COHEN

DECISION SCIENCE CONSORTIUM, INC.

# ON THE ART AND SCIENCE OF HEDGING A CONCLUSION: ALTERNATIVE THEORIES OF UNCERTAINTY IN INTELLIGENCE ANALYSIS


by:

Marvin S. Cohen
David A. Schum
Anthony N.S. Freeling
James O. Chinnis, Jr.


Submitted to:

AMRD/ORD


Submitted by:

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



December 1985

Technical Report 84-6

TABLE OF CONTENTS

# 1.0 INTRODUCTION

## 1.1 The Analyst's Dilemma

In intelligence analysis, as in most reasoning tasks, people draw conclusions, offer explanations, and make predictions based on a collection of evidence; and inevitably, they experience varying degrees of uncertainty about their conclusions, explanations, or predictions. We would argue 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.

In many contexts, however, there is a further difficulty: expressions of uncertainty are often unpopular among persons for whom the analysis is intended. Consumer A may expect a conclusive judgment from analyst B; Consumer A may react with some impatience to any attempt by Analyst B to qualify or to "hedge". Analyst B will be comforted only slightly by knowledge that such hedging is both natural and sensible behavior when available evidence does not justify an unqualified judgment.

Thus arises the analyst's dilemma: In few, if any, analyses can it be claimed that the evidence set is complete, conclusive, and absolutely reliable. If the analyst fails to qualify or hedge conclusions under these conditions, he (or she) will quickly be told that he has overstepped the evidence. Yet if the analyst does hedge in some way, his reasoning may be dismissed as invalid simply because he has made uncertainty explicit.

## 1.2 Uncertainty: A Fact of Life in Intelligence Analysis

On one view, a well-structured argument consists of compelling evidence that "speaks for itself", and concern over the manner in which uncertainty is assessed and expressed is unnecessary. Unfortunately, arguments of this kind do not appear often. In general, a hedged, qualified, or probabilistically-expressed conclusion is both reasonable and proper whenever:

- the evidence set is less than complete on all relevant matters affecting the conclusion, or
- the assembled evidence is inconclusive, i.e., it is to some degree consistent with the truth of more than one possible conclusion, or
- the evidence comes from imperfect or unreliable sources, or
- there may be possible conclusions other than those which have been specifically entertained.

Sometimes it will be true that with more effort, more or better evidence could have been obtained. But in our view, the conditions outlined above are not usually <u>flaws</u> in an analysis. They are facts of life. Problems arise only when they are not acknowledged and appropriately handled. For example, it may seem prudent to suspend analysis until "better data" arrive, or a new and better sensor is developed. Unfortunately, however, there is no sensor for intentions. More generally, it is unlikely that you will ever have all the possible evidence on a given problem. You are more likely to have all the evidence you can handle. The real challenge is sorting out relevant evidence from the large amount available and extracting its implications.

Of course, there are ways to avoid the trouble of considering and expressing uncertainty. One way is to refrain from drawing any conclusions at all. Purely descriptive or noninferential analyses are often useful (sometimes requested). However, some customers who lack substantive knowledge and who depend upon the analyst's assessment of the significance of the evidence may well be unhappy at the necessity of having to draw their own conclusions.

Another way to minimize uncertainty is to limit the number of possible conclusions which are entertained; the limit, of course, is one. It is often very tempting to commit oneself to a certain conclusion, particularly if this conclusion seems popular for one reason or another. Evidence which favors the conclusion is sought after; evidence against the conclusion is avoided, explained away, or viewed as anomalous. The many hazards of these and other similar strategies for avoiding or suppressing uncertainty will be apparent.

## 1.3 Toward a Resolution of the Dilemma

What can the practicing analyst _do_ about uncertainty--both to communicate it effectively to consumers and to enhance personal understanding of the analytical problem? This report is an effort to address that question. We do so, in part, by asking some more fundamental questions: what _is_ uncertainty: are there different kinds, as well as different degrees, of uncertainty? in what way (or ways) should it be conceptualized and measured? An implicit theme is that different _theories_ of uncertainty can shape, direct, or more subtly influence the _art_ of performing and reporting intelligence analyses. Theory may inform and improve practice.

In recent years, a variety of alternative approaches to inference have been proposed or defended: Bayesian probabilities, fuzzy probabilities, possibilities, belief functions, and others. They often have dramatically divergent implications for the assessment, aggregation, and/or reporting of uncertainty. They differ in the concepts they attempt to capture (e.g., chance, imprecision, completeness of evidence), in the degree to which appropriate normative justifications have been achieved, and in the demands they impose on the analyst for assessments and computations. Perhaps most importantly, however, they differ (or purport to differ) in compatibility with the decision processes of analysts and consumers: i.e., in the readiness with which they prompt questions and represent distinctions which are natural or illuminating to a particular analyst or problem domain, and in the extent to which they satisfy the needs of policy makers. Acceptance among practitioners may well hinge on this factor. So a second implicit theme of this report is that practice is the ultimate test of theory.

The bulk of this report is in the form of a dialogue: a series of briefings and conversations among intelligence analysts who must somehow cope with uncertainty on a daily basis. Technical debate among competing theorists has been lively in recent years; here, our goal is to translate that debate--among logicians, statisticians, and psychologists--into the context and language of a "real-world" application. In this dialogue, it is the analysts who have mastered the essential details of one or another of the competing positions, and who must hash out some of the advantages and disadvantages of different viewpoints.

A common thread in this discussion is the issue of the knowledge underlying the assessment and reporting of uncertainty. For each inference framework two specific questions can be asked:

- o Does it demand inputs that match, or exceed, or fall short ofthe knowledge of the user?
- o Does it provide some meaningful measure of the knowledge, or weight of evidence, underlying the outcome of an analysis?

These two issues are, of course, related. For example, one way to ease the task of assessing inputs is to require intervals (e.g., "the probability is beween .2 and .6") rather than precise numbers. And one way to represent the amount of evidence behind the conclusions of an analysis is to provide "confidence" intervals in addition to, or in place of, precise numbers. Variations among different theories of inference can be understood in significant measure as differences in the way they address these two questions.

## 1.4 Objectives

Those who might benefit from "overhearing" this dialogue include:

- o Practicing intelligence analysts,
- o Agency researchers in the Office of Research and Development, and elsewhere who support analysts by identifying or developing new aids for inference and related tasks, and
- o Agency educators who present courses on inferential issues in intelligence analysis.

Among the anticipated benefits are the following:

- o Increased familiarity with the concepts and underlying rationales for several major current theories of inference: Bayesian probability theory, Glenn Shafer's theory of belief, Lotfi Zadeh's fuzzy set and possibility theory, and L.J. Cohen's theory of inductive probability;
- o An understanding of how these theories can be put to work in a con crete analytical problem; and
- o An introduction to some of the current issues and controversies among these alternative viewpoints.

