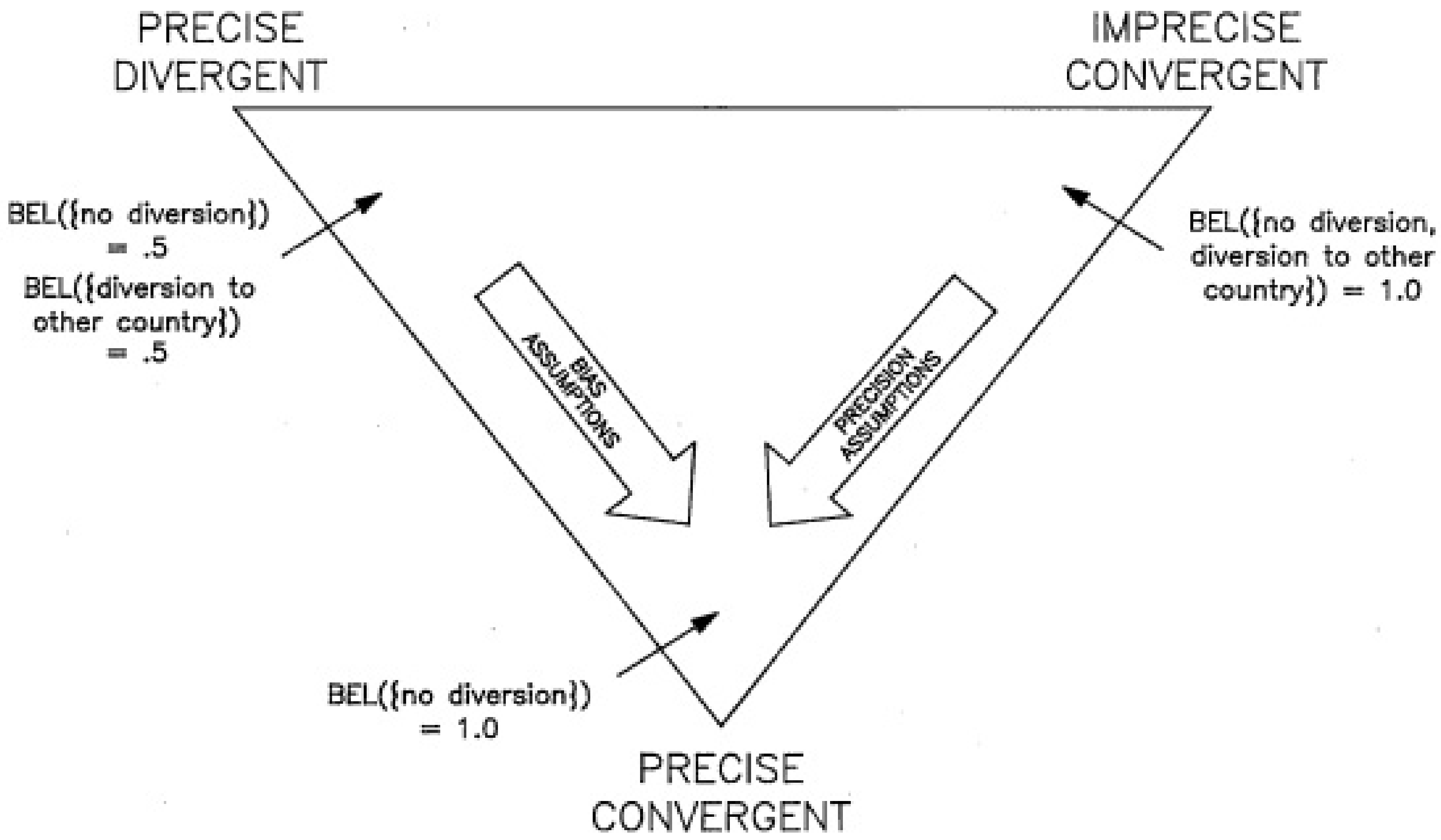
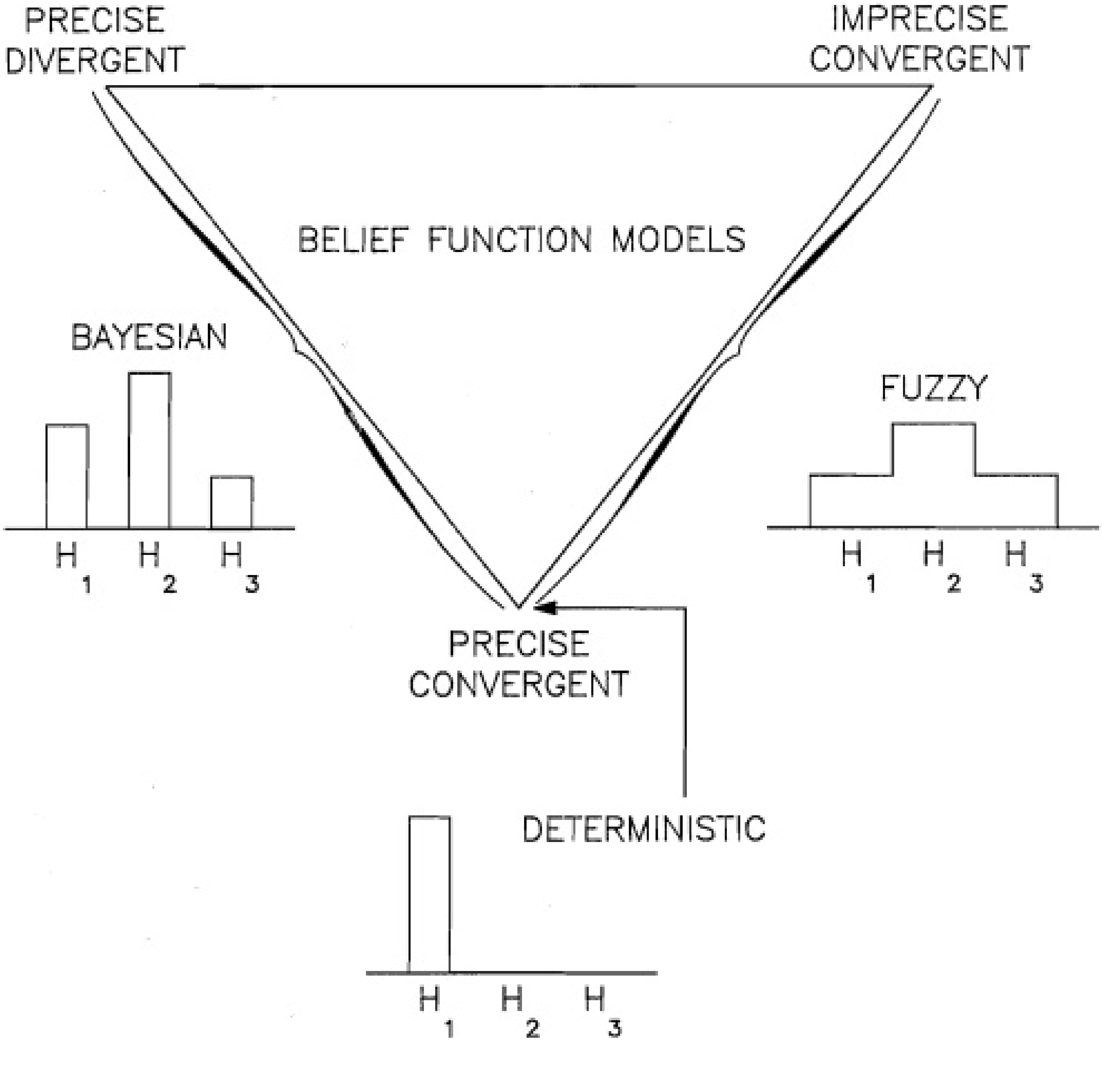


Figure 4: Multiple Models of Uncertainty



culpability of different assumptions and beliefs in conflict more appropriately assessed. In this section, after briefly reviewing Dempster's Rule, we review the role of numerical measures in these functions. A more detailed description may be found in Appendix A.

Combination of arguments. Dempster's Rule as a mechanism for combining arguments is a natural generalization of the approach in the non-numerical mode. It assigns support to the intersections, or common components, of the conclusions of the arguments being combined. If $m_1(A)$ is the support given to A by argument 1, and $m_2(A)$ is the support given by argument 2, the support that should be given to A by the two pieces of evidence is:

$$m_{12}(A) = \frac{\sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2)}{1 - \sum_{A_1 \cap A_2 = \emptyset} m_1(A_1) m_2(A_2)}.$$

The numerator here is the sum of the products of support for all pairs of subsets A_1, A_2 whose intersection is precisely A. The denominator is a normalizing factor which ensures that $m_{12}(\cdot)$ sums to 1, by eliminating support for impossible combinations.

In causal terms, an argument supports a subset of hypotheses if it provides a causal chain linking that subset to the evidence. A given argument may support more than one causal chain, if there is uncertainty about premises. To compare two such arguments, we consider all combinations of the causal chains from each argument. The meaning of each combination is the intersection of the subsets they support; and the chance of both of them being valid is the product of the support those causal chains receive from their respective arguments.

Contribution to a conclusion. In the "Conclusion" section of Figure 2, reasons pro and con are provided for each possible conclusion. The numerical mode makes it possible to give a more sophisticated account of why a conclusion is believed. It displays a measure of the *relative* contribution of any PRO argument to current belief in that conclusion. (The measure is the normalized sum of the products for all n-tuples of subsets which (a) have the specified conclusion as their intersection *and* (b) would not have the relevant conclusion as their intersection if the given argument were not included.) If the analyst wishes to resolve conflict by establishing a consistent environment for a particular conclusion, he or she can use this measure to avoid the discrediting of arguments that provide key support, remaining free to reject arguments which provide only incidental support.

Degree of conflict. The numerical mode provides a natural measure of conflict among arguments. This is 1 minus the normalization constant in Dempster's Rule, i.e., the sum of the products for all subsets whose intersections are empty. This measure is the probability of an impossible state of affairs (i.e., nothing being true), given all our current beliefs and assumptions. The higher it is, therefore, the greater the pressure to modify some of those beliefs and assumptions to reduce the conflict (just as in classical hypothesis testing, if the probability of *not* obtaining the actually observed sample is very high given the null hypothesis, we tend to reject the null hypothesis). In short, SED interprets the conflict measure as *evidence* of mistaken beliefs and assumptions.

In SED, the user may set a threshold on the conflict measure, indicating how much chance of error he is willing to tolerate. SED uses this measure in deciding when to initiate conflict resolution procedures -- i.e., recommending tests, querying the user, or automatically revising assumptions.

Contribution to conflict. We saw that SED can provide a measure of the relative contribution of an argument to a conclusion. In precisely the same way, it can provide a measure of the contribution of an argument to conflict (since conflict is "belief in the null set"). The analyst may use this measure to focus his or her attention on components of conflict where revision of beliefs or assumptions may do the most good.

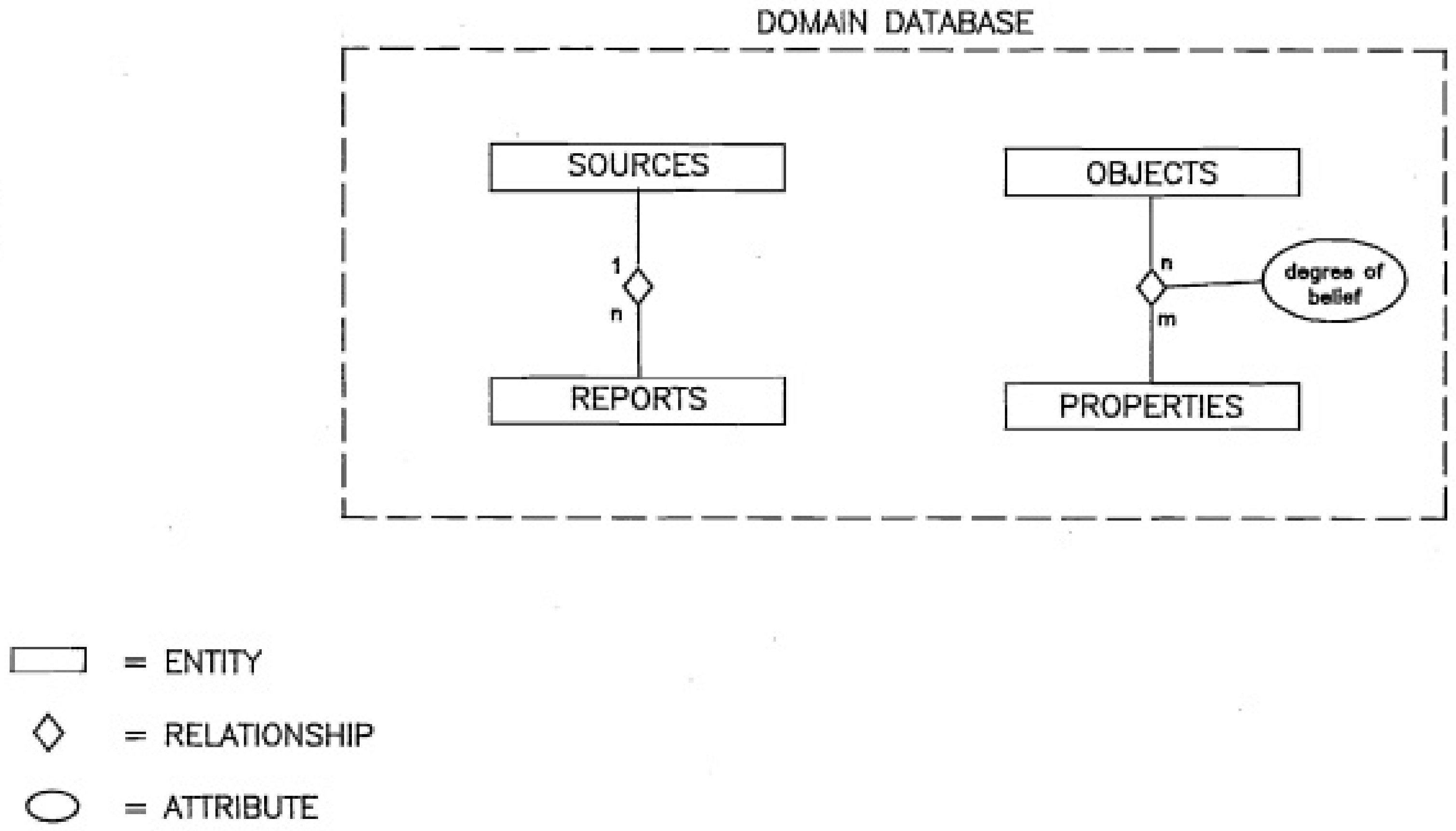
The benefit of testing or changing assumptions and beliefs. In conflict resolution, SED may make recommendations regarding the collection or analysis of further information. Such recommendations are based on an evaluation of each test in terms of its costs and benefits. The numerical mode makes possible measures of benefit which represent the potential reduction in conflict obtainable by performing that test.