The goal is not to provide a cookbook for solving inference problems, or a full working knowledge of any of the rival viewpoints. Our expectation, rather, is that a qualitative grasp of basic concepts is a valuable first step, and may by itself bear fruit in more reliable <u>and</u> more defensible analyses of intelligence data.

## 1.5 Outline of the Report

Chapter 2 contains an illustrative problem in intelligence analysis to which we shall return periodically throughout the paper. Chapter 3 is a dialogue in which the systems of inference associated with Bayes, Zadeh, Shafer, and L.J. Cohen are discussed. Three of these theories are applied to the sample problem in Chapter 4.

## 2.0 AN ILLUSTRATIVE PROBLEM IN INTELLIGENCE ANALYSIS

Art, a weapon systems analyst for a U.S. government agency, has been asked for a quick assessment about the following situation. An interested policy-making "customer" requires a judgment about whether weapon developers in Malbridgia are now attempting to build a prototype of a tactical weapon system called ZAP. System ZAP, requiring several novel subsystems, could replace an existing System YAP now deployed by Malbridgia. If System ZAP is developed successfully, it would give Malbridgia a decided tactical advantage wherever such systems might be used. This customer requires a briefing in three days.

As Art begins work on the problem he has a fairly strong expectation that developers in Malbridgia are <u>not</u> currently building a prototype ZAP system. This expectation is based mainly upon a recent briefing he heard given by a nationally-recognized American scientist. This scientist discussed why it is not yet technologically feasible to develop certain subsystems which ZAP would require. In addition, the scientist remarked about the very costly nature of development of such subsystems, should they become technically feasible. Nevertheless, Art seeks, as evidence in this task, specific information about development efforts in Malbridgia. Of the several items of information he is able to obtain from his files, and from those of his colleagues, the following five evidence items seem to be the most relevant.

### THE EVIDENCE

$E_1$: From open-source literature is a scientific paper published one year ago by a scientist in Malbridgia. One conclusion of this paper is that the technology for developing subsystems of the sort that ZAP would require is "at least five years away." (Apparently, their scientists agree with ours on this matter.) This paper, by the way, was written in Malbridgian; the available copy is but a translation offered by a person whose credentials are unknown.

$E_2$: From a "very reliable" source in South Contraria is a report, dated 6 months ago, that representatives from South Contraria and Malbridgia negotiated a contract for the immediate purchase by Malbridgia of a large quantity of Xyleum, a material vital to the development of a required subsystem for ZAP.

$E_3$: Ten months ago a mid-level government employee in Malbridgia, supplied us with a copy of a document allegedly containing minutes of a meeting (held one year ago) of military weapon planners in that country. These minutes record a decision by these planners to transfer 15 technologists from their then-current work locations to a known weapon development site. Such transfer was to be completed within six months of the meeting date. The 15 technologists were named in an appendix to these minutes. The source was also able to identify 10 of the 15 persons as having been instrumental in the development of their current YAP system. The source is rated as "usually dependable;" however, an update on their records reflects that he has made no contact with us for six months. In addition, the document, a copy of which we have, is rated as "probably authentic."

$E_4$: A national from South Contraria named L. Melfata recently reported to us about a discussion she held two months ago with a Malbridgian technologist. Melfata asserted that this technologist had rather boastfully described recent advances in the development of several of the novel subsystems required by System ZAP. Melfata has many contacts in Malbridgia and is allowed to travel freely there. There is room for speculation that she may also work for Malbridgia.

$E_5$: P.F. Muldip, an influential political figure in Malbridgia, asserted one year ago to a member of the press from Malbridgia that he, Muldip, would strongly back the development of new weapon technologies, in particular, the development of several subsystems among which are two of those which System ZAP would require. Our knowledge of a recent political power struggle in Malbridgia causes us to wonder whether or not Muldip is still in a position to influence weapon system development.

This collection of five evidence items is hardly impressive, but it is the best that Art can muster on quite short notice. The evidence has three characteristics in common with most, if not all, evidence collections upon which analysts must base conclusions. The evidence items are inconclusive, the collection of items is by no means complete, and the items of evidence come from sources whose credibility and competency are less than perfect. In short, in his briefing Art knows he will somehow have to qualify or hedge his judgments.

## 3.0 PRELUDE: DECIDING HOW TO DECIDE

Art calls upon certain of his colleagues for varying degrees of assistance in his work on this problem. Two days before the briefing, he meets with two fellow analysts, Sy and Phyllis.

<u>Sy</u>: Well, this is not a very impressive lot of evidence. Some of it points one way, some of it the other.

<u>Phyllis</u>: Yes, we even have two views of Malbridgia's technical capability to build ZAP: $E_1$ and $E_4$.

<u>Art</u>: There does seem to be <u>more</u> evidence in favor of building ZAP than against it. But I wonder if Malbridgia wants us to <u>believe</u> they are building ZAP? If that were true, the evidence in favor would not mean much.

<u>Sy</u>: I'm not sure it means much anyway, Art. Let's assume all our sources are reliable: forget about the possibility of deception, and also forget about the conflicting evidence $E_1$. $E_2$, $E_3$, $E_4$, and $E_5$ <u>still</u> don't prove that Malbridgia is building ZAP.

<u>Art</u>: How's that?

<u>Sy</u>: Well, it's entirely possible that Malbridgia has the technical capability to build ZAP ($E_4$), that there is some political support for doing it ($E_5$), and that various resources have been mobilized ($E_2$, $E_3$), but that they aren't in fact building ZAP. The technicians and material could be there for a different purpose, after all, and that politician Muldip could have been overruled by other figures in power. We have no idea what other hypotheses might also explain this evidence, for example, development by Malbridgia of some other system that is technically similar to ZAP.

<u>Art</u>: When we add back in the unreliability of the sources and the conflicting evidence $E_1$, we don't have much, do we?

<u>Sy</u>: I don't see how we can draw any conclusions at all from this data. Why don't we just report our evidence and let the customer draw her own conclusions?

Art: I don't think the customer would like that very much. We're supposed to be the experts here, after all. Our job is more than just compiling a lot of raw data; it includes assessing their significance.

Sy: Hmmm. I don't know if all analysts would agree with that. But if you see it that way, and the best evidence you can muster falls short of certainty, how can you avoid simply guessing at what it all means?

### 3.1 The Inference Theory Approach and Some Initial Doubts

Phyllis: Let's not give up too soon. There *is* another possibility, a compromise between guessing at conclusions and sticking to known facts, but I'm not yet sure how practical it is.

Art: Well, don't keep it a secret!

Phyllis: I've been hearing about various theories, or formal frameworks, for reasoning about uncertainty. Perhaps there is a scientifically respectable way of tracing the implications of your data, Art. One thing these formal theories have in common is a strategy called "divide and conquer." If the problem is too complicated or confusing to deal with as a whole, you break it down into simpler elements, make some assessments regarding those elements, and then use a calculus provided by the theory to compute your degree of certainty in the various possible answers.

Art: Interesting, but it does sound rather mechanical. Put in some numbers and out come the answers! In the first place, I never have felt very comfortable about expressing my beliefs as numbers--that's asking for more precision than is really there. Secondly, does a theory of this sort capture the way *I* naturally think about the problem? Much as I appreciate the help, I do think I have acquired some rather unique expertise in my years as an analyst. Why should I trust the output of a process that doesn't reflect that experience?

Sy: I have some doubts about this, too, but I don't object to its being "mechanical." In fact, it doesn't sound mechanical enough. You haven't gotten very far beyond guesswork. Although you no longer have to guess at the conclusion itself, I gather from your description that the assessments you provide for elements of the problem are subjective. Moreover, what is the justification for using any particular "calculus" to combine these assessments? Surely, there's nothing comparable here to physical laws or mathematical proof. In short, why suppose the conclusion derived in this way is any better established than if you guessed?

Phyllis: Hold it! I'm being attacked on two fronts at once. You have raised two kinds of objections that appear, at least on the surface, to be at cross-purposes. Art is concerned about descriptive issues: how close a theory of inference and its required inputs come to replicating your own reasoning processes. Sy is concerned about normative issues: the degree to which such a theory is justified as a recipe for how reasoning ought to be conducted.

Art: Well, maybe they're both important.

Phyllis: I gather that there is disagreement both about how various theories measure up in regard to these criteria, and about the relative importance of the criteria themselves...Wait a minute! I see just the four specialists we need coming down the hall: Barbara, Zara, Shawn, and Colette. They support abbreviated and somewhat modified versions of four current positions on inference--the views of Bayes (and his contemporary followers), Lotfi Zadeh, Glenn Shafer, and L.J. Cohen, respectively.* Let's explain our problem to them.

(Art, Sy, and Phyllis recount their conversation to Barbara, Zara, Shawn, and Colette.)

---

* In all these cases, the most accurate assumption is that the speakers have been influenced by the views of the theory for which they are named, rather than reflecting them exactly.

3.2 <u>The Bayesian Answer</u>

<u>Barbara</u>: Well, I'm glad I got here in time to clear up this confusion. Thanks to an eighteenth century English clergyman named Thomas Bayes and to extensions of his ideas by many others, it is possible, Sy, to provide a rigorous justification for reasoning about probabilities. I'll say a little bit about that in a moment. With all due respect, Art, we are much less concerned with describing how people "really" think. If ordinary reasoning were already consistent with Bayesian precepts, a normative theory could not improve it.

<u>Art</u>: But I've got to be able to <u>use</u> the normative theory, don't I?

<u>Barbara</u>: That's right. So it <u>is</u> important that people be able to understand and assess the inputs required by a theory. This is one of the major strengths of the Bayesian framework as developed, for example, by Ramsey (1926), de Finetti (1937), and Savage (1954). "Probabilities" are specified not as abstract, intangible quantities, but as reasons for action. So I can determine your degree of belief (or numerical subjective probability) for a proposition simply by asking you about actions you would choose; the catch is that the outcomes of your actions will depend on whether or not the proposition turns out to be true. Your subjective probability for a proposition is reflected in the odds at which you would be willing to bet on it. For example if you would pay no more than 70 cents for a gamble in which you receive 1 dollar if the proposition is true, then your probability for the proposition is .70. Some elaborations of this procedure take into account your attitude toward risk.

<u>Art</u>: What do I do with these probabilities once I've assessed them?

<u>Barbara</u>: Let's take the topic of your briefing: whether Malbridgia is building System ZAP. Bayesian theory gives you a variety of ways to break down <u>that</u> probability into probabilities that you find easier to assess.

<u>Art</u>: Such as?

Barbara: Well, the most natural way to handle this problem would be to use Bayes' rule. You would start by assessing your prior probability or odds for the hypothesis that Malbridgia is building ZAP, before considering any of the evidence. Then you would quantify the impact of each bit of evidence. This quantification involves the assessment of a "likelihood ratio," which is simply the probability of the evidence given that the hypothesis is true divided by its probability given that the hypothesis is false. Then Bayes' rule can be used to combine your original beliefs with your assessment of the impact of the evidence, to derive what your new beliefs ought to be.

Art: OK. Let's say I start off thinking the odds are about 5 to 1 against Malbridgia's building ZAP. Then I discover evidence $E_4$--apparently a Malbridgian technologist has boasted about technical advances relevant to building ZAP. Now what?

Barbara: The question is, how much more likely would this evidence be if Malbridgia is building ZAP then if it is not building Z?

Art: I'd say it's about twice as likely if Malbridgia is actually building ZAP.

Barbara: Then your posterior odds for the hypothesis, after receiving $E_4$, is just the prior odds times the likelihood ratio for $E_4$: $1/5 \times 2/1 = 2/5$. This corresponds to a probability of $2/(2+5) = 2/7 = 29\%$. So the new evidence $E_4$ does not outweigh your prior expectation that Malbridgia is not building ZAP.

Phyllis: You said there was more than one way to express the probability that Malbridgia is building ZAP in terms of other probabilities?

Barbara: That's correct. For example, evidence often bears on the hypothesis of interest indirectly, through a series of intermediate hypotheses. That's certainly the case in our problem. Take $E_4$, for example. The datum represented by $E_4$ is not that a Malbridgian technologist boasted about recent technical advances, but rather L. Melfata's report to that effect. So the evidence bears on the hypothesis that Malbridgia is building ZAP indirectly, via the intermediate hypothesis that Melfata's report is true. In this case, instead of directly assessing the likelihood ratio for $E_4$, you might assess

$E_4$'s impact on the intermediate hypothesis and on its complement (that Melfata's report is false), and then assess the impact of the intermediate hypothesis and its complement on the hypothesis of interest. This process involves what is known as cascaded, or hierarchical, inference (Schum, 1980).

Phyllis: In that case, you've got a lot more assessments to make and, I would guess, a more complex computation to perform. But the advantage is that you've broken down the direct likelihood ratio into components that you feel more confident about assessing?

Barbara: That's right. This way of analyzing a problem is especially useful when we are concerned about the credibility of a source. It lets us focus separately on credibility issues and on the evidential value of what the source said, assuming that it were true. These are lumped together when we assess a direct likelihood ratio for $E_4$.