3.9 Data Structures in SED

The models of uncertainty supported by SED may all be represented within the constraints of the standard relational model of data (Appendix C). Here we briefly describe a set of data structures that appear to be adequate for this job. The convention we have adopted is a variation of the "entity-relationship" diagrams introduced by Chen (1976). This provides a higher-level characterization than standard relational schemas, and, as a result, a cleaner conceptualization of the basic semantic relationships in the database. However, there is an immediate, unambiguous translation into standard relational structures.

The convention of these diagrams is as follows. A box stands for an "entity relation"; roughly, this is a data table with a single key field and an arbitrary number of non-key fields which are referred to as "attributes." For example, in Figure 5, a table concerning human sources of information might use the name of the source as the key field (i.e., a unique identifier for each object or row in the table); it might also include such attributes (not shown) as age, address, length of service, etc. Attributes may be represented in the diagram by circles, which are connected by lines to the relations to which they belong. Diamonds are "relationship relations", i.e., data tables with compound key fields. For example, a table might have the name of the source as one component key and the number of a report as a second component key. Then each row of the table concerns the combination of that source with that report. In the diagram, the component keys in a relationship relation can be determined by looking at the entity relations which are connected by lines to the diamond. Small numbers next to those lines represent the semantic mapping between the keys. For example, in Figure 5 the "1" next to the

Figure 5: A Minimal Intelligence Database



Source line and the "n" next to the Reports line indicate that each source will be paired with multiple reports in that relationship relatiôn. In the relationship relation involving objects and properties, however, each object may be paired with multiple properties, and each property with multiple objects.

Figure 5 is a minimal intelligence database: it permits the analyst to store information about data and their sources, and also to represent hypotheses (i.e., object-property combinations) together perhaps with some measure of degree of belief (see Appendix C). What it lacks is any means for *linking* the data and the hypotheses by means of evidential arguments.

Figure 6 provides an overview of the data structures introduced by SED for this purpose. Diagramming in this way dramatizes the role that arguments play in linking data (reports) to hypotheses via premises.

A separate table stores information at the level of problems (e.g., "determine whether West Germany diverted nuclear material"). Attributes here might include who is working on the problem, who requested the study, deadline for solution, etc. A relationship relation keeps a record of all the arguments that have been associated with each problem. Subsets of hypotheses are derived from the domain database of hypotheses. Each argument/subset combination is associated with a degree of belief, i.e., the current support for that hypothesis from that argument. Similarly, each problem/subset combination is associated with a degree of belief, i.e., the combined support from all the arguments associated with that problem. Each argument is associated with premises, although the same premise may appear in more than one argument. Premises are associated with tests (to determine their truth or falsity), and tests are associated with outcomes.

Figure 7 presents a somewhat richer picture of the attributes required for processing information in SED. Thus, each argument may be characterized by a conflict threshold (indicating how much conflict triggers various conflict resolution processes among the arguments associated with that problem) and a responsibility threshold (indicating degree of contribution to the conflict that is required before an assumption will be tested or revised).

Figure 6

DATA STRUCTURES IN SED: BASIC SYSTEM

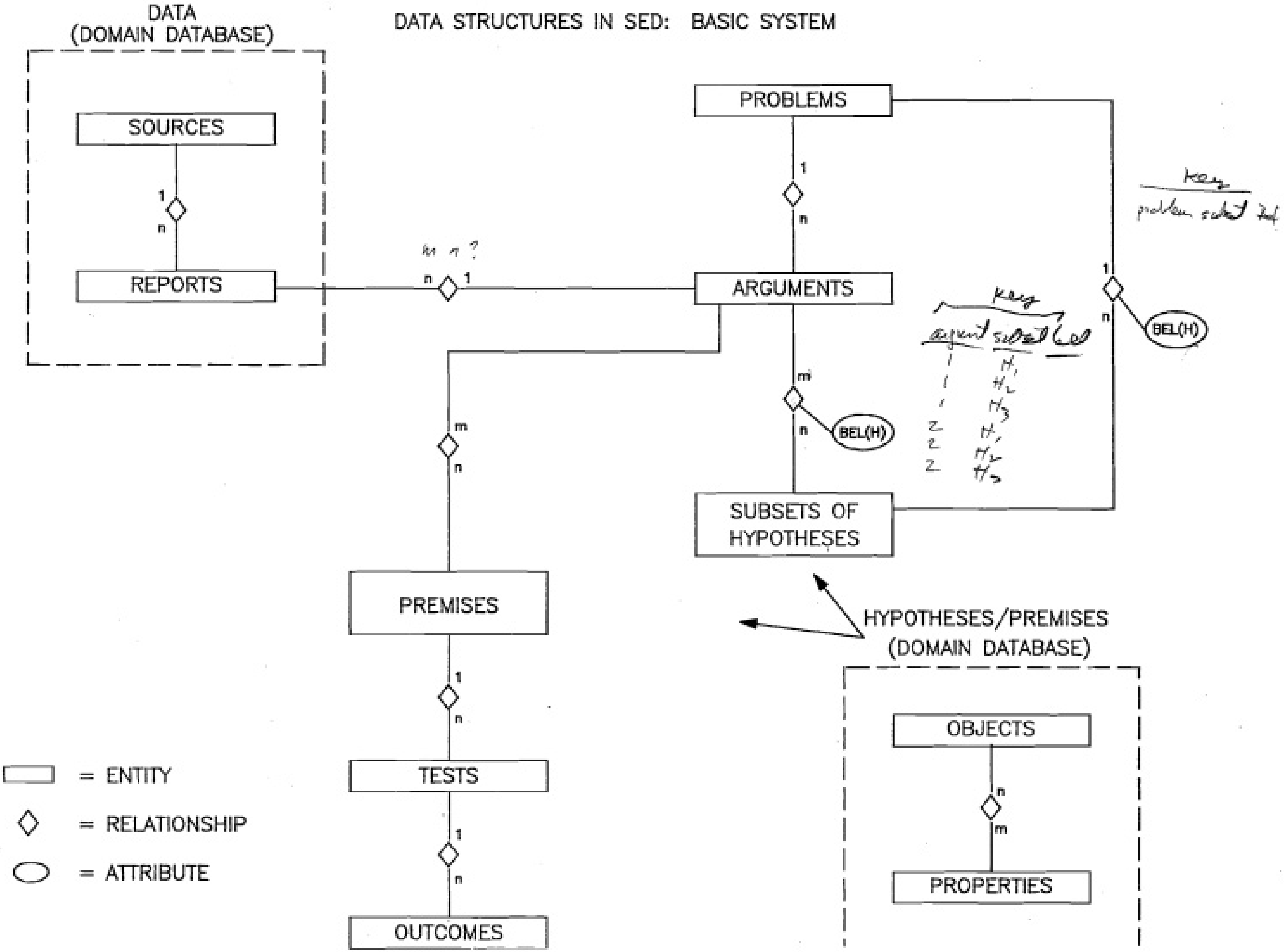
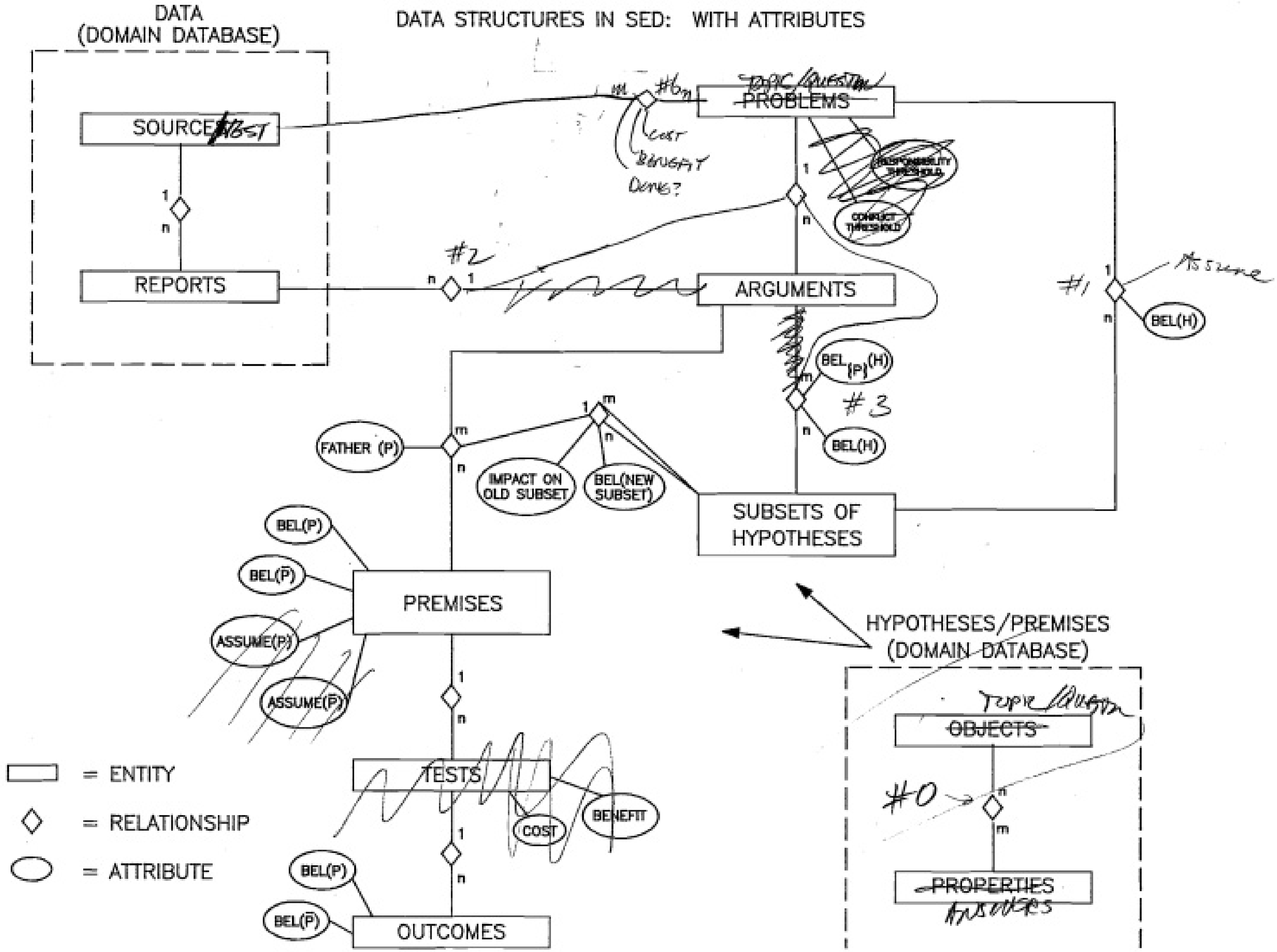


Figure 7

DATA STRUCTURES IN SED: WITH ATTRIBUTES



Argument/subset combinations need to be characterized not only by the current belief, but by the belief assessed in the core argument as well ($Bel_{(p)}(H)$). Each premise/argument combination is characterized by the location of the premise in the causal chain associated with that argument (i.e., what is the preceding or "Father" premise). Premises themselves are characterized by degrees of belief and assumption, independently of the arguments they enter into. A very high level relationship relation characterizes the impact of negating each premise in each argument. This characterization includes pairs of subsets, one of which is reduced in support and the other of which is increased; the relationship relation has attributes corresponding to the degree of impact on each. Finally, tests are associated with costs and benefits. And outcomes of tests are associated with the degrees of belief in the premise and its negation which they support.

4.0 CONCLUSIONS

A major goal of SED is to support the processes of reasoning and methods of organizing data that are preferred and practiced by successful intelligence analysts. It does so in a variety of ways:

- by emphasizing qualitative causal models of how evidence is linked to conclusions;
- by permitting the development of simplified problem representations through the adoption (and testing) of assumptions;
- by providing an "evidential spreadsheet" in which the implications of changes in beliefs and assumptions become readily apparent;
- by supporting an iterative process of argument construction, in which conflict is used as a cue to reexamine and revise assumptions; and
- by supporting a strategy of hypothetical reasoning in which the plausibility of an overall coherent "story" can be assessed.

Numerical measures, from this point of view, are quite incidental; they are means to the end of facilitating an effective and natural process of reasoning. Indeed, SED minimizes numerical assessment and subordinates it to the process of constructing and manipulating qualitative arguments.

A second major goal of SED, complementary to the first, is to capture what we perceive to be the strengths of recent theoretical work on uncertainty, without some of the weaknesses.

The underlying representation of uncertainty in SED is Shaferian. The uncertainty calculus is a modification of belief function calculus which allows for assumptions and a non-monotonic backtracking capability. SED's belief functions arise out of a highly differentiated knowledge structure, patterned after Toulmin's model of argumentation. The underlying qualitative models giving rise to belief functions, associated with specific types of premises affecting their reliability, can be thought of as endorsements, in the style of Paul Cohen.

Because it incorporates both belief function theory (with Bayesian theory as a special case) and possibility theory, SED can deal with uncertainty due to chance, to incompleteness of evidence, and to imprecision. The existence of numerical belief measures allows SED to overcome the inability of non-numerical theories (such as Toulmin's model, the theory of endorsements, or non-monotonic logic) to represent and manipulate gradations of uncertainty. The incorporation of qualitative theories allows SED to take advantage of the strengths of the qualitative reasoning ability of traditional AI systems.

Although its reasoning processes derive from belief function theory, SED incorporates innovations that ameliorate some of the shortcomings of the Shaferian theory. An important difficulty for Shafer's theory is that it does not allow reassessment of the reliability of a source based on what the source says. SED allows *assumptions* about source reliability to be made; actual source testimony may then prompt reexamination of these assumptions. In the ideal case, any actual reassessment of source credibility is based on additional evidence--the role of source testimony is as a trigger to test for that evidence.

A second major difficulty of belief function theory is the inability of Dempster's Rule to handle cases where arguments are not based on independent evidence. In our experience, it is often reasonable to model non-independence as shared premises. Thus, SED provides an important feature that is lacking in traditional Shaferian systems: namely, a capability for representing non-independence between evidential arguments.

A third property of belief function models, at least as they have appeared in the literature thus far, is "flatness." That is, application of Dempster's Rule assumes all input belief functions to be defined on the same hypothesis space. Thus, the theory appears not to support hierarchical inference. However, the tools of conditional embedding, minimal extension, and marginalization allow the definition of a mechanism within Shaferian theory for hierarchical inference. The details of the Shaferian hierarchical inference model are given in Appendix B.