Art: In fact, even if Melfata were telling the truth, perhaps we should be worried about the credibility or motives of the scientist she said she heard boasting.

Barbara: If you want to deal with that concern separately, you can just insert another intermediate hypothesis in your analysis. You assess the impact of the boasting (assuming that it occurred) on your belief in the intermediate hypothesis that technical advances have taken place; then you assess the impact of the latter hypothesis on the claim that Malbridgia is building ZAP. Without going into a lot of detail, let me say that structures of any degree of complexity can be created within this Bayesian framework and can be made to capture a wide diversity of inferential subtleties (Schum, 1980, 1981).

Art: You know, even without doing the more complex analysis, I realize now that I overestimated the impact of $E_4$. Taking into account these doubts about the credibility of Melfata and of the scientist, I'd now say $E_4$ was about 1.5 times as likely if Melfata is building ZAP than if it isn't building ZAP. That makes my posterior odds $1/5 \times 1.5/1 = 3/10$ and gives a probability of $3/(3+10) = 3/13 = 23\%$ that Malbridgia is Building ZAP.

Sy: Then is Bayes' rule the only method for combining probabilities in an inference task?

Barbara: By no means. It is only one of several useful formulae that can be derived from the probability calculus. Bayes' rule seems natural where the evidence is a "symptom," or a causal effect, of the hypothesis. In other cases we may prefer to analyze the problem in terms of intermediate uncertainties that are antecedents or preconditions of the hypothesis. For example, we don't know whether Muldip, the politician referred to in $E_5$, did in fact support the building of ZAP. But we can assess the probability that he did based on our evidence, and then assess two conditional probabilities: that Malbridgia would build ZAP given that Muldip supported it and that Malbridgia would build ZAP given that Muldip did not support it. Now we can combine these assessments into an estimate of the probability that Malbridgia is building ZAP using a rule called the Law of Total Probability.

Phyllis: Does that mean that Bayes' rule couldn't have been used?

Barbara: Not at all. We could have constructed a quite different analysis here, involving Bayes' rule. Such an analysis would be just as correct from a formal point of view. But it would require us to assess the probability that Muldip supported the proposal to build ZAP given that Malbridgia is now building ZAP (and also given that Malbridgia is not building ZAP). That seems much less natural to us than the probability that Malbridgia would build ZAP given that Muldip supported it.

Art: That's very interesting. Even though you're trying to improve ordinary reasoning rather than duplicate it, the selection of an appropriate Bayesian structure does depend on how we ordinarily think. You try to decompose the problem into elements that match the way we store information, to make it easier for us to provide inputs for the analysis.

Barbara: Correct. And the ability to do that is a large part of the art, as opposed to the mathematics, of Bayesian decision analysis.

Sy: But I gather that improvements in reasoning would be due to the math?

Barbara: That's right. The Bayesian approach transfers the burden of combining these inputs from the decision maker's head to his calculator or computer.

Phyllis: So far, we've learned that the Bayesian theory provides a behavioral interpretation of its inputs (in terms of betting) and affords a rather wide diversity of analytical structures to capture intuitions about evidential relationships. I guess that leaves out one thing: Do Bayesian probabilities tell us anything about the actions we ought to take?

Barbara: They certainly do. The link is simple and direct, and is in fact presupposed by the betting interpretation. Suppose you have a choice among actions, and the outcome for each action depends on uncertain events or states of affairs. Bayesian theory says you should assess both your utility (or degree of preference) and your probability for each possible outcome of each action. You compute the expected utility of each act by summing the products of the utilities and probabilities. This represents a sort of "average" preference for each act. Then you select the act that maximizes expected utility.

Sy: This is all very impressive, Barbara, but I guess I still have trouble understanding why we should believe the result.

Barbara: Then let me get to the justification. A powerful argument can be given that your degrees of belief, as reflected in your choices among bets, ought to be consistent with the probability calculus. It turns out, as de Finetti (1937) showed, that unless your beliefs are probabilistically coherent in this way, a devious adversary could arrange a set of gambles which you would accept, but in which you were sure to lose. A set of gambles of that sort is called a "Dutch book".

Sy: Well, I wouldn't worry too much about a Dutch book. I don't know many adversaries clever enough to figure all that out.

Phyllis: Perhaps actually protecting yourself against a Dutch book isn't the real point, Sy. You must admit that if a theory of probability leaves you open to a Dutch book, it might be symptomatic of something wrong with the theory.

Barbara: I'm glad you agree, Phyllis. Now let me sum up my answers to Art and Sy. First, Art: Bayesian theory is an idealization of how people actually think, not a literal description. In some cases, for example, people may simply not know how they would bet; their choices may be indeterminate at the level of precision required by the theory. This was acknowledged by Savage (1954, pp. 57-58) and by de Finetti (1937, p. 60). Moreover, the ensemble of their choices among bets may fail to be consistent with the probability axioms. In fact, it is for these reasons that the theory is of value. It provides a way of computing degrees of belief for propositions which one finds hard to evaluate directly, in terms of propositions which are easier to assess. The result is that one's beliefs become more consistent with one another and more precise than they otherwise would be.

To Sy, I would argue that subjectivity is inescapable in dealing with uncertainty. If you have a customer who needs to make timely recommendations regarding action or policy, you may not have the luxury of waiting for conclusive evidence--the "smoking gun." Not acting is itself a decision, after all, and it may not be the best one. If we take seriously the fact that we must make decisions under some degree of uncertainty, then Bayesian theory is the right approach. It enables us to bring to bear assessments of uncertainty that are relatively precise upon assessments which cannot be confidently made, and it is the only guide to action that guarantees us against the expectation of a sure loss.

### 3.3 <u>The Problematic Relation Between Probability and Knowledge</u>

<u>Sy</u>: I'm a bit confused, Barbara. Let me state my understanding of your position, at the risk of some exaggeration. You agreed that we should select a probability model whose inputs match the way we store information?

<u>Barbara</u>: Right.

<u>Sy</u>: Well, that seems to imply that we have in our heads "true" (or psychologically definite) assessments for some probabilities and not for others. In the case of Bayes' rule which you described, for example, we would have determinate assessments for prior probabilities and likelihoods, but not for the "posterior probability" (the probability which reflects both prior and new information). So we let the model compute the latter from the former.

<u>Barbara</u>: I suppose I did imply that. But it sounds a bit less plausible when you state it so explicitly.

Sy: I think it's implausible, Barbara, because we are in fact able to provide assessments that fit more than one model. I can directly assess the probability that Malbridgia is building ZAP, by the betting paradigm, or I can derive it from assessments of probabilities for other propositions. And, as you just noted, there will be more than one way to express the probability of a proposition in terms of other propositions. Now I may feel more confident in some of these approaches than in others, but I certainly wouldn't reject any of them as totally meaningless.

Barbara: I think you have a good point.

Sy: Well, then, here's the problem: The Bayesian approach doesn't tell us what model for a particular problem is "the" right one. I suppose a perfectly rational being, from a Bayesian point of view, would arrive at the same answer by all these different routes. An ordinary mortal like myself, however, would be able to supply inputs for more than one model, but they might well turn out to be inconsistent.

Barbara: That's certainly true.

Sy: Suppose I directly assess the probability that Malbridgia is building System ZAP and also provide assessments that allow the probability to be computed indirectly. If the results agree, I had no need for the Bayesian computations in the first place; I get the same answer by direct judgment. But if they disagree, Bayesian theory gives me no way to decide among them. So it doesn't help then, either.

Barbara: Well, not so fast. You're quite right that Bayesian theory doesn't dictate what you ought to believe. Strictly speaking, what the mathematical part of it does is alert you when a set of probabilities is internally inconsistent.

Sy: Is that all it does?

Barbara: Well, in that regard, it's no different from logic in general. You can then choose among a variety of possible revisions in your set of beliefs to eliminate the conflict and restore coherence. For example, in traditional two-valued logic, suppose you believe a hypothesis H and also believe other things, I, J, and K, that are shown to imply not-H. To remove the inconsistency, you may decide to adopt any of not-H, not-I, not-J, or not-K. Similarly, in probability theory, if you assign probability .6 directly to H, but indirectly derive a probability of .5 (e.g., by assigning .3 to H given C, .7 to H given not-C, and .5 to C), you can either revise your direct estimate from .6 to .5 or else adopt any number of ways to make the indirect estimate of the probability of H come out to .6.

Art: Hmmm. One thing is becoming clear: our comparison of inference theory to a machine that just cranks out answers was way off the mark. In fact, the choice of what to believe is still quite subjective, even if it is constrained by the maxims of probability theory. Instead of accepting the conclusion of an analysis, you could hold on to your "gut feeling" for the probability of A and adjust the results of the analysis to agree with it.

Sy: Well, Art may be pleased by this, but it bothers me that Bayesian theory provides so little guidance in deciding what to believe. Just because the conclusion isn't dictated by logic or probability theory, Art, doesn't mean it's entirely arbitrary. There might be very good reasons for favoring one set of assessments over another.

Barbara: I agree, Sy. But you're forgetting what I said earlier: selecting an appropriate structure is part of the "art" of decision analysis, not the mathematics. In practice, you would use the assessments you felt most confident with, and derive other probabilities from them.

Phyllis: We remember, Barbara. But a Bayesian should be the last person to say that something can't be dealt with formally because it's subjective. If I understand you correctly, Barbara, you are now saying that the function of Bayesian decision theory is to help us police our set of beliefs for consistency, so as to avoid a potential Dutch book. But earlier you emphasized a very different function: to enhance precision or confidence in our beliefs by taking the assessments we feel less confident about judging directly, and deriving them indirectly from more confident ones. The problem is, Bayesian theory has a lot to say about consistency or coherence, but virtually nothing to say regarding this pivotal concept of "precision" or "confidence."

Art: But I thought that's what probability theory was all about! The more confident I am in a conclusion, say, that Malbridgia is building ZAP, the higher the probability I assign it.

Phyllis: That's not quite right, Art. From a Bayesian point of view, the conclusion of your analysis is not that Malbridgia is or is not building ZAP, but the probability that it is. So we need to know your confidence in the probability. For example, suppose you assign a 50% chance that a coin will land heads and a 70% chance that Malbridgia would build ZAP given that Muldip supported it. The second probability is higher than the first; but the first is relatively sharp and firm, while the second is vague and labile. You can easily imagine receiving new information, or remembering old information, about Muldip's standing in Malbridgia that would cause that probability to shift, but you do not expect to alter your belief that the coin is fair. Traditional Bayesian theory treats both probabilities the same.

Barbara: And quite rightly, Phyllis. From a normative point of view, differences in confidence don't matter. In order to avoid the possibility of a Dutch book, you must use your vague 70% probability just as you would a sharp 70% probability. In other words, the normatively recommended decision, which maximizes your expected utility, will be unaffected.

Sy: I'm not so sure, Barbara. Let me return to my original point: Since there is more than one way to arrive at any probability estimate, we don't know what our probability for an event is without some consideration of which assessments we trust. In fact, the recommended action could easily be different depending on which competing set of probabilities we decided was more credible.

Phyllis: The problem, it seems to me, is that Bayesian theory works quite well as a description of an "ideal" decision maker. We can then comfortably assume that all relevant information is utilized in all assessments. What troubles me is an implicit assumption about ordinary decision makers.

Bayesians seem to be assuming that the ordinary decision maker also utilizes all relevant information, (but differs by being limited to a more restricted set of assessments). I think this may often be wrong. One reason, at least, why we cognitively fallible organisms need the theory in the first place is that we <u>don't</u> automatically make use of all relevant information. Different formulations of the problem may trigger different chains of associations or direct my attention in different ways. As a result, I may be able to assess a probability in a variety of direct and indirect ways, but I will very likely draw on different portions of my store of knowledge each time. This is what makes some assessment strategies better and more natural than others, and is what makes a good analysis good.

<u>Art</u>: So most of the time the real problem is how to pull together all our knowledge into a single conclusion?

<u>Phyllis</u>: Exactly.

<u>Art</u>: I think what you just said has a very important implication, Phyllis. We may not want to rely on just <u>one</u> analysis, even if it is the one we have most confidence in. If the goal is to tap as much of our knowledge as we can, we should foster inconsistency, at least temporarily, by approaching each problem in more than one way (Brown and Lindley, 1982).

<u>Phyllis</u>: I think that's right. Consistency could be achieved very quickly by <u>restricting</u> the knowledge your assessments draw on: I could just select one assessment model and ignore the rest. If we want to increase the total amount of knowledge utilized in reasoning, it clearly will not be an automatic by-product of achieving consistency.

<u>Sy</u>: I would guess, though, that some measure of the amount of knowledge incorporated in the different approaches would be needed to combine them in a formally justifiable way.

<u>Phyllis</u>: I would think so. In short, the Bayesians tell us that an ideally rational person would <u>already be</u> a Bayesian; but the normative theory seems to tell <u>us</u> nothing about how to get there from here.

### 3.4 Probabilities Upon Probabilities I: Precision

Barbara: You may be pleased to know that there is some formal work on this in the Bayesian tradition, although it is controversial. This work recognizes that we may be uncertain about what our true probabilities are and utilizes probability theory itself, at a second level, to measure this uncertainty.