As noted above, the SED system bears an important relationship to Doyle's (1979) non-monotonic logic. In non-monotonic logic, an AI rule-based system is endowed with the capacity to make assumptions and draw conclusions based on the assumptions. When the system encounters a contradiction, it responds by retracting an assumption to resolve the contradiction. Interest in Doyle's system has grown because of its similarity to human reasoning processes.

SED endows an uncertainty calculus (the theory of belief functions) with a non-monotonic reasoning capability. As such, it is a very powerful theory. SED is endowed with a flexible self-reconciling capability that incorporates the strengths of non-monotonic logic. Yet SED counterbalances the weaknesses of non-monotonic logic by providing a measure of the strength of evidential support and an explicit prioritization for belief revision.

Within Doyle's system, an assumption is a statement that is believed without proof. Assumptions are different from axioms (which are also believed without proof) because, although the system believes assumptions, it remembers that they are "only" assumptions, and drops them readily when it discovers they are responsible for contradictions.

In SED, an assumption is defined as a *rule for allocating uncommitted belief*. The SED system reasons with the "assumed" belief functions (those with uncommitted belief allocated according to assumption), but it keeps track of the original belief functions. When it encounters conflict, the SED system goes through its assumptions, searching for tests to justify a reallocation of the uncommitted belief.

In Doyle's system, the entities with which the system reasons are propositions, or statements. In SED, the fundamental objects are belief functions. Note that in SED one can assign belief 1 to a conclusion. This is equivalent to a proposition (if evidence, then conclusion). In this sense, SED is more general than Doyle's system.

The SED model can be though of as a very general "executive" for keeping track of the process of building and testing an evidential model. As such, SED has the potential to provide a formal theory for the (typically informal) process of model building and testing. DeGroot (1982) has said that a good statistician always reserves "a little pinch of probability" for the event that the model is wrong. In general, a good scientist will examine the output of a model, and, if the results are anomalous, will investigate alternative models. Previous attempts to model the "little pinch of probability" as a Bayesian pinch have run into grave difficulty. How much is a big enough pinch? What is the space of alternate models, and how should one allocate the pinch among them? In the typical situation, the modeler adopts a parsimonious model; allowing even a small pinch of probability to *arbitrary* alternate models might result in a very unparsimonious model being adopted because it fits the data exactly.

These problems disappear in the SED model. One need not worry about how much of a pinch to allocate; one simply assigns initial plausibility of 1 to the current working model. Model diagnostics (e.g., outliers) then play the role of "triggers" to search for alternate models. Indeed, alternate models may not even have been developed initially; they may be generated as a response to the existence of conflict.

Of course, the credibility factor model has its own problems, such as finding a rationale for how much conflict is "enough" to trigger the search for alternate models. Still, it provides a valuable and very general framework for theorizing about the evidential reasoning process.

APPENDIX A: MATHEMATICAL BASIS FOR THE SELF-RECONCILING EVIDENTIAL DATABASE

Premises form the basis for the non-monotonic aspect of the Self-Reconciling Evidential Database. This is so because premises may be assumed true except to the degree that there is evidence for their negation. A problem in the argument (typically conflict, although as we shall see, there can be other problems) acts as a trigger to reexamine such assumptions and recommend further information collection in a non-monotonic belief revision process.

A precision premise is a special case of a premise which when negated typically acts to *discount* an argument (i.e., it weakens the *strength* of the argument without changing its direction). We begin our examination of the theory by examining the simplest special case: a precision premise whose effect is *total* discrediting of the associated belief function.

Special case 1: Total discrediting. Consider a belief function Bel_X over the set of hypotheses X. The associated basic probability function m_X is a probability distribution over the power set of X, and determines the belief function as follows:

$$Bel_X(A) = \sum_{B \subset A} m_X(B), \qquad (1)$$

for all subsets A of X.

Let us suppose the existence of a premise p whose rejection would invalidate the belief function Bel_X. In other words, the belief function Bel_X applies only when p is accepted. If p is false, there is no information about the set X. Mathematically, this can be expressed by extending the belief function Bel_X to the set X×P, where $P = \{p,\overline{p}\}$ is the set indicating truth or falsity of the premise p. Denote the extended belief function by Bel_e. Each focal element A of Bel_X is extended to the focal element $(A\times\{p\})\cup(X\times\{\overline{p}\})$ of Bel_e, with the basic probability assignment $m_X(A)$. The interpretation is that belief $m_X(A)$ is assigned to the set A if p is true, but to X (the universal set) if p is false. Belief assigned to X reflects the degree to which evidence does not discriminate among the hypotheses in X.

Now, suppose there is a belief function Bel_P over the set P, with corresponding basic probability function m_P. The function Bel_P represents belief in the truth or falsity of the premise p. This function, too, is extended to the set X×P. The new focal elements are $X \times \{\overline{p}\}$, $X \times \{p\}$ and X×P, with basic probability assignments $m_P(\{\overline{p}\})$, $m_P(\{p\})$, and $m_P(P)$, respectively. (This is called the minimal extension (Shafer, 1982), and reflects the judgment that the evidence on which Bel_P is based contains no information about X.)

Assuming that the two functions Bel_X and Bel_P are based on independent evidence, we may combine the extended belief functions using Dempster's Rule. The combined belief function, denoted by Bel_*, has three types of focal element. These focal elements are displayed in Table A-1. The first type of focal element represents belief in focal elements A of Bel_X and {p} of Bel_P. Thus, to the extent that the evidence supports p, belief is assigned to focal elements of Bel_X. The second type of focal element represents belief in the falsity of the premise p; since Bel_X would be invalidated if p is rejected, all belief is assigned to X. The third type of focal element represents belief committed to neither p nor $\overline{p}$. In words, these focal elements represent belief in "A and p, or X and not p."

Table A-1: Focal Elements for Combined Belief Function (Total Discrediting)

	Focal Element of Bel_*	Basic Probability
Premise true	$A \times \{p\}$	$m_X(A)m_P(\{p\})$
Premise false	$X \times \{\overline{p}\}$	$m_P(\{\overline{p}\})$
Premise unknown	$A \times (\{p\}) \cup (X \times \{\overline{p}\})$	$m_X(A)m_P(P)$

Since we are interested in the premise P only insofar as it affects belief about the hypotheses in X, the next step is to reduce Bel_* to a new belief function over X. The standard Shaferian operation would be to form the marginal belief function. Since the second and third types of focal element are consistent with the truth of any hypothesis in X, the result of marginalizing -s to assign their belief to the set X. That is,

$$\begin{aligned} m_M(X) &= m_p(\{\bar{p}\}) + \sum_{A\subset X} m_X(A)m_P(P) \\ &= m_P(\{\bar{p}\}) + m_P(P) \\ &= 1 - m_P(\{p\}) \end{aligned}$$

A focal element $A\times\{p\}$ of the first type is consistent only with hypotheses in A. Thus, marginalization results in basic probability

$$m_M(A) = m_X(A)m_P(\{p\})$$

assigned to proper subsets of A. Thus, the marginal belief function is a *discounted* version of Bel_X, with discount rate $\delta = 1\text{-}m_p(\{p\})$ equal to the pausibilty that the premise is false.

While this procedure is a correct application of belief function theory, it does not accord well with the human expert's process of provisionally adopting the belief function Bel_X, searching for discrediting factors only when prompted by conflict with other evidence. Marginalization treats the case where the truth of the premise is unknown as equivalent to the case where it is known to be false; i.e., it assigns all belief to the universal set X except to the extent that there is positive belief in the truth of the premise (in particular, when Bel_P is vacuous, so is the marginal of Bel_*).

In contrast, a human expert might adopt the belief function Bel_X except to the extent that there is positive belief in the falsity of the premise. Mathematically, this amounts to treating the third type of focal element in Table A-1 as if it were the first type. Equivalently, we may think of the belief function Bel_P as being provisionally replaced by a function which gives the universal set P no weight, but assigns belief $m_P(\{p\})+m_P(P) = \text{Pl}(\{p\})$ to the focal element $\{p\}$. In other words, adopting p as an assumption is equivalent to assigning uncommitted belief to p. The result for the target hypotheses X is a discount rate $\delta = m_P(\{\bar{p}\})$ equal to the belief committed directly to the falsity of the premise, $\bar{p}$.

Thus, by making assumptions, the analyst may choose to adopt the belief function Bel_X so long as there is no direct evidence to the contrary. When conflict occurs, SED helps the analyst in the process of examining and revising such assumptions, or in searching for evidence that would increase belief in $\overline{p}$, thereby causing the discount rate to increase and conflict to decrease.

This framework may be extended to two or more discrediting factors as follows. Suppose there are n premises, $p_1, \ldots, p_n$, such that the falsity of any one would discredit the belief function Bel_X. That is, conditional on $p_1 \wedge p_2 \ldots \wedge p_n$ the belief function Bel_X applies, but conditional on $\overline{p}_1 \vee \ldots \vee \overline{p}_n$ the vacuous belief function applies.

Suppose belief functions Bel_i are defined over the $P_i = \{p_i, \overline{p}_i\}$, and that they are judged to be based on independent evidence. Each Bel_i can be extended (using the minimal extension described above) to the product space $P_1 \times \ldots \times P_n$. Combining these by Dempster's Rule produces a belief function Bel_P over the product space. Its focal elements are of the form $Q_1 \times \ldots \times Q_n$, where Q_i is a focal element of Bel_i. Their basic probabilties are defined by $m_P(Q_1 \times \ldots \times Q_n) = m_1(Q_1) \cdots m_n(Q_n)$.

Recall that the belief function Bel_X applies only when *all* premises hold; otherwise, the vacuous belief function applies. Suppose, as above, that the analyst wishes to assume the premises are true except to the extent they are directly contradicted by the evidence. The result would again be a discounted belief function, but this time the discount rate would be equal to the belief committed directly to the falsity of at least one of the p_i. That is,

$$\delta = Bel(\overline{p}_1 \vee \overline{p}_2 \vee \cdots \vee \overline{p}_n) = 1 - Pl(p_1 \wedge p_2 \wedge \cdots \wedge p_n).$$

The following theorem demonstrates that δ is defined by:

$$1 - \underset{i}{\delta} = \Pi\ (1-\delta_i). \qquad (2)$$

<u>Theorem</u> Consider n belief functions, $Bel_1, \ldots, Bel_n$ concerning n premises $p_1, \ldots, p_n$. Suppose each Bel_i is extended to the product space $P_1 \times \cdots \times P_n$ (where $P_i = \{p_i, \overline{p}_i\}$), and these belief functions are combined according to Dempster's Rule into a belief function Bel. Then

$$Pl(p_1 \wedge p_2 \wedge \cdots \wedge p_n) = \prod_{i=1}^{n} Pl_i(P_i).$$

<u>Proof</u>: The result is clearly true for $n = 1$. Suppose it holds for $n-1$. Now, the plausibility that all premises hold is the sum of the basic probabilities of all subsets *intersecting* $(p_1,\cdots,p_n)$ - that is, of all subsets consistent with the truth of all the premises. That is,

$$\mathrm{Pl}(p_1\wedge\cdots\wedge p_n) = \sum_{(p_1,\cdots,p_n)\in Q} m(Q) = \sum_{p_i\in Q_i} m_1(Q_1)\cdots m_n(Q_n)$$

$$= \sum_{\substack{p_i\in Q_i \\ i \le i \le n-1}} [m_1(Q_1)\cdots m_{n-1})](m_n(\{p_n\})+m_n(P_n))$$

$$= \mathrm{Pl}(p_1\wedge\cdots\wedge p_{n-1})\mathrm{Pl}_n(p_n)$$

$$= \prod_{i=1}^{n} \mathrm{Pl}_i(p_i)$$

Since $1-\delta$ is defined to be equal to $\mathrm{Pl}(p_1\wedge\cdots\wedge p_n)$ and $1-\delta_i$ to be $\mathrm{Pl}(p_i)$, Equation (2) follows.

A general theory of credibility factors. Thus far, the only consequence of rejecting a premise has been to discredit a belief function, i.e., to increase support for the universal set at the expense of an across-the-board reduction of belief in all other sets. Rejecting a premise may, however, have other effects; e.g., in changing the direction of an argument or causing support to shift from one focal element to another.

As a prelude to introducing the general theory, let us reflect a bit more closely on the structure of the above model. The model says that conditional on the truth of premise p, belief function Bel_X applies; conditional on the falsity of premise p, the vacuous belief function applies. The combined belief function Bel_e is formed via the *conditional embedding* of these conditional belief functions into the space X×P (conditional embedding is discussed in detail in Appendix B).

In the general model, we postulate that the falsity of premise p leads to belief function Bel_X^c. Our model, in words, is "if p is true, then Bel_X; if p is false, then Bel_X^c." In the special case described above, Bel_X^c is vacuous.

As before, we use conditional embedding to obtain a belief function on the product space X×P. The focal elements are of the form $(A\times\{p\})\cup(B\times\{\overline{p}\})$, where A is a focal element of Bel_X and B is a focal element of Bel_X^c. The corresponding basic probability assignment is $m_X(A)m_X^c(B)$.

As before, we introduce a belief function Bel_P over the space P, and again use the minimal extension to extend it to X×P. Combining by Dempster's rule yields the combined belief function Bel_*. Table A-2 gives the resulting focal elements (compare with Table A-1 above).

The focal elements of the marginal belief function are of two types: (i) a focal element of either Bel_X or Bel_X^c and (ii) the intersection of a focal element of Bel_X and a focal element of Bel_X^c. The basic probability of a set A is given by:

$$m_M(A) = m_X(A)m_P(\{p\})+m_X^c(A)m_P(\{\overline{p}\})+\sum_{B\cap C=A} m_X(B)m_X^c(C)m_P(P)$$

Table A-2: General Credibility Factor Model

	<u>Focal Element of Bel_*</u>	<u>Basic Probability Assignment $m_*(\cdot)$</u>
<u>Premise true</u>	$A\times\{p\}$	$m_X(A)m_P(\{p\})$
<u>Premise false</u>	$B\times\{\overline{p}\}$	$m_X^c(B)m_P(\{\overline{p}\})$
<u>Premise unknown</u>	$(A\times\{p\})\cup(B\times\{\overline{p}\})$	$m_X(A)m_X^c(B)m_P(P)$

(Note: A denotes a focal element of Bel_X
B denotes a focal element of Bel_X^c)

If we choose to assume the truth of the premise to the extent consistent with our evidence, SED shifts all uncommitted belief to p before marginalizing. Belief in focal element A is now given by $m_M(A) = m_X(A)Pl_p(\{p\}) + m_X^c(A)Bel_p(\{\overline{p}\})$.