Art: I see. Suppose my analysis concludes that there is a 70% chance of Malbridgia's building System ZAP. I can now turn around and ask, at a second-order level, what is the probability that 70% is in fact the true probability? I can also ask about the probability that 69% is the true probability, and so on.

Barbara: That's right. Tani (1975), for example, defines our "authentic probability" as the one which best describes all our relevant knowledge. This is to be contrasted with probabilities that are actually assessed, which he calls "operative" probabilities. Similarly, Watson, Brown, and Lindley (1977) and Lindley, Tversky, and Brown (1979) regard elicited probabilities as noisy measurements of the true ones, which are defined as the probabilities you would provide after infinite time for thought and introspection. On both views, uncertainty about the true probability is expressed as a set of second-order probabilities for possible values of the true first-order probability.

Art: Then my confidence in a first-order probability judgment is represented by the range of first-order probabilities that must be considered probable, in a second-order sense.

Barbara: That's right. You could even measure your confidence by using a 95% uncertainty interval; i.e., the range of probabilities within which you feel 95% sure the true probability falls.

Art: I see. So my confidence in the probability for a coin landing heads could be represented by a very narrow interval around .5, such as .49 - .51; while my confidence in the probability that Malbridgia will build ZAP given that Muldip supported it is a much wider interval around .7, such as .2 - .9.

Barbara: Exactly. Now we can say a little more about the goal of a decision analysis. It is not merely to arrive at a consistent set of first-order probabilities, but to arrive at a consistent set of probabilities which is as close as possible to the "true" values, i.e., to maximize higher-order precision.

Sy: So if I directly assess the probability that Malbridgia is building Zap and also derive it from an analysis, I'd expect the uncertainty interval for the second estimate to be smaller than for the first?

Barbara: That's right. The analysis gives you a more accurate estimate of your own true probability. As one Bayesian (Lindley) has said: inside every real decision maker is a rational man fighting to get out. These higher-order probabilities, together with Bayes' rule, can also be used to combine the results of different analyses when they disagree (Lindley, Tversky, and Brown, 1979). The resulting conclusion will reflect more of your information about your probabilities--and be more precise--than any of the individual approaches taken by itself.

Phyllis: Frankly, Barbara, the idea of second-order probabilities sounds pretty obvious. Why did you say it was controversial?

Sy: I think I can see one objection, Phyllis. There is no reason to assume second-order probabilities are precisely assessed either, so the application of probabilities to probabilities seems to generate an infinite regression. I doubt if convergence could be demonstrated.

Art: I'm disturbed about the assessment burden as we ascend this hierarchy. It might be harder to assess my confidence in my probabilities than it was to assess the probabilities in the first place.

Sy: I would guess, too, that the application of probability theory gets harder to justify normatively as you ascend. It's hard to see how a decision maker could make much sense of choices among bets about his first-order probabilities, or that he would worry much about a Dutch book on what his true probabilities are. For one thing, no one will ever really know what they are unless the decision maker applies an "infinite" amount of thought and introspection. Even then, given the possibility of a non-converging regress, I'm not sure it makes sense to assume there are such things as "true" probabilities somehow inside his head.

Shawn: I share your skepticism, Sy. In my view, probability assessment is not a process of uncovering pre-existing "real" degrees of belief. Degrees of belief are not "measured"; they are created (Shafer, 1981). I regard a probability model as an argument; it's a good argument if it helps us capture and organize some portion of our evidence in a cogent, insightful, thorough, and reliable way (Shafer and Tversky, 1983).

Barbara: Hold on a minute here. Second-order probabilities may be a very useful fiction despite all these difficulties, if they help us express our confidence in first-order judgments or, if you prefer, "arguments."

Sy: That depends on whether the conclusions they lead to make sense, doesn't it, Barbara? I have some doubts here, too. Consider the following example. Suppose I perform two probabilistic analyses of the hypothesis that Malbridgia is building ZAP. I start with a prior probability of .2; the first analysis yields a posterior probability of .8, the second yields a posterior probability of .6. According to the measurement analogy (as interpreted by Lindley, Tversky, and Brown, 1979), reconciliation gives me an estimate of my "true" probability that typically lies between .6 and .8--e.g., .7.

Barbara: OK, what's wrong with that?

Sy: Well, Barbara, it ignores the possibility that my two analyses could have mentally tapped independent sources of information, or evidence. In that case, both collections of evidence appear to favor the hypothesis that Malbridgia is building ZAP. Thus, the second analysis should cause us to increase our probability (e.g., from .8 to .9), not decrease it (from .8 to .7)!

Phyllis: In fact, Barbara, isn't probability a function of two things: the hypothesis and the evidence upon which the probability is based?

Barbara: That's right.

Phyllis: Well, if they are based on different evidence, the probabilities provided by two analyses are not estimates of the same "true" probabilities. For example, the "true" probability that Malbridgia is building ZAP, given one bit of evidence (say $E_2$), is not the same number as the probability that Malbridgia is building ZAP given some other evidence (say $E_3$). Infinite thought and introspection applied to $E_2$ will give a different answer from infinite thought and introspection applied to $E_3$. The analogy to reducing the noise in a set of measurements doesn't seem at all suited to the picture of different probability analyses capturing different pieces of our total knowledge.

Sy: This sounds like a serious problem, Barbara; after all, our original reason for considering more than one probability model was to increase the amount of knowledge utilized in our reasoning. To be specific, most of the time we are interested in maximizing knowledge about some real-world hypothesis (e.g., whether Malbridgia is building ZAP), not knowledge about our "true" probabilities--even if they existed.

Art: Hmmm. Second-order probabilities appear to be a fictional device aimed at a fictional problem.

### 3.5 Probabilities Upon Probabilities II: Completeness of Evidence

Barbara: Maybe so, Art. But I'm not quite ready to give up on this approach. It seems clear that we went wrong at the start, when we focused on "thought and introspection" rather than additional evidence. What we should look at instead is how different our current probability will be from the one we would have when we get more data.

Art: It sounds like that's worth a try. Suppose I do a Bayesian analysis of the chance that Malbridgia is building ZAP, but I possess only part of the evidence, say $E_1$ and $E_2$. The probability that emerges from this analysis, say, is .23, but I know there's some other potential evidence. For example, an intelligence source thinks he can obtain minutes from meetings of military weapons planners in Malbridgia ($E_3$). I also remember a source, L. Melfata, who is acquainted with Malbridgian technologists; she too might have some useful information ($E_4$); and so on. Unfortunately, before I can check these possibilities out, the customer wants a report.

Barbara: Good example. You can report the .23 probability that Malbridgia is building ZAP, but you might also report how firm that belief is. In other words, how likely is it that the probability will assume various other values when additional relevant evidence is obtained -- e.g., $E_3$, $E_4$, and $E_5$? We might express that firmness by a range, say, 23% ± 15%, within which we feel 95% sure the probability will fall after looking at the other evidence.

Art: Wait a minute, Barbara. Aren't we talking about an awful lot of work here? The assessment task would involve anticipating all the evidence that could possibly be relevant! We certainly don't think $E_1$ through $E_5$ exhaust the possibilities, do we? And even if we just look at $E_4$, for example, before I actually talk to L.Melfata, I have no idea what she will say. If I have to weigh all these possibilities in my mind before assessing the firmness of the probability, I'm afraid I'm sunk.

Shawn: I have some good news and some bad news, Art. The good news is that it's quite easy to assess the potential impact of all the evidence. The bad news is that the result is trivial. The probability of any verifiable proposition based on all the data will be 0 or 1. In other words, it will turn out either that the proposition is true or that it is false. Not very informative, I dare say!

Phyllis: Maybe we're not interested in all the evidence, Shawn, just some part of it. What we really want to know is how the probability might change as a result of evidence we are likely to obtain.

Sy: That reminds me of some computerized aids I've seen recently. They look at the cues with which the system is prepared to deal, but for which there is as yet no data; and they calculate where the probability of the hypothesis could end up as those data come in (e.g., Speigelhalter, 1985).

Art: That sounds good, Sy, except for one little problem: in most cases such a system doesn't exist. And in most cases, at least, it hardly seems worthwhile to try to list all the cues we think we might obtain and figure out how we would react to them.

Barbara: I quite agree, Art. A related approach, which is much less burdensome, is to assess directly what the likely probability range would be after different amounts of time and effort spent on information collection (e.g., Brown, 1985). We'd expect the range of likely values to be larger, the more time and effort we devote. But we don't need to specify explicitly the information we expect to obtain.

Art: Well, I'm still having trouble with this. We're being asked to make judgments about things we don't know and, I daresay, can't know. How can I ever be sure that I have factored into my thinking all the relevant data that I might encounter?

Phyllis: I'm pretty discouraged, too. We already saw that second-order probabilities run into trouble as representations of imprecision. It now seems they have problems representating completeness of evidence, as well. If we take completeness to mean all possible evidence, the intervals are trivial (0 to 1). If we specify completeness more narrowly, as an explicit list of factors or in terms of the evidence we implicitly expect to discover, the intervals seem quite ad hoc, and might fluctuate significantly with factors that we happen to include or leave out.

Shawn: I quite agree, Phyllis. In fact, I think there is another way to appraoch completeness of evidence that is more promising.

Phyllis: Oh?

Shawn: It focusses on the completeness of specific arguments, rather then completeness in any broader sense. When we have a particular item of evidence, we consider what argument or arguments might be constructed based on that evidence for (or against) the truth of an hypothesis. Then it is a relatively simple matter to consider the completeness or reliability of this argument. We don't need to entertain obscure worries about other hypothetical evidence we might or might not obtain that would support other hypothetical arguments. The important question is: to what extent is this given argument a complete proof that the hypothesis is true? I hope we have a chance to discuss this in more detail later.

### 3.6 What (or Who) is Rational

Barbara: In light of all these difficulties, I think I'd prefer to drop the idea of second-order probabilities. I'll just retreat to my earlier position: selecting and combining probability models is part of the art, not the science, of decision analysis. Let me go further, though, and say that we need methodologies to help us bridge the gap between normative ideals and descriptive realities. I think we should distinguish, along with Keeney and Raiffa (1976), between a normative theory and a prescriptive theory. The former describes an ideal decision maker. But the latter tells real decision makers how to use the normative theory effectively.

Art: For example?

Barbara: A prescriptive theory would contain a variety of pragmatic tools, e.g., methods of sensitivity analysis and methods for generating, comparing, and reconciling multiple analyses. Sensitivity of conclusions to small changes in specific inputs would lead the decision maker to analyze further those particular inputs. Inconsistencies among different analyses would serve as a prompt for the decision maker to dig more deeply into his store of knowledge, to explain and eliminate them. "Prescriptive theory" in this sense is, of course, no more than a codification of the art of constructing a decision analysis.

Colette: The introduction of a new kind of "prescriptive" theory, to make up for the shortcomings of the normative theory, sounds to me like an admission of defeat. If a theory almost always demands more precise inputs than people feel comfortable with, or suggests conclusions people frequently disagree with, then the theory is wrong, not ordinary judgment. I think Bayesian theory is guilty on both of these counts.

Zara: Quite right, Colette. And simply abandoning the idea of higher-order assessments will only make matters worse, in my opinion. For example, higher order uncertainty can have an important impact on decision making.

__Barbara__: I disagree there, Zara. It makes no difference in decision making whether your probabilities are vague or precise. In other words, when it comes time to act, all those second-order probabilities are entirely summarized by a single number: the "average" first-order probability. (The only exception is when you are deciding whether to collect more information.)

__Zara__: Wait a minute, Barbara: that's what the __theory__ says. But in a decision maker's mind, the original ambiguity may be very important.

__Barbara__: I think you're confusing normative and descriptive issues, Zara. Vagueness of probabilities may make a difference in ordinary, unaided decision making. What I claim is that it __ought__ not to matter.

__Colette__: Let me come to Zara's defense here. I would be very surprised if even the staunchest Bayesian were indifferent if he had to choose between two gambles with identical probabilities and utilities for their outcomes, but where one set of probabilities is well-supported by data (e.g., the probability 0.5 that a coin will land heads) and the other set represents ignorance (e.g., the probability 0.5 that McEnroe will win Wimbledon). For all I know the probability of McEnroe's winning might be as low as .3 or as high as .8. Why shouldn't that possibility influence my degree of interest in the gamble? If I am at all cautious, I will prefer to bet on the coin rather than on the tennis match.

__Phyllis__: So your willingness to bet ought to depend on how much knowledge is incorporated into your beliefs?

__Colette__: Exactly! No matter how coherent I am in terms of probability theory, I will lose if I bet with people who have more relevant knowledge than I do (L.J. Cohen, 1980). Following Bayesian norms will lead to disaster! But amount of knowledge is the very factor which Bayesian decision theory, as we have just seen, almost entirely fails to appreciate.

__Barbara__: Well, Colette, I would not claim that the Dutch book is the only argument in support of Bayesian theory. In fact, the requirement that beliefs conform to the probability calculus can be based simply on the inherent plausibility of various constraints on our beliefs associated with the probability calculus--but having nothing whatever to do with betting. Shimony (1970) outlines a set of such constraints, or axioms, e.g., that our degrees of belief in logically equivalent propositions be equal; that our degree of belief in A _or_ B is a monotonically increasing function of our belief in A and our belief in B; etc.