Special case 2: Partial discrediting. It may be the case that the rejection of premise would only *partially* discredit a belief function. For example, even if we were sure that IEAE inspectors are incompetent, we may not wish to completely discount their report; instead, we might wish to discount the belief function for no diversion based on that report by a discount rate β (where $\beta<1$). This corresponds to the case where m_X^c is a *discounted* version of m_X. That is, $m_X^c(A) = (1-\beta)m_X(A)$ for all proper subsets A, and $m_X^c(X) = (1-\beta)m_X(X) + \beta$.

In this case, the focal elements of the marginal belief function are the same as those of Bel_X. Belief for proper subsets is given by

$$\begin{aligned} m_M(A) &= m_X(A)Pl_P(\{p\}) + m_X(A)(1-\beta)Bel_P(\{\overline{p}\}) \\ &= m_X(A)[1-Bel_P(\{\overline{p}\}) + (1-\beta)Bel_P(\{\overline{p}\})] \\ &= m_X(A)[1-\beta Bel_P(\{\overline{p}\})]. \end{aligned} \tag{3}$$

Thus, the result is a discounted belief function, with discount rate equal to the product of β and the belief in the falsity of p. (Note that this is mathematically equivalent to a model which postulates that conditional on $\overline{p}$, there is belief β in the presence of some other factor which *totally* discredits Bel_X. Substantively, however, the partial discounting model often makes more sense.)

Special Case 3: Discrediting subsets. Sometimes we do not wish to discredit an entire argument; rather, we find that the presence of a discrediting factor simply results in our being unable to distinguish two hypotheses we had thought we could distinguish.

In this case, the belief function Bel_X^c would be as illustrated in Figure A-3. Increasing belief in $\overline{p}$ then results in a proportionate decrease in belief in U and H, with the decrease being added to belief in {U,H}; and a proportionate decrease in belief in {A,U} and {A,H}, with the decrease being added to belief in {A,U,H}.

Basic Probability Assignment

Focal Element	Conditional on p	Conditional on $\overline{p}$
A	$m_X(A)$	$m_X(A)$
U	$m_X(U)$	0
H	$m_X(H)$	0
A,U	$m_X(AU)$	0
A,H	$m_X(AH)$	0
U,H	$m_X(UH)$	$m_X(U)+m_X(H)+m_X(UH)$
A,U,H	$m_X(AUH)$	$m_X(AU)+m_X(AH)+m_X(AUH)$

Figure A-3: Example of Subset Discrediting

Special case 4: Crediting It may be the case that we wish to make the provisional assumption that an argument is *unreliable* to some degree. In this case, belief in the falsity of a premise would serve to *increase* our confidence in the argument. We would then define the belief function $m^c_X(X) = 0$, and $m^c_X(A) = m_X(A)/(1-m^c_X(X))$ for proper subsets A. That is, $Bel^c_X(\cdot)$ is an *undiscounted* version of Bel_X.

In this special case, increasing belief in the absense of the premise would result in relative belief assignments for all proper subsets remaining the same, but the mass on the universal set decreasing.

Again, there is no reason why the falsity of a premise should *completely* remove uncommitted mass; we could have partial crediting in analogy to partial discrediting. Similarly, we could credit just a *subset*, reducing m(A) and allocating mass proportionately to subsets of A (where A is a proper subset of X).

The General Model: Multiple premises. Again, the general model can be extended to the casé of multiple premises relating to an argument. Here, though, there are complexities that did not arise in the case of simple discrediting. Let us say that the belief function Bel_X applies under premises p_1 and p_2. Invalidating p_1 leads to a new belief function Bel_X^1; invalidating p_2 leads to Bel_X^2. Now, though, what happens when *both* premises are invalidated? The answer was simple in the case of total discrediting: if invalidating either premise resulted in shifting all belief to the universal set X, then so would invalidating both premises. But in the case of general belief functions Bel_X^1 and Bel_X^2 the answer is not immediately obvious.

Let us step back to basic principles. In our original belief function Bel_X, assigning support $m_X(A)$ to the set A amounts to asserting, "the evidence underlying Bel_X means A with probability $m_X(A)$". Invalidating a premise underlying Bel_X would mean that *some* of the subsets we had thought the evidence might mean in actuality mean something else.

Thus, the analyst might select any number of pairs (A,δ_A) of subsets and discount rates, with the meaning, "reduce belief in A to $(1-\delta_A)m_X(A)$". The result will be a certain amount of probability that has been "removed" from subsets of X. (A default of $\delta_A = 0$ means removing no mass.) The user then specifies a "recipient belief function" $\mathrm{Bel}_X^r(\cdot)$ to which this belief is to be allocated proportionately among its focal elements. Thus, the belief function conditional on $\overline{p}$ is given by

$$m_X^c(A) = (1-\delta_A)m_X(A) + \alpha\ m_X^r(A), \quad \text{where} \quad A = \sum \delta_A m_X(A)$$

The interpretation is that there is probability $(1-\delta_A)m_X(A)$ that the evidence means A and probability α that the evidence means the belief function m_X^r.

How does this interpretation of the belief function Bel_X^c help us? Let us consider the case of two premises, p_1 and p_2. For each A we have a discount rate δ_A for p_1 and γ_A for p_2. We also have recipient belief functions Bel_X^r for p_1 and Bel_X^s for p_2.

As was the case for total discrediting, it seems reasonable to reduce belief in A to $(1-\delta_A)(1-\gamma_A)m(A)$. This results in belief

$$\alpha = 1 - \sum_{A}(1-\delta_A)(1-\gamma_A)m(A)$$

to be reallocated. How should this belief be reallocated?

We have evidence (i.e. $\overline{p}_1$) indicating that belief should be reallocated according to Bel_X^r and evidence (i.e. $\overline{p}_2$) indicating that belief should be reallocated according to Bel_X^s. The evidence for $\overline{p}$, is assumed independent of the evidence for $\overline{p}_2$; yet $\overline{p}_1$ and $\overline{p}_2$ themselves are not independent in their impact on Bel_X, since they are part of the same causal chain. The natural action, if the premises were independent in their impact, would be to reallocate according to the orthogonal sum (i.e. apply Dempster's rule). However, instead of taking intersections of subsets, as in Dempster's Rule, we take unions, in accordance with the causal structure discussed in Section 3.5. Thus, define Bel_X^+ as arising from the componentwise union of Bel_X^r and Bel_X^s. Then we have four conditional belief functions as described in the second column of Table A-4.

<u>Conditioning Premises</u>	<u>Conditional Basic Probability of A</u>	<u>Weight</u>
(p_1,p_2)	$m_X(A)$	$Pl_1(p_1)Pl_2(p_2)$
$(\overline{p}_1,p_2)$	$(1-\delta_A)m_X(A) + \alpha_1 m_X^r(A)$	$Bel_1(\overline{p}_1)Pl_2(p_2)$
$(p_1,\overline{p}_2)$	$(1-\gamma_A)m_X(A) + \alpha_2 m_X^s(A)$	$Pl_1(p_1)Bel_2(\overline{p}_2)$
$(\overline{p}_1,\overline{p}_2)$	$(1-\delta_A)(1-\gamma_A)m_X(A) + \alpha m_X^+(A)$	$Bel_1(\overline{p}_1)Bel_2(\overline{p}_2)$

Table A-4

Given independent belief functions over the premises, we combine them by Dempster's Rule. If both premises are assumed valid to the extent warranted by the evidence, the resulting belief function is a weighted average of the four belief functions in the above table, with weights given by the third column of Table A-4.

We may think of the $(1-\gamma_A)(1-\delta_A)m_X(A)$ as the portion of Bel_X which is unaffected by the premises, and which should therefore be "protected" from combination with respect to $\overline{p}_1$ and $\overline{p}_2$. That is, $\overline{p}_1$ and $\overline{p}_2$ constitute evidence as far as Bel_X^r and Bel_X^s are concerned, but not for the "protected" portion of Bel_X.

Premises affecting more than one belief function. To this point, it has been assumed that premises affect only one belief function. In general, this might not be the case; for example, incompetence of IAEA inspectors might discredit beliefs about reactor safety as well as materials diversion.

Let us begin by considering the simplest case: suppose we are combining two belief functions, Bel_1 and Bel_2, defined over the space X. Suppose that the premise p, if invalidated, would act to discount Bel_1 by discount rate δ_1 and Bel_2 by discount rate δ_2. Suppose that there is belief b in the falsity of premise p.

The combined belief function over X is obtained as follows. First we combine belief functions Bel_1 and Bel_2 by Dempster's Rule. It turns out that the computations are simpler if we carry them through without normalizing (normalization would serve to force all final basic probabilities to sum to 1). The basic probability assigned to the set A *before normalization* is given by

$$m_{12}(A) = \sum_{A_1 \cap A_2 = A} m_1(A_1)m_2(A_2). \qquad (4)$$

This is an unnormalized version of the combined belief function *conditional on p* To obtain the combined belief function conditional on $\overline{p}$, we combine the discounted belief functions m_1^c and m_2^c. Again before normalization, we obtain

$$\begin{aligned} m_{12}^c(A) &= \sum_{= A_1 \cap A_2 = A} (1-\delta_1)(1-\delta_2)m_1(A_1)m_2(A_2) + \delta_1 m_2(A) + \delta_2 m_1(A) \qquad (5) \\ &= (1-\delta_1-\delta_2+\delta_1\delta_2)m_{12}(A) \ + \ \delta_1 m_2(A) + \delta_2 m_1(A). \end{aligned}$$

The final marginal belief assignment is obtained by giving weight 1-b to (4) and b to (5).

$$m_M(A) = (1-b\delta_1-b\delta_2+b\delta_1\delta_2)m_{12}(A) + b\delta_1 m_2(A) + b\delta_2 m_1(A). \qquad (6)$$

(Note that (6) still represents beliefs before normalizing.)

If instead we had discounted Bel_1 by discount rate $b\delta_1$ and Bel_2 by discount rate $b\delta_2$ (acting as if two independent premises were operating, instead of a single premise), we would replace δ_i in Equation (5) by $b\delta_i$ to obtain

$$\tilde{m}_M(A) = (1-b\delta_1-b\delta_2+b^2\delta_1\delta_2)m_{12}(A) + b\delta_1 m_2(A) + b\delta_2 m_1(A). \qquad (7)$$

The only difference between (6) and (7) is that the term $b\delta_1\delta_2$ in (6) is replaced by the quadratic term $b^2\delta_1\delta_2$ in (7). This term becomes more important as the "worst case" discount rates δ_1 and δ_2 become closer to 1.

Thus, when the same premise affects different belief functions, we cannot discount the belief functions individually and then combine them. Rather, we combine them with no discounting, combine them again with maximum discounting (i.e. assuming p is false), and form a weighted combination of the two resultant belief functions.

Here we considered only a simple special case. In general, a given premise might apply to different belief functions at different points in separate lines of inference in a hierarchical structure. The same general principle of combination applies: combine conditional on p, combine conditional on $\overline{p}$, and form a weighted combination of these two belief functions.

Initiation and prioritization of the search for information. The nonmonotonic conflict resolution procedure in SED is invoked when there is conflict between items of evidence being combined. In Doyle's (1979) nonmonotonic logic, conflict is an all or nothing proposition--it occurs when both a proposition and its negation have been proven. In SED, there are gradations of conflict--two arguments conflict to the extent that they point toward contradictory hypotheses, even though the system cannot "prove" either hypothesis for certain. Thus, SED requires a measure of the *degree* of conflict between evidential arguments.

Fortunately, the Shaferian calculus provides a natural measure of conflict. When two or more belief functions are combined via Dempster's Rule, support is allocated to a hypothesis according to the extent that each individual belief function lends support to the hypothesis. To the extent that the belief functions support contradictory hypotheses, support is allocated to the null set, and the resultant belief function is then normalized so that belief in non-null sets sums to 1. This mass assigned to the null set provides a natural measure of the conflict between lines of reasoning. When belief functions $\mathrm{Bel}_1, \ldots, \mathrm{Bel}_n$ are combined, the *weight of conflict*, denoted symbolically by μ_c, is given by the formula

$$\mu_c = \sum_{A_1 \cap \cdots \cap A_n = \emptyset} m_1(A_1) \cdots m_n(A_n), \qquad (8)$$

where $m_i(A_i)$ is the basic probability assignment of belief function i to the set A_i.

When, during its process of combining evidence, SED encounters conflict greater than a threshold t_c, the conflict resolution mechanism is invoked. This "trigger" threshold can be set in advance, or adjusted dynamically by the user. Alternately, the user may reject the idea of a fixed threshold, preferring to examine the output of inferences and decide, in context, whether the level of conflict is acceptable.

The conflict resolution procedure is, in effect, a mechanism for reaching within the arguments leading to each of the five belief functions in an attempt to identify potential weaknesses in the arguments. When such weaknesses are identified, the corresponding belief functions are discounted, leading to reduction in conflict. As long as potential weaknesses can be identified, the process of conflict resolution continues until conflict is reduced to below t_c.

The first step of the conflict resolution procedure, the search for discrediting factors, is initiated when conflict exceeds the threshold t_c, and consists of the following five stages.

i. Decide which discrediting factor for which belief function is the provisional "culprit" and which test to perform on the culprit. (This step is the crux of the algorithm, and the selection criteria, based on a benefit/cost tradeoff, are described in detail below.) If no culprit can be found, move to Steps 2 and 3 of the conflict resolution procedure.

ii. Perform the test and revise belief in the appropriate discrediting factor.

iii. Compute a revised discount rate and apply it to the culprit belief function, resulting in a new belief function.

iv. Recombine the belief functions according to Dempster's Rule. The result is a new combined belief function, and a new measure of conflict.

v. If conflict is below t_c, stop. Otherwise, return to *i*.

The choice of which test to perform is based on a tradeoff between benefit and cost. Recall that each test has associated with it a cost of performance; this cost is to be traded off against benefit as measured by the potential for conflict reduction (the benefit measure is discussed in detail below). There are many possible tradeoff functions (e.g., a simple linear weighting of benefit and cost, or choosing the test with the highest benefit per unit cost, subject to a minimum benefit threshold). Alternatively, the system could present a list of the most promising tests (i.e., ones clearing thresholds on both benefit and cost), letting the user trade off informally.

The benefit of performing a test is to be defined as its potential for conflict reduction. This can be separated into two parts: the impact of the test result on the discount rate, and the impact of changing the discount rate on conflict. Thus, a test is beneficial to the extent that it has potential for increasing the discount rate, and to the extent that increasing the discount rate reduces conflict. We discuss the impact of discounting on conflict first.

The impact of changing the discount rate on conflict is measured by the partial derivative of conflict with respect to the discount rate. Mathematically, the reduction in conflict when a belief function is discounted is a linear function of the discount rate δ. To see this, rewrite Equation (8) as follows.

$$\mu_c(0) = \sum_{\substack{A_1\cap\cdots\cap A_n=\emptyset \\ A_1\neq X}} m_1(A_1)+\cdots+m_n(A_n) + m_1(X) \sum_{A_2\cap\cdots\cap A_n=\emptyset} m_2(A_2)\cdots m_n(A_n), \tag{9}$$

where we write $\mu_c(0)$ for the level of conflict when there is no discounting.