_Zara_: I don't think any of us will deny the interest or desirability of many features possessed by the Bayesian system. Just remember, there is no proof that this particular method for representing uncertainty is the _only_ one that has interesting or desirable formal properties. Indeed, this is far from the case. I would argue that the probability calculus remains unsuited for representing many species of uncertainty or imprecision.

_Phyllis_: The appropriateness of different normative theories may vary as a function of the concepts and goals that characterize a particular problem domain or application?

_Zara_: That's right. One example of a concept not captured by probability theory is the one we have been discussing: weight of evidence or knowledge. A related concept, which I hope we can discuss shortly, is something called "possibility."

_Colette_: I would go much further on this point. Normative theories are no more than systematizations, or idealized descriptions, of ordinary intuitions about what's reasonable in particular cases. Why should we evaluate an inference theory in terms of its _axioms_? What counts is whether it makes sense when we apply it to actual problems.

_Art_: Well, I feel vindicated. As I said earlier, it makes sense to me that a good theory of reasoning would be one that describes how people actually reason. It seems to me that that is the philosophy that lies behind the recently emerging "expert systems" technology in artificial intelligence.

_Barbara_: Let me remind Colette that many Bayesians have regarded _their_ theory as an "idealization" of actual practice.

_Sy_: I hate to be a wet blanket, but the experimental literature in cognitive science suggests that _none_ of the current normative models fits actual practice very well (e.g., Kahneman, Slovic, and Tversky, 1982; Schum and Martin, 1980; Goldsmith, 1983). The discrepancies seem to be systematic rather than random in any meaningful sense, though I suppose one could stretch "idealization" to cover almost anything.

Shawn: Perhaps so. But a more accurate statement may be that before we construct an analysis, our beliefs are usually not precise enough or definite enough to be identified with or distinguished from various inference frameworks (Shafer, 1981).

Phyllis: Along these same lines, Art might be interested to know that in the design of expert systems, computer scientists who specialize in eliciting expert knowledge report that an expert's model of his subject area is not simply "copied" into the computer. It is a moving target, which is transformed in the process of knowledge elicitation. The expert is typically exposed to constraints deriving from the architecture of the system being built, including the mechanisms for performing inference; and these constraints influence the way he expresses his knowledge. Conversely, of course, the selection of system architecture is influenced by its success in capturing what the expert knows.

Shawn: Your comment, Phyllis, fits my conception of inference theory as constructive. We are better off leaving aside abstract arguments about axiomatic derivation or about descriptions of how people "really" think. The ultimate test of a theory is whether its formal properties and its ease of use combine to make it an effective tool. The real question is the extent to which the scheme prompts us to ask questions that either match what we already know or lead to new knowledge. We should ask, how productively can people utilize it to achieve given ends in a particular type of problem?

Art: That sounds good. I guess the next question is, are there alternatives to the Bayesian viewpoint that address issues like precision of inputs or weight of evidence more adequately?

Zara, Shawn, Colette: Yes!

### 3.7 Fuzzy Reasoning

Zara: With the permission of my colleagues, I would like to contribute something on both of those points, drawing on an approach recently developed by a scientist at the University of California in Berkley, named Lotfi Zadeh (1965, 1978). Fuzzy set theory is a direct effort to model the inexactness in human judgment and reasoning. The traditional all-or-none concept of set membership is generalized into a membership function which represents the degree to which an element belongs to a set. This concept of graded set membership turns out to be a very powerful tool for modeling a large number of different types of imprecision within a common framework. Zadeh's work is closely tied to an analysis of how uncertainty is actually expressed in natural language, so it has a strong descriptive component. He argues that in most cases the uncertainty is "possibilistic" rather than (or in addition to) probabilistic.

Barbara: Well, I don't know about natural language, but I doubt there is any meaningful uncertainty to which Bayesian probabilities could not apply. The test, after all, is simply whether or not one could formulate a bet on the outcome of the uncertainty.

Zara: Let me give a simple example that I think does not fit within your framework, Barbara. Assume that you know that the actual number of members in a terrorist organization called Pink Thursday is, say, 250. Yet you are uncertain about the claim that this organization is large. Notice that there is no uncertain outcome to bet on! The problem is that "large terrorist organization" is a fuzzy rather than a crisp predicate. Different sizes have different degrees of membership or "belongingness" in the fuzzy set denoted by "large terrorist organization". For example, 400 is more a member of the set than is 250. The degree of truth in the claim that Pink Thursday is large is expressed by the degree of membership of its size in the fuzzy set "large terrorist organization" (Zadeh, 1978). Note that we are interested in a degree of truth, not in the probability that some proposition is or is not true.

Barbara: I'm not so sure a probabilistic analysis couldn't be applied here, Zara. The relevant uncertainty is whether or not the meaning of "large terrorist organization" includes 250.

Sy: Well, how do we decide that?

Barbara: We could determine the probability that a randomly selected speaker of English would agree that a terrorist organization with 250 members was large.

Zara: I don't deny that you could gimmick up some such analysis, Barbara. But look how unnatural it is. You have to force your respondents to make an all-or-none decision about largeness. I suspect each one of them would be more inclined to regard the meaning itself as a matter of degree.

Sy: This is all very interesting, Zara, but what use is it? In your example, we know Pink Thursday's exact size to begin with, so why should we worry about the degree to which it is large?

Zara: Good question, Sy. One important reason is that much of our knowledge about the world is "fuzzy". So it may be that what makes Pink Thursday's size relevant or interesting to us is (in part at least) what it implies when we use this fuzzy knowledge. For example, we may believe that "large terrorist organizations experience internal organization conflict". It would be senseless to try to translate this into something non-fuzzy by means of exact cut-offs, e.g., any organization above precisely 330 members is large. After all, there isn't an abrupt change from not-large to large at some specific size! So before we can use this knowledge to draw conclusions about Pink Thursday's tendency to experience organizational conflict, we need to know the degree to which Pink Thursday is large. Zadeh's system of fuzzy logic enables us to draw inferences of just this sort.

In the second place, our facts are not always so exact. For example, returning to Art's problem, note that evidence $E_2$ tells us that "a large quantity" of Xyleum was purchased by Malbridgia. We are not told the exact amount. Maybe the source of this report relied on evidence that pointed to a large quantity without specifying how much -- for example, quick visual observations of packages of Xyleum ready for shipment, or the role in contract negotiations of high-level officials. According to Zadeh, though, "large quantity" acts as a constraint which induces a set of possibility measures for different exact amounts. The