Now, suppose Bel_1 is discounted by the rate δ. This means that $m_1(A)$ is replaced by $(1-\delta)m_1(A)$ for all proper subsets A, and $m_1(X)$ is replaced by $(1-\delta)m_1(X)+\delta$. The new weight of conflict becomes

$$\mu_c(\delta) = (1-\delta) \sum_{\substack{A_1 \cap \cdots \cap A_n = \emptyset \\ A_1 \neq X}} m_1(A_1)+\cdots+m_n(A_n) + [(1-\delta)m_1(X)+\delta] \sum_{A_2 \cap \cdots \cap A_n = \emptyset} m_2(A_2)\cdots m_n(A_n),$$

which can be rewritten (combining terms) as

$$\begin{aligned} \mu_c(\delta) &= (1-\delta) \sum_{A_1 \cap \cdots \cap A_n = \emptyset} m_1(A_1)\cdots m_n(A_n) + \delta \sum_{A_2 \cap \cdots \cap A_n = \emptyset} m_2(A_2)\cdots m_n(A_n) \\ &= (1-\delta)\mu_c(0) + \delta \sum_{A_2 \cap \cdots \cap A_n = \emptyset} m_2(A_2)\cdots m_n(A_n) . \end{aligned} \tag{10}$$

This function is linear in the discount rate δ.

The partial derivative of μ_c with respect to δ is then

$$\frac{\partial \mu_c}{\partial \delta} = -\mu_c(0) + \sum_{A_2 \cap \cdots \cap A_n = \emptyset.} m_2(A_2)\cdots m_n(A_n), \tag{11}$$

In words, this is the difference between the conflict when all (undiscounted) belief functions *except* Bel_1 are combined, and the conflict when all (undiscounted) belief functions are combined. This derivative is always negative.

The other aspect of benefit is the test's potential for increasing the discount rate. The change in the discount rate will depend on which test result occurs. Benefit must therefore be defined in terms of the outcome of an uncertain event.

Information about which test result will occur is expressed by a belief function over test results. This belief function will be vacuous if there is no such information. We may interpret the basic probability m(R) for a set R of test results as the probability that our evidence means the result will be in R. Suppose there were some number $\delta(R)$ such that belief in R justified belief in the discount rate $\delta(R)$. We could then say our evidence justified belief in the discount rate

$$\delta(t) = \sum \delta(R)m(R) \qquad (12)$$

for the test t.

Unfortunately, R will in general contain many test results, each with a different discount rate. There is no *unique* way to define a discount rate implied by belief in R.

We may, however, adopt one of several possible *decision attitudes*. Each of the following decision attitudes has implications for how $\delta(R)$ is chosen.

1. *Pessimism.* Belief in R should justify belief in a discount rate of *at least* $\delta(R)$. We therefore define $\delta(R)$ as the minimum discount rate associated with a result in R.
2. *Optimism.* Belief in R should imply a discount rate that could be as high as $\delta(R)$. We therefore define $\delta(R)$ as the maximum discount rate associated with a result in R.
3. *Conservatism.* To the extent that there is no information to distinguish them, all results should be treated equally. We therefore define $\delta(R)$ as the average discount rate associated with a result in R.

We are now ready to define the benefit measure. The test t tests for presence of some discrediting factor associated with belief function Bel_i. The initial discount rate for Bel_i is δ_i. The derivative of the conflict μ with respect to δ_i is μ_i. Assuming one of the above decision attitudes, the discount rate associated with test t is $\delta(t)$, as defined in Equation (12). The benefit of performing test t is then defined as

$$b(t) = -(\delta(t) - \delta_i)\mu_i. \qquad (13)$$

The negative sign occurs because conflict varies inversely with discount rate (i.e., $\mu_i<0$.)

The benefit of performing test t can be thought of as the value of acquiring the information contained in the test result. Thus, we can think of Equation

(13) as an analogue to a Bayesian value of information computation. Note that the goal here is conflict reduction. We might argue that conflict reduction is not itself of intrinsic worth--we wish to reduce conflict in order to improve the quality of inferences, which, in turn, might be used to make decisions that affect an ultimate goal. Should we not then compute the value of information in terms of this ultimate goal? We respond by noting that this computation would be exceedingly complex. At any rate, the intelligence analyst process does not always have full access to the policy decisions that will be effected by A. Moreover, is not the abovementioned "ultimate goal" itself merely a subgoal of some still higher goal? Slicing the problem so that benefit is defined in terms of conflict reduction has the advantage of modularity and simplicity.

A second comment is the absence of a *unique* way of measuring value of information (or any other kind of expectation) in a Shaferian system. Each of the decision attitudes described above corresponds to a rule for allocating the uncommitted belief so as to collapse the belief function into a probability distribution. Pessimism corresponds to allocating all uncommitted mass to the lowest discount rate consistent with the evidence. Optimism corresponds to allocating it to the highest rate. Conservatism corresponds to portioning uncommitted mass equally among all compatible test results.

Benefit is traded off against the cost $\gamma(t)$ of performing test t. One possible tradeoff function is

$$u(t) = \begin{cases} 0 & \text{if } b(t) < \text{threshold}_{\text{benefit}} \\ b(t)/\gamma(t) & \text{otherwise .} \end{cases} \tag{14}$$

If a test is found for which $u(t) > \text{threshold}_{\text{utility}}$, then NMP performs the one for which u(t) is maximized (i.e., for which benefit per unit cost is maximized). Otherwise, no test meets criteria. This could happen because no test has the potential for significant conflict reduction, because all tests cost too much to perform, or because all possible tests have already been performed.

The second phase in the conflict resolution procedure is to search for additional information (other than discrediting factors). This could include testing for *credibility factors* whose presence would *increase* (rather than decrease) the strength of an argument or change its overall direction. This could also mean collecting unrelated evidence (as represented by a new belief function to be combined with the other belief functions).

To prioritize the search for information, a measure of the impact of the new evidence is needed. In the case of credibility factors, this is a straightforward extension of measuring the impact of discrediting factors (since credibility factors are generalized discrediting factors). In the case of independent evidence, we require a model of beliefs about the impact of the evidence.

Typically, the result of this phase will be an *increase* in conflict, but we hope that the information will help us discover which of the component arguments is most likely to be flawed. Thus, our criterion in evaluating the value of the new information will be its ability to discriminate competing arguments.

This could be operationalized by clustering the belief functions into groups, such that within-group conflict is low and between-group conflict is high. New information would be sought that was judged likely to provide information that distinguished these groups. That is, we seek information that would be expected to keep within group conflict low. We would expect (and therefore, we would not penalize) a concommitant increase in between group conflict.

Information search. Thus far, we have considered only conflict as a possible trigger for the search for information about the presence of credibility factors. Other triggers are possible. One important situation concerns the case when a final belief function has too much uncommitted mass. In other words, we are unable on the basis of current evidence to arrive at a definite conclusion. This event might trigger the search for crediting premises (or, perhaps more commonly, for additional independent evidence). Similarly, corroboration between two sources (i.e. *more* agreement than we had expected) might trigger the search for crediting factors (i.e. evidence that the arguments were *more* reliable than we had thought).

Implications for belief function framework. We noted in Section 2 that a serious shortcoming of the Shaferian approach was the inability to reassess the quality of a line of evidential reasoning based on the content of the evidence (i.e. on whether sources corroborate or contradict each other). The credibility factor model gives us the tools for overcoming this shortcoming.

Suppose we have two belief functions Bel_1 and Bel_2, which represent our current assessment of the impact of two lines of argument. Suppose we judge that one or both of them are affected by certain premises, but we have no evidence that they are invalid.

Now, suppose we combine Bel_1 and Bel_2 by Dempster's Rule, and suppose they are in conflict. The conflict triggers the search for evidence against underlying premises. Such evidence would, if found, downgrade the assessed reliability of the arguments. Similarly, excess corroboration might trigger the search to invalidate other premises to upgrade reliability.

As described here, the content of the evidence (i.e. conflict or corroboration) served only as a *trigger* for the search for new evidence. It is this new evidence, and not the content of the original evidence, that changes our estimate of the reliability of the sources. Thus, we maintain the independence assumption upon which Dempster's Rule is based.

There may be times when we cannot find "hard" evidence against a premise, but the existence of conflict leads us to question the reliability of our sources. The overall discounting phase of conflict resolution provides a case in point. We assign a premise to each argument, termed "conflict with other arguments is not too great". When there is significant conflict, belief in this premise decreases and all arguments are discounted.

This procedure can be explained within the language of non-monotonic logic. In a non-monotonic system, there are certain default assumptions (e.g. the reliability of certain arguments). When conflict renders these become inoperative, and the system typically specifies alternate assumptions (in our system, alternate allocations of uncommitted belief) to replace them. Thus, using conflict to revise beliefs (absent "hard" evidence) amounts to replacing one default assumption about how to allocate belief in $\{p,\overline{p}\}$ (allocate all of it to p) with another (allocate some of it to $\overline{p}$).

APPENDIX B: HIERARCHICAL INFERENCE IN A BELIEF FUNCTION MODEL

B.1 A Single Chain of Evidence

It is often the case that an evidential argument proceeds in a series of steps. For example, we receive a report that leaves have been canceled for Warsaw Pact troops. The report lends support to the hypothesis that leaves have in fact been canceled, the degree of support depending on the credibility of the source. The event that leaves have been canceled lends support to the hypothesis that Warsaw Pact personnel are in a state of readiness, which, if true, lends support to the hypothesis that Warsaw Pact troops have been mobilized.

Shafer's theory of belief functions provides a natural representation of the idea of evidential support. For example, the degree to which cancellation of leaves supports personnel readiness would be represented by a belief function over the possible degrees of readiness, conditional on the event that leaves have been canceled. Such conditional belief functions can be specified at each stage of the hierarchy. This section explains how to "chain up" the hierarchy to obtain a belief function over the set of hypotheses at the terminal point of the chain. Also considered are the assumptions underlying the procedures.

Figure 1 illustrates a chain of argument linking observed evidence e (e.g., the report of leave cancellation) to a set H of hypotheses of interest (e.g., whether or not Warsaw Pact troops have been mobilized). The evidence e and hypotheses H are linked through a number of sets of intermediate hypotheses (X, Y, ..., Z). The idea of a *chain* is formalized as follows. The evidential link between consecutive sets in the chain is formalized by a *conditional belief function* (e.g., $\mathrm{Bel}_{Y|X}$ in Figure 1). These conditional belief functions represent degrees of belief over Y implied by each element $x \epsilon X$. We also have a bottom-level belief function Bel_X, representing the implication of the evidence e for X.

The validity of the methodology to follow depends on the following assumptions.

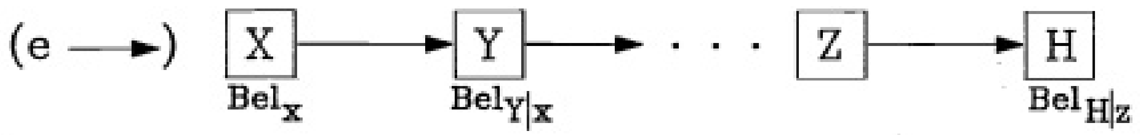


Figure 1: A Single Chain of Evidence

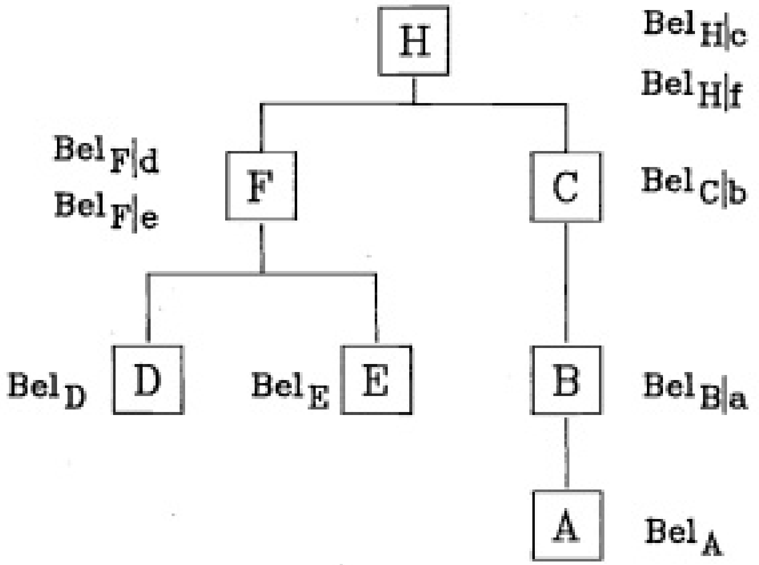


Figure 2: An Example of a Hierarchial Inference Structure

1. The evidential link between any two adjacent sets is completely described by the conditional belief functions linking those two sets.

2. The evidence e affects hypotheses further up in the chain only through its impact on belief in X.

3. Each set of intermediate hypotheses affects belief in hypotheses further along in the chain only through its relation to the set of hypotheses immediately following it.

4. All of the conditional belief functions are based on independent evidence, and on evidence independent of that on which the lowest level belief function, Bel_X, is based.

To illustrate the methodology, we consider a chain of length 2 (i.e., $X\rightarrow Y$). We are given a belief function Bel_X over X, and, for each $x\epsilon X$, a conditional belief function $Bel_{Y|x}$ over Y. Of interest are the implied beliefs for hypotheses in Y.

When belief functions are based on independent evidence, Dempster's Rule is the natural combination tool. In order to incorporate the *link* between the sets X and Y, we need to consider the product space $X\times Y$. Our first step is to extend Bel_X to $X\times Y$, using the minimal extension (meaning that our evidence about X tells us nothing directly about Y). Each focal element A of Bel_X is extended to a new focal element $A\times Y$, with the same belief $Bel_X(A)$ and basic probability assignment $m_X(A)$.

The second step is to embed the conditional belief functions into the product space. We start by creating a new belief function over $X\times Y$ for each $Bel_{Y|x}$. Each focal element $B_x\subset Y$ is associated with a new focal element $(\{x\}\times B_x)\cup(\{\bar{x}\}\times Y)$ of the new belief function. The new focal element is assigned belief $Bel_{Y|x}(B_x)$ and basic probability $m_{Y|x}(B_x)$. It represents belief in B_x if x is the case, but no information about Y otherwise.

Combining the new belief functions by Dempster's Rule gives the *conditional embedding* (Shafer, 1982) of the $Bel_{Y|x}$ in $X\times Y$. For each choice of focal elements B_x of the $Bel_{Y|x}$, there is an associated focal element of the conditional embedding, given by $\bigcup_{x\epsilon X}(\{x\}\times B_x)$. The basic probability corresponding to this focal element is $\prod_{x\epsilon X} m_{Y|x}(B_x)$.

Third, the conditional embedding of $Bel_{Y|x}$ in Y×X is combined via Dempster's Rule with the extension of Bel_X to X×Y. The focal elements of the combined belief function can be described as follows. For each focal element A of Bel_X and each choice of focal elements B_x of the $Bel_{Y|x}$, there is a corresponding focal element $C = \bigcup_{x\epsilon A}(\{x\}\times B_x)$. Its basic probability assignment is the sum, over all A and B_x giving rise to the set C, of $m_X(A) \prod_{x\epsilon X} m_{Y|x}(B_x)$.

The final step is to marginalize over Y. The focal elements of this new belief function (which we shall call Bel_Y) have the form $D = \bigcup_{x\epsilon A} B_x$, for some choice of A and the B_x. The basic probability assignment is given by the sum, over all A and B_x giving rise to D, of $m_x(A) \prod_{x\epsilon X} m_{Y|x}(B_x)$. In other words, belief in focal element D of the marginal belief function Bel_Y is given by

$$m_Y(D) = \sum_{\bigcup_{x\epsilon A} B_x = D} m_X(A) \prod_{x\epsilon X} m_{Y|x}(B_x). \qquad (1)$$

It is clear that the marginal belief function Bel_Y can now be used as input to the next link up the chain. We can proceed in this manner up the chain until a belief function over the final hypothesis space H is obtained.

In fact, under the assumptions given above, this is the correct thing to do. Consider a three-stage chain, X→Y→Z. The following theorem establishes the equivalence of chaining up one step at a time and working in the full product space X×Y×Z.

<u>Theorem</u>: Consider the belief functions Bel_X, $Bel_{Y|x}$ $(x\epsilon X)$, and $Bel_{Z|y}$ $(y\epsilon Y)$. Suppose the belief function Bel_{XY} over X×Y is formed as above, by

a. forming the minimal extension of Bel_X into X×Y;

b. forming the conditional embedding of $Bel_{Y|x}$ into X×Y;

c. combining by Dempster's Rule.

Then the following procedures yield equivalent marginal belief functions over Z.

Procedure 1 (chaining up):

d. form the marginal Bel_Y of Bel_{XY};

e. form the minimal extension of Bel_Y into Y×Z;

f. form the conditional embedding of $\mathrm{Bel}_{Z|y}$ into Y×Z;

g. combine by Dempster's Rule to form Bel_{YZ};

h. form the marginal of Bel_{YZ} over Z.

Procedure 2 (working over the product space):

d'. form the minimal extension of Bel_{XY} into X×Y×Z;

e'. form the conditional embedding of $\mathrm{Bel}_{Z|y}$ into Y×Z;

f'. form the minimal extension of the conditional embedding into X×Y×Z;

g'. combine by Dempster's Rule to form Bel_{XYZ};

h'. form the marginal of Bel_{XYZ} over Z.

Proof: Procedure 1: As derived above (Equation (1)), the belief function Bel_Y has, for each choice of A, B_x:

Focal element $$D = \bigcup_{x \epsilon A} B_x$$

Belief $$m_Y(D) = \sum_{\bigcup_{x \epsilon A} B_x = D} m_X(A) \prod_{x \epsilon X} m_{Y|x}(B_x).$$

Now, for each choice of focal elements F_y of $\mathrm{Bel}_{Z|y}$ and D of Bel_Y (which, in turn, was derived from some choice of A, B_x focal elements of Bel_X and $\mathrm{Bel}_{Y|x}$), the marginal belief function Bel_Z (formed by combining Bel_Y and $\mathrm{Bel}_{Z|y}$, and marginalizing) has

Focal element $$G = \bigcup_{y \epsilon D} F_y = \bigcup_{x \epsilon A} \bigcup_{y \epsilon B_x} F_y$$

(for some particular choice of A, B_x, F_y);

Belief $m_Z(G) = \sum_{\bigcup_{y\epsilon D} F_y = G} m_Y(D) \prod_{y\epsilon Y} m_{Z|y}(F_y)$

$$= \sum_{\bigcup_{y\epsilon D} F_y = G} \sum_{\bigcup_{x\epsilon A} B_x = D} m_X(A) \prod_{x\epsilon X} m_{Y|x}(B_x) \prod_{y\epsilon Y} m_{Z|y}(F_y)$$

$$= \sum_{\bigcup_{x\epsilon A} \bigcup_{y\epsilon B_x} F_y = G} m_X(A) \prod_{x\epsilon X} m_{Y|x}(B_x) \prod_{y\epsilon Y} m_{Z|y}(F_y). \quad (3)$$

Procedure 2: Belief function Bel_{XY} has, for each choice of A, B_x:

Focal element $\bigcup_{x\epsilon A} (\{x\}\times B_x)$

Belief $m_X(A) \Sigma \prod_{x\epsilon X} m_{Y|x}(B_x)$,

where the sum is over all choices of B_x yielding the same focal element (i.e., B_x for $x\epsilon A$ does not affect the focal element).

Extending this to X×Y×Z gives

Focal element $\bigcup_{x\epsilon A} (\{x\}\times B_x\times Z)$,

with the same belief. The conditional embedding of $\text{Bel}_{Z|y}$ has, for each choice of F_y focal elements of $\text{Bel}_{Z|y}$,

Focal element $\bigcup_{y\epsilon Y} (\{y\}\times F_y)$

Belief $\prod_{y\epsilon Y} m_{Z|y}(F_y)$.

Extending to X×Y×Z gives

Focal element $\bigcup_{y\epsilon Y} (X\times\{y\}\times F_y)$,

with the same belief. Now, combining by Dempster's Rule requires intersecting these, to yield

Focal element $\bigcup_{x\epsilon A} \bigcup_{y\epsilon B_x} (\{(x,y)\}\times F_y)$. (4)

Belief in this focal element is the sum of the products of belief in focal elements that intersect to form the given focal element (4). The belief is the sum of terms of the form

$$m_X(A) \prod_{x\epsilon X} m_{Y|x}(B_x) \prod_{z\epsilon Z} m_{Z|y}(F_y),$$

where the sum is over all choices of A, B_x, F_y giving rise to the focal element (4). Finally, we marginalize, to get belief function Bel_Z', with

Focal element $G' = \bigcup_{x\epsilon A} \bigcup_{y\epsilon B_x} F_y$

Belief $m_Z'(G) = \Sigma\, m_X(A) \prod_{x\epsilon X} m_{Y|x}(B_x) \prod_{y\epsilon Y} m_{Z|y}(F_y)$,

where again the summation is over all A, B_x, F_y giving rise to focal element G. Clearly these match the focal element (2) and belief (3) from Procedure 1, so the two procedures are equivalent.

It is clear from this analysis that a single chain of evidential argument would be evaluated by "chaining up," one level at a time, until the final output, the implication of the evidence for the top-level hypothesis, is obtained.

A.2 A Hierarchical Tree of Evidence

Figure 2 is an example of the kind of evidential structure commonly found in problems of hierarchical inference. Given some evidence about the low-level events (A, D, E) and some evidence relating events to those immediately above them in the tree, the problem is to find the implications for the top-level hypothesis set H.

Following the ideas in the previous section, we express the impact of the evidence on knowledge about low-level events as belief functions (Bel_A, Bel_D, Bel_E). The relation between adjacent intermediate level events is expressed as conditional belief functions ($Bel_{B|a}$, $Bel_{C|b}$, etc.). The analysis outlined here requires certain independence assumptions. Assumptions 1-4 must be satisfied within each single chain (e.g., D→F→H). Moreover, separate arguments for a given conclusion (i.e., lines of reasoning that do not share a common hypothesis space except for the conclusion) are based on independent evidence (i.e., D→F→H and E→F→H are not independent because they share the argument from F to H, but each is independent of A→B→C→H. Moreover, D→F and E→F are independent).

The procedure for analyzing an argument such as that given in Figure 2 is as follows. First, chain up each argument individually until encountering a node shared with another line of argument. Then combine, using Dempster's Rule, all belief functions for a given node. Continue chaining up and combining until the top-level node is reached.

For example, in Figure 2, we would propagate the belief function Bel_D through the chain D→F to obtain belief function Bel_{F1} over F. Then propagate Bel_E through E→F to obtain Bel_{F2} over F. Combine by Dempster's Rule, obtaining Bel_F. Propagate through F→H to obtain Bel_{H1} over H. Now, propagate Bel_A through A→B→C→H to obtain Bel_{H2}. Combine these using Dempster's Rule to obtain the final belief function Bel_H.

APPENDIX C: UNCERTAINTY IN DATABASE SYSTEMS

In this section, we turn to aspects of currently existing database technology which contribute to the design of an evidential database for intelligence analysts. In the course of this discussion, we ask what methods exist or have been proposed to represent incomplete or uncertain data. Although the most prominent database formats and query languages in *use* today offer little or nothing to assist users in representing uncertainty, nevertheless, considerable technical attention is being given to the problem, and a variety of apparently viable approaches have been defined.

The situation is quite different with respect to a second question: what methods exist or have been proposed to represent the *evidential arguments* that underlie such representations of uncertainty? Few researchers have addressed the issues of model building and structuring of evidential arguments that are essential to the intelligence analyst's *understanding* of his data. To have the capability of representing, storing, and retrieving numerical probabilities (or possibilities, or degrees of belief) corresponding to different hypotheses (e.g., about the identity of the perpetrators of a terrorist act) is of some use. Of far greater value, we believe, is the ability to represent, record, and retrieve the reasons and assumptions upon which those assessments depend, and to manipulate such reasons and assumptions in a variety of ways to resolve conflict and construct a convincing evidential argument.

Our discussion is in two parts: (1) a brief review of the relational model of data and alternative ways of organizing data (network, hierarchical); and (2) recent efforts to modify or extend the relational system to include representations of uncertainty.

C.1 <u>The Basic Relational Model</u>

A number of different ways of organizing large databases have been developed, each with its advantages for certain classes of problems and disadvantages for others. In theory, structure is a crucial issue, affecting design of the data manipulation language, including the types of inferential and uncertainty-related queries that can be posed. In practice, selection of a data structure may make implementation of a particular user operation easier or harder, but never makes it impossible.

Most prominent among representation schemes at the user level are hierarchical, network, and relational approaches. The relational approach is by far the best developed, and the one for which there has been the most work on adaptations to deal with incomplete information. In the next subsection, we examine modifications to the relational approach that allow representation of incomplete information. As a prelude, we present an overview of the relational approach.

In a relational system, the fundamental objects are two-dimensional tables known as *relations*. By convention, the columns correspond to the *attributes*, or dimensions along which the data are characterized; and the rows are the *records*, corresponding to the different objects or occurrences being modeled in the database. An attribute or set of attributes whose values uniquely identify the records in a relation may be used as a *key*. All records are assumed to be distinct; no duplicates are allowed.

To be useful, a database must not only store information, it must permit access and manipulation of the information. The user communicates with the database management system by means of an *access language*. The access language allows the user to manipulate relations to form other relations, and to pose *queries* to the database management system.

The chief advantage of the relational model is that the access language can be data independent: the language need only address the general operations on relations and their elements, not details of data storage. The access language may be defined and viewed in different ways, but a common approach is to construct an algebra based on five primitive operators. These five operators form new relations from existing relations. The first three are the usual set theoretic operators, *union, set difference*, and *cartesian product*. Union and set difference must be restricted only to relations that are *union compatible*, that is, relations having attributes which can be placed in one-to-one correspondence with each other, such that each attribute in one relation has the same domain as the corresponding attribute in the other relation. Cartesian product is unrestricted in application. The fourth operator, *projection*, forms a new relation by dropping all but a specified set of attributes, then removing duplicate rows. Finally, the *selection* operator selects all records for which the value of a specified logical expression has truth value T (e.g., select all records for which AGE takes on values between 40 and 65). Based on these five operators, other useful ones can be defined. A common one is *join*, which concatenates rows of one relation with rows of another, based on a logical expression relating an attribute of each. An example of the *natural join* operator is shown in Figure 2-1 (other joins are possible, based on logical comparisons other than equality).

We might think of any of these five operators as a type of query, but selection most closely resembles the intuitive meaning of the term query. Thus, the kinds of queries supported by a relational database system include questions such as, "find all records for which country of origin is France," or "for which individuals is the city of birth the same as the city of residence?" Selection followed by projection allows the following kind of query: "Show me the employee number, salary, and length of service for all employees in the chemistry department." A final type of query results not in a relation (or set of records), but in a truth value (T or F). An example of such a "yes-no" query is, "are there any contacts who are fluent in Russian and live in Baltimore?"

C.2 Incomplete Information in the Relational Model

Uncertainty in its various forms is ubiquitous in data taken from the real world. Nevertheless, until recently database theorists have devoted little attention to studying the representation of uncertainty in databases and the special requirements it imposes on structure and information access. This lack of attention has been reflected in database software products, which typically do not provide support for incomplete, ambiguous, or imprecise data.

Within the relational model, some kinds of uncertainty are handled more easily than others. Most authors make the assumption that the number of attributes, the attributes themselves, and the attribute domains are known precisely. (That is, we know that there are 25 attributes, that one of the attributes is SALARY, and that the domain of SALARY is the set of nonnegative integers). Thus, the only allowable kind of missing information concerns the *values* of attributes for particular records. Because the key field is supposed to be a unique identification field for a record, it is assumed that the value of the key is known (this restriction applies only to *base relations*; relations derived by means of operators such as projection may contain uncertain values in the key). Implicitly, therefore, it is also assumed that the number of objects (records) and their identify are not uncertain.

C.2.1 Null values. Codd (1979) develops extensions of the relational algebra to deal with *null values*, where a null value (denoted by #) is taken to mean "value at present unknown." Given that null values are allowed to exist, the data access language must be extended to deal with them. Specifically, a means must be provided for determining the truth values of expressions containing null values, and the operators must be redefined so that they operate on relations containing null values.

Codd defines a three-valued logic for determining the truth value of expressions containing nulls. The value of any logical expression may take on one of three values: T (true), F (false), or # (unknown). Truth tables for logical operators, as well as rules for set inclusion, are based on the *null substitution principle*. The null substitution principle states that an expression has the value # if and only if substituting (possibly distinct) non-null values for each occurrence of # in the expression can result in either the value T or the value F. Thus, the expression $\#\vee T$ has the value T, because substituting either T or F for # yields T. On the other hand, the expression $\#\wedge T$ has the value #, because substituting T yields T, but substituting F yields F.

The five operators of the relational algebra are extended based on the null substitution principle and on an extension of the rule for removing redundant rows. For purposes of removing duplicate rows, a null in one row is considered equal to a null in another (in contrast to a truth value of # for the expression # = #. With these rules, the extensions of union, set difference, cartesian product, and projection are well-specified. The select operator, however, depends on the truth value of a logical expression. Two extensions are possible, one that selects only those records for which the expression evaluates to T, and another that selects only those for which the expression does not evaluate to F. Codd resolves this indeterminacy by defining "true" and "maybe" versions of the select operator. (His maybe version selects *only* those records for which the expression evaluates to #; selecting non-F records requires performing a true select followed by a maybe select.)

C.2.2 Lipski's information system. In Codd's system, we either know the value of an attribute exactly or we know nothing about it. Frequently, however, an intermediate level of information is available. We now turn to one kind of generalization of Codd's relational algebra with null values. Lipski (1979) considers the case where the value of each attribute is known to lie within a specified subset of the attribute domain. A known value corresponds to a singleton, and Codd's null value corresponds to the entire domain. Lipski's system allows knowledge intermediate between these two extremes.

The only kind of incomplete information that can be represented in Lipski's system is knowledge that a given attribute of a given record is constrained to lie within a subset of the attribute domain. In particular, information such as TAX < SALARY cannot be represented (unless, of course, one of TAX or SALARY is known exactly, in which case this information would be equivalent to constraining the other to lie within a subset). Moreover, an attribute value either is possible or it is not--there is no way of representing *degrees* of possibility.

Lipski is concerned with developing a logic for queries in a database consisting of a single relation. He is not concerned with the operators for forming new relations out of existing relations (except, as we shall see, for select). For Lipski there are two kinds of queries, corresponding, in our language, to the select operator (find all records satisfying a given property) and to yes-no queries (does the following property hold?). Intersection and union can also be expressed in Lipski's language (as AND and OR, respectively). However, projection, division, and cartesian product cannot be expressed.

For every object and every attribute, then, the system stores not a unique value, but a subset consisting of all the values which the attribute *could* assume for the given object. Lipski is concerned with extending the usual logical definitions for a query language to this case of incomplete information.

There are two natural responses to a query in an incomplete information system. For selection queries, the first is to list the set of objects for which the property is *known* to hold; the second is to list the set of objects for which the property *might* hold, or for which the property is not known not to hold. These two responses define respectively the *lower* and *upper* values of a selection query. For yes-no queries the lower and upper values are defined analogously. Codd also considers a third query response, corresponding to what he calls the "internal interpretation" of a query.

Lipski defines a modal logic for dealing with queries in an incomplete information system. The "possible worlds" of Lipski's modal logic are the possible *completions* of the database. A completion of a database is a complete information system (i.e. all attribute values are definite) formed by selecting, for each record and each attribute, one of the set of possible values for the attribute. The lower value of a selection query is the set of all objects for which the query is satisfied in *all* possible completions of the database; the upper value is the set of records for which the query is satisfied for *some* completion of the database.

Thus, Lipski's system allows the formulation of queries such as "list all sources who are known to have contacted us within the last five years," or "list all employees who possibly earn more than $30,000." Indeed, the capability for an internal interpretation of queries allows combinations such as, "list all contacts who are known to be fluent in Russian and who may have visited the Soviet Union."

For atomic values, the internal interpretation is the same as the lower value. For a compound query, such as, finding all red or blue objects, there is a difference. The lower value would retrieve all objects whose color is known to lie in the set {red, blue}. In the internal interpretation, the lower values of the atomic queries are processed as in ordinary logic. Thus, the result would be blue. (The name "internal interpretation" comes about because the response is a statement of the system's internal knowledge, rather than about the external world being modeled).

Lipski's logic is not truth-functional (i.e. the truth of an expression cannot be determined solely from the truth values of its components). He notes a paradoxical element in Codd's truth-functional three-valued logic. Consider, for example, the expression $p \wedge \neg p$, where the value of p is #. In Codd's logic, the truth value is unknown, as is the truth value of the conjunction of *any* two null-valued expressions. Nonetheless, we know that p and not-p cannot both be true. Thus, argues Lipski, the expression $p \wedge \neg q$ should have truth value # when the expressions p and q are unrelated, but when q=p, the value of the expression should be F. In general, Lipski recommends modifying the null substitution principle in the following way: when substituting for null values in an expression, always substitute the *same* value for different occurrences of the same variable (distinct values may be given to distinct variables).

The fact that the value of an expression cannot be determined by the truth values of its components presents problems for computing query responses. Fortunately, Lipski presents theorems by which arbitrary expressions can be transformed into canonical forms from which the upper and lower values of queries can be determined from the upper and lower values of the component expressions.

C.2.3 Fuzzy constraints on attribute values. Prade (1984) generalizes Lipski's approach using ideas from Zadeh's possibility theory. In Lipski's system attribute values are constrained to lie in *crisp* subsets of the attribute domain. That is, a given value either is or is not possible--there is no room for *degrees* of possibility. In Prade's system, although the attribute domains are crisp sets, attribute values are *fuzzy* subsets of the attribute domain. Thus, Prade allows for differing *degrees* of possibility for different attribute values.

As with Lipski, Prade is concerned with a query language for a database system with incomplete information. Since he is concerned only with a single relation, he does not consider the full relational algebra (i.e. cartesian product or join). Rather, he considers the problem of how to respond to select or yes-no queries when information about attribute values is fuzzy.

Prade's fuzzy relational database consists of a single relation, with rows corresponding to records and columns corresponding to attributes. The value of attribute i for object x is a *possibility distribution* $\pi_{i(x)}$ which maps elements in the attribute domain D_i into the interval [0,1]. A value of 1 for $\pi_{i(x)}(u)$ means that u is undoubtedly possible as a value for attribute i on object x; a value of zero means it is impossible; and intermediate values mean intermediate degrees of possibility. (Note that $\pi_{i(x)}(u) = 1$ does *not* mean that u *is* the value; it simply means it is definitely a possible value.) If an attribute's value is known to lie in a crisp (i.e. ordinary) subset, then $\pi_{i(x)}$ is the characteristic function of that subset (equal to 1 for elements of the subset, and zero otherwise). Thus, Lipski's representation is a special case of Prade's.

The formulation of incomplete information in terms of possibility theory allows expression of a different kind of incomplete information than "value at present unknown." By specifying $\pi_{i(x)} = 0$, we may represent information of the form, "attribute i does not apply to object x," (since all values in the attribute domain have possibility zero). Prade requires that if an attribute applies to a given object, then there exists at least one value that is undoubtedly possible (i.e. the maximum value of $\pi_{i(x)}$ is either 1 or 0). This rules out the situation where attribute i is only "partially applicable" to object x.

Prade extends Lipski's results on the lower and upper values of queries to the case of fuzzy information. For a selection query, Lipski's upper value is the (crisp) set of those objects which are not excluded from having the property; in Prade's system, the upper value is a possibility distribution for the fuzzy set of responses to the query. Correspondingly, Lipski's lower value is the (crisp) set of those objects which are known to satisfy the query; in Prade's system, it is the *necessity distribution* corresponding to the above-mentioned possibility distribution (the necessity measure is defined as 1 minus the possibility of the complement, and takes on the value 1 if and only if the object *must* satisfy the query).

Prade's system has several advantages over Lipski's. First, he claims that the min/max operators are easier to manipulate than the "possible completions" semantics of Lipski's modal logic. Second, Prade's system can handle more complex types of dependencies than can Lipski's system. Any kind of fuzzy relation between two or more attributes can be modeled (including the possibility that for some values of attribute i, attribute j may be inapplicable). Finally, Prade's system supports a more expressive query language. Specifically, one may ask not just for objects whose values are known to be possible ($\pi_{i(x)} = 1$), but for objects whose value are at or above a specified level of possibility.

Prade assumes that possibility distributions are already specified as numerical quantities; his system manipulates these numerical possibility distributions to compute possibility and necessity measures as query responses. Prade does not consider the translation of linguistic terms (e.g. "tall" man) into possibility distributions (e.g. a possibility distribution for the HEIGHT attribute with possibility values $\pi(5'4") = 0$; $\pi(5'10") = .5$; $\pi(6'3") = 1.0$).

C.2.4 A fuzzy relational algebra. On the surface, Buckles and Petry (1982) would seem to be addressing the same problem as Prade: generalization of Lipski's framework to incorporate fuzzy information about attribute values. However, the authors actually address different aspects of the incomplete information problem. Prade generalizes Lipski to allow attribute values to be fuzzy (not crisp) subsets of the attribute domains. He is concerned with responding to selection and yes-no queries when attribute values are fuzzy sets. In contrast, Buckles and Petry retain Lipski's requirement that attribute values be modeled as crisp sets. They are concerned not with a query language, but with a relational algebra consisting of commands such as projection, union, intersection, and join. Their use of fuzzy set theory is the introduction of a fuzzy similarity relation on the cross product of the attribute domain. The similarity relation is used for the operation of removing redundant tuples from the database. For purposes of removing redundant tuples, two attribute values are considered equivalent if their similarity exceeds a user-specified threshold. They require the similarity relation to satisfy a form of transitivity in order to ensure uniqueness in the resultant relation.

Buckles and Petry state that operators are allowed to have a conditional clause (i.e. form the intersection of RELATION1 and RELATION2, where DEPARTMENT=CHEMISTRY). However, they do not consider the issue of the truth value of expressions containing fuzzy-set-valued attributes, and indeed, all their examples of conditional expressions involve attributes with known values.

C.2.5 Another approach to a fuzzy query language. Another approach to incorporating fuzzy logic into relational database theory is developed by Zvieli, Zvieli criticizes Buckles and Petry because of the restriction that attribute values must be described by crisp subsets; like Prade, he allows fuzzy attribute values. But, unlike Prade, he considers both a relational algebra and a query language.

Zvieli's relational algebra and query language are based on FFOL (fuzzy first order logic). FFOL incorporates "fuzzified" versions of the logical operators of first-order predicate calculus (Section 2.? discusses how to "fuzzify" logical expressions). Additional operators are included, called *transformer operators*. These operators modify the possibility distribution associated with an expression (e.g. *concentration*, defined by $\mu_{CON(A)} = \mu_A^2$, is a possible interpretation of the adjective "very"). Finally, FFOL contains a *transitive closure* operator (the transitive closure of an expression is the conjunction of the expression with itself infinitely many times).

His relational algebra consists of operators union, intersection, difference, cartesian product, projection, selection, join, division, and assignment; as well as the relational transformer operators. Zvieli proves that his relational algebra is *complete*, in the sense that it can express anything expressible in his relational calculus, or query language.

Zvieli expresses fuzziness in a different way than does Prade, at the record level rather than at the individual attribute level. Consider, for example, a database with relation CONTACT-LOCATION, with fields CONTACT-ID, COUNTRY-OF-ORIGIN and CURRENT-RESIDENCE. In Prade's system, attribute values may be fuzzy sets. Thus, the COUNTRY-OF-ORIGIN of Contact-385 might be expressed by the possibility distribution ((France, Germany, Austria); (1, .8, .2)) (i.e. France is undoubtedly a possible country of origin; Austria is possible only at the .2 level). In contrast, Zvieli expresses fuzziness at the relation level, by specifying a fuzzy membership function of each object in the relation. Looking at just the projection of the CONTACT-LOCATION relation on the attributes CONTACT-ID and COUNTRY-OF-ORIGIN, Zvieli would express the same information as described in Table ZVIELI. In contrast to Prade, Zvieli would say that (Contact-385, Austria) belongs to the projection of CONTACT-LOCATION at level of possibility .2.

Table ZVIELI: Structure of a Fuzzy Relation in Zvieli's system

CONTACT-ID	COUNTRY-OF-ORIGIN	μ
Contact-385	France	1
Contact-385	Germany	.8
Contact-385	Austria	.2

Zvieli's representation is more expressive, but at the cost of a vastly greater storage requirement. In large systems with a high degree of fuzziness, his representation would quickly become unworkable.

To see this, note that it is quite naturalin Zvieli's system to represent the knowledge that Contact-385 no longer lives in her country of origin. We would simply assign zero possibility to the tuples (Contact-385, France, France), (Contact-385, Germany, Germany), etc. Note that, e.g., France might still be a possible current residence: we might have μ(Contact-385, Germany, France) = .8. In Prade's system, we can express contraints by defining a separate constraint relation, defined as a fuzzy subset of the cross product of the attribute domains. We can thus specify that France is not a possible attribute for CURRENT-RESIDENCE when COUNTRY-OF-ORIGIN = France. However, Prade discusses only constraints between two attribut$^{e}_{\wedge}$s. Constraints among three attributes would be required to represent knowledge that the above contraint applied only to Contact-385.

The cost of the greater expressiveness of Zvieli's system is that he must store the entire joint possibility distribution on CURRENT-RESIDENCE and COUNTRY-OF-ORIGIN. As the number of attributes in the relation increases, the storage load of Zvieli's system grows exponentially, while that of Prade's system grows linearly. For example, if there are ten contacts, each with three possible values for COUNTRY-OF-ORIGIN and CURRENT-RESIDENCE, ninety (10×3*3) tuples, with four fields each (three attributes and a possibility), must be stored (unless some tuples have zero possibility). The total storage would be 360 fields. Prade must store only ten tuples, although the COUNTRY-OF-ORIGIN and CURRENT-RESIDENCE fields of each will have three (<value>, <possibility>) pairs. The result is ten tuples with 1+6+6 fields each, or 130 fields. The problem becomes more acute as the size of the tuples increases.

Clearly, Zvieli's system is unworkable in a system of any size unless the amount of fuzziness is strictly controlled. Indeed, he criticizes the approach of Buckles and Petry because merging similar tuples "increases the fuzziness of the data (we prefer to control and minimize fuzziness)." However, he never says how fuzziness can be "control[led] and minimize[d]."

C.2.6 A general fuzzy relational database system: Representation of fuzzy linguistic expressions. Zemankova-Leech and Kandel (1984) develop a fuzzy relational database system that is similar to Prade's system, but more general in some respects.

Like Prade's system, their system can represent fuzzy data, manipulate fuzzy logical expressions, and compute possibility and necessity measures for satisfaction of queries. They also derive some fuzzy relational operators based on fuzzy similarity and proximity relations (these correspond to operators such as IS, GREATER THAN, etc.). Finally, their system that can process linguistic expressions (e.g. "John is young.").

Fuzzy relational operators allow derivation of properties based on their similarity and proximity to other properties. An example given by Zemankova-Leech and Kandel concerns deriving the possibility that Bob's hair is brown from the possibilities that it is red or blond, and the similarity of those colors to brown. Assume that Bob's hair is blond with possibility .3 and red with possibility .7 (note that there is no possibility equal to 1; this reflects the fact that no color completely describes Bob's hair). If blond is .4 similar to brown and red is .5 similar to brown, then the possibility that Bob's hair is brown is given by max{ .3×.4, .7×.5 } = .35.

The popularity of fuzzy sets has stemmed in large part from their promise for representing the imprecision in natural language. Of the authors considered here, Zemankova-Leech and Kandel are the only ones to propose a mechanism for processing linguistic expressions in a fuzzy relational database. To do this, they require a relation that "translates" any given linguistic expression into a possibility distribution. For example, they give a relation YOUNG which assigns a possibility to each age (e.g. 10 has possibility 1; 40 has possibility .4).

The storage requirements for such a scheme would be enormous, if the system were to be able to handle a large number of linguistic expressions. Moreover, although they do recognize that users might want to customize these "translation" relations, they do not consider that a single user might require different meanings for the same term, depending on context. Consider for example, the different connotations evoked by the phrases, "young business executive," and "young newlyweds," or by the phrases, "aging Politburo member," and "aging terrorist."

C.2.7 A statistical approach to incomplete information. An alternative approach to incomplete information is to allow for *probabilistic* information about attribute values. Wong (1982) considers a framework within which a number of different types of information can be represented. He considers a database consisting of a single relation. He presents results for two types of query: queries that select sets of objects and queries that extract values (projection is an example of the second type of query).

Wong creates a taxonomy, classifying incomplete information databases by the type of incomplete information and whether there is prior information about attribute values. There are two types of incomplete information. In the first type, the observed database is a *distortion* of the ideal database. There is a known distortion function which transforms a vector of attribute values in the ideal database to the vector (possibly lower dimensional) of observed attribute values. Examples are when material and labor costs have been collapsed into total cost, or when height is required in inches, but is given only as tall, average, or short. In the second type of incomplete information, the observed attribute value is a realization of a random variable whose distribution depends on the true value of the attribute (and is assumed to be independent across objects, conditional on the true value). Note that the noise may be dependent *within* objects. Thus, if we have only last year's sales and last year's salaries, we may model this year's sales and salaries as random increments from last year's. Conditional on this year's sales, the values for last year's sales are assumed independent, but sales and salaries may be dependent.

Wong considers two cases with respect to prior information. The first is no prior information. The second case is that the all vectors of attribute values are *a priori* independent and identically distributed.

Wong's formulation allows for a rich representation within the types of incomplete information it supports, but it cannot handle the full range of uncertainty modeled in Lipski's system. For example, it cannot handle the situation where more is known about a given attribute for some objects than for others (e.g. SALARY is missing for Smith but not for Jones).

For selection queries, in the no prior/known distortion function case, Wong's system would respond with Lipski's lower and upper values.

For the other three cases, where probabilistic information is available, Wong derives statistical tests based on trading misses and false alarms. When there is no prior information, a likelihood ratio test is used to decide whether an object satisfies the criterion provided in a query. When there is prior information, losses for misses and for false alarms are specified. The minimum expected loss response to a query is to report that an object satisfies the query if the posterior (to observing the data) probability that it satisfies the query exceeds a threshold depending on the ratio of the two loss values. Lipski's lower (upper) value corresponds to avoiding false alarms (misses) at any cost. Wong's framework allows investigation of cases intermediate between these two.

In selection queries, the system is to find which objects satisfy the given criteria, and the response may be simply an object identifier. In Wong's second type of query, the response will be attribute values or functions of attribute values. Wong suggests using the maximum likelihood estimate of the value when there is no prior information and the mode of the posterior distribution when there is prior information. Wong does not discuss the issue of representing the range of uncertainty associated with a query. Such a representation would be necessary in many intelligence applications, and useful in most.

In the case of selection queries with prior information, the ratio of the miss/false alarm loss values is directly related to the posterior probability that will be accepted as a positive response to the query. Thus, varying this ratio will allow selection with different degrees of stringency. Indeed, users might feel more comfortable specifying queries in the form, "select all individuals for which CITIZENSHIP = GREECE with probability greater than .7," than specifying miss/false alarm tradeoffs directly. In general, it would be useful for the results of a query to include both the items satisfying the query and their posterior probabilities.

When there is no prior information, a similar type of sensitivity analysis can be performed by varying the p-value for acceptance in the likelihood ratio test. Indeed, if equal priors are assumed, the p-value corresponds to the posterior probability.

For Wong's second type of query, again it would be useful to provide a representation of the uncertainty associated with an estimate. This could be accomplished by providing a confidence interval about the maximum likelihood estimate (in the case of no prior information), or a credible interval about the posterior model (in the case of prior information).

C.2.8 Summary

The work discussed in this section covers a very specific type of uncertainty: unknown *values* for fixed unknown objects. Prade allows an "attribute does not apply" designator; Zvieli allows attributes that only *partially* apply. None of the authors considers uncertainty about what are the relevant attributes. In general, a unique, certain key field is required, at least in base relations.

The earliest work (Codd, 1979) allows null values, with meaning "attribute value at present unknown." A number of authors have considered generalizing the query language of Codd's system. Lipski (1979) extends Codd's query language to deal with finer gradations of information than Codd's "everything or nothing" framework. Specifically, he allows the case where attribute values are subsets of the attribute domains; within his system, singletons correspond to complete information, and the entire domain corresponds to Codd's null value. Lipski provides a mechanism for computing *lower* and *upper* responses to queries, corresponding to what is known and what is possible, respectively.

These systems are general, in that the value may be unknown due to imprecision, incompleteness of evidence, or stochastic uncertainty. Nevertheless, they do not allow representation of *degrees* of uncertainty. One cannot represent the knowledge that while several values are possible, one may be in some sense a better match than the others.

In extending Codd's and Lipski's work to allow gradations of uncertainty, most authors have concentrated on representing imprecision, by means of fuzzy set theory. We have reviewed a sampling of work in this area (Prade, 1984; Buckles and Petry, 1982; Zvieli; Zemankova-Leech and Kandel, 1984). In Prade's system, attribute values are fuzzy subsets of the attribute domain. In addition to the queries allowed by Lipski, Prade allows the user to ask for those records for which a given condition holds with a certain level of possibility. Most authors deal with manipulating the mathematical structures of fuzzy set theory; only Zemankova-Leech and Kandel treat the important issue of "translating" back and forth between linquistic expression and possibility distributions. This, we feel, is a key problem limiting the success of fuzzy set theory. Like other fuzzy set theorists, Zemankova-Leech and Kandel fail to provide an adequate method for accomplishing this.

Wong has treated another aspect of uncertainty in attribute values: that of probabilistic uncertainty. He treats querying a database with incomplete information as a problem in statistical estimation. His approach is ideally suited to the case in which uncertainty in attribute values can be viewed as arising from a probabilistic process generating the uncertain values. (It is important to emphasize that Wong's approach is not limited to cases where we believe attribute value actually are generated probabilistically; it applies when we believe the analogy fits well enough for practical purposes). Wong's approach allows the user to ask for objects satisfying a criterion with a given probability (in the case where there is prior information); it also allows an explicit tradeoff between false alarms and misses.

Although we are not aware of any work in this area, it is clear that on extension of Lipski's model could be developed to handle the problem of incomplete evidence. In such a model, attribute values would be characterized not by probability or possibility distributions, but by Shaferian belief functions.

Queries would retrieve objects for which the degree of support or plausibility for a hypothesis was above a given level.

While useful, the efforts discussed in this section address only part of an analyst's problem. The analyst requires support that goes beyond the level of representing and manipulating numerical measures of uncertainty, to the level of support in developing the numbers themselves. This entails a system which can help the user in *structuring* an evidential argument, and *constructing* numberical assessments of uncertainty from the reasons and assumptions on which they are based. The Self-Reconciling Evidential Database described in Secion 3.0 will, we argue, provide the analyst with both the capability for representing and manipulating uncertainty, and the support needed for assessing the numerical inputs for the system.

REFERENCES

Barclay, S., Brown, R.V., Kelly, C.W. III, Peterson, C.R., Phillips, L.D. and Selvidge, J. *Handbook for decision analysis*(Technical Report TR-77-6-30). McLean, VA: Decisions and Designs, Inc., September 1977.

Buckles, B.P. and Petry, F.E. *A Fuzzy Representation of Data for Relational Databases*. Fuzzy Sets and Systems 7, North Holland Publishing Co., 1982.

Cheeseman, P. *In defense of probability*. In proceedings of the 9th International Joint Conference on AI. Los Angeles, CA, 1985.

Chen, S. *The Entity-Relationship Model-Toward a Unified View of Data*. ACM TODS 1, No. 1, 1976.

Codd, E.F. *Extending the database relational model to capture more meaning*. ACM Transactions on Database Systems, Vol.4, No. 4, 1979.

Cohen, P.R. *Heuristic reasoning about uncertainty: An artificial intelligence approach*. Boston, MA: Pitman, 1985.

Cohen, M.S., Schum, D.A., Freeling, A.N.S., and Chinnis, J.O., Jr. *On the art and science of hedging a conclusion*: *Alternative theories of uncertainty in intelligence analysis* (Technical Report 84-6). Falls Church, VA: Decision Science Consortium, Inc., August 1985.

Cohen, M.S., Laskey, K., McIntyre, J. and Thompson, B. *An expert system framework for adaptive evidential reasoning: application to in-flight route re-planning*. Falls Church, VA: Decision Science Consortium, Inc., March 1986.

de Dombal, F.T. Surgical diagnosis assisted by computer. *Proceedings of the Royal Society*, 1973, *V-184*, 433-440.

deFinetti, B. Foresight: Its logical laws, its subjective sources. English translation in H.E. Kybert, Jr., and H.E. Smokler (Eds.), *Studies in subjective probability*. NY: Wiley, 1964. (Original: 1937)

deKleer, J. An assumption-based truth maintenance system, *Artificial Intelligence*, 1986, 29, 241-288.

deGroot, M.H. Comment (on Lindley's paradox). *Journal of the American Statistical Association*, June 1982, 77(378), 336-339.

Doyle, J. A truth maintenance system. *Artificial Intelligence*, 1979, 12(3), 231-272.

Edman, M. Adding independent pieces of evidence. *Modality, Morality and Other Problems of Sense and Nonsense: Essays dedicated to Soren Hallden*. Lund, 1973.

Edwards W. (Ed.). Revisions of opinions by men and man-machine systems. *IEEE Transactions on Human Factors in Electronics*, 1966, 7(1).

Gardenfors, P., Hansson, B., and Sahlin, N. (Eds.), *Evidentiary value: Philosophical, judicial, and psychological aspects of a theory*. Lund, Sweden: C.W.K. Gleerups, 1983.

Hallden, S. *Indiciemaekanismer, Tidsskrift fur Rettsvitenskap*, 1973, 86, 55-64.

Harmon, Gilbert. *Thought*. Princeton University Press, 1973.

Kyberg, H.E. *Knowledge and uncertainty*, Proceedings of the AAAI Workshop on Uncertainty in Artificial Intelligence, 1986, 151-157.

Kyberg, H.E. *Interval-based decisions for reasoning systems*, Proceedings of the AAAI Workshop on Uncertainty and Probability in Artificial Intelligence, 1985, 193-200.

Lipsky, W. *On semantic issues connected with incomplete information databases*. ACM Transactions on Database Systems, Vol.4, No.3, 1979.

Lindley, D.V., Tversky, A., and Brown, R.V. On the reconciliation of probability assessments. *Journal of the Royal Statistical Society*, Series A, 1979, 142(2), 146-180.

Lagomasino, A., and Sage, A.P., *Imprecise knowledge representation in inferential activities*. United States Army Research Institute for the Behavioral and Social Sciences, 1985.

McDermott, D., and Doyle, J. Non-monotonic logic I. *Artificial Intelligence*, 1980, 13, 41-72.

Nau, R.F. *A new theory of indeterminate utilities and probabilities*. Durham, NC: Duke University, The Fuqua School of Business, 1986.

Nilsson, N.J. *Probabilistic logic* (Technical Note 321). Menlo Park, CA: SRI International, February 1984.

Nozick, R. *Philosophical explanations*. Cambridge, MA: Harvard University Press, 1981.

Peterson, C.R., Randall, L.S., Shawcross, W.H., and Ulvila, J.W. *Decision analysis as an element in an operational decision aiding system (Phase III)* (Technical Report 76-11). McLean, VA: Decisions and Designs, Inc., October 1976.

Prade, H. *Lipski's approach to incomplete information data bases restated and generalized in the setting of Zadeh's possible theory*. Inform. Systems Vol 9, No. 1. pp 27-42, 1984.

Reiter, R. A logic for default reasoning. *Artificial Intelligence*, 1980, 13, 81-132.

Schum, D.A., Current developments in research on cascaded inference processes. Chapter 10 of T.S. Wallsten (Ed.), *Cognitive process in choice and decision behavior*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Schum, D.A., and Martin, A.W. *Probabilistic opinion revision on the basis of evidence at trial: A Baconian or a Pascalian process?* (Report 80-02). Houston, TX: Rice University, 1980.

Shafer, G. *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press, 1976.

Shafer, G. Jeffrey's rule of conditioning. *Phil. of Sci.*, 1981, 48, 337-362.

Shafer, G. *Constructive Decision Theory*. Lawrence, KS: University of Kansas Business School, 1982.

Shafer, G., and Tversky, A. *Weighing evidence: The design and comparison of probability thought experiments*. Stanford, CA: Stanford University, June 1983.

Spiegelhalter, D.J., and Knill-Jones, R.P. Statistical and knowledge-based approaches to clinical decision-support systems, with an application in gastroenterology, *Journal of the Royal Statistical Society*, Ser. A, 1984, 147, 35-77.

Stallman, R.M., and Sussman, G.J. Problem solving about electrical circuits. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, August 22-25, 1977. Cambridge, MA: 299-304.

Tani, S.N. A perspective on modeling in decision analysis. *Management Science*, 1978, *24*, 1500-1506.

Toulmin, S.E. *The uses of argument*. Cambridge, England: Cambridge University Press, 1958.

Toulmin, S.E., Rieke, R., and Janik, A. *An Introduction to Reasoning*.NY: Macmillan Publishing Co., Inc. 1984.

Wong, Eugene. *A statistical approach to incomplete information in database systems*. ACM Transactions on Database Systems, Vol.7, No. 3, 1982.

Yager, Ronald R. *On the Dempster-Shafer Framework and New Combination Rules*. Tech Report MII-504, New Rochelle, NY: Iona College, 1984.

Zadeh, L.A. The concept of a linguistic variable and its application to approximate reasoning. *Information Science*, 1975, 8, 199-249; 301-357; 9, 43-80.

Zemankova-Leech, Maria and Kandal, Abraham. *Fuzzy relational data bases - a key to expert systems*. The Florida State University, 1984.

Zvieli, Aire. *On complete fuzzy relational query languages*. Baton Rouge, LA: Dept. of Computer Science, Louisiana State Univ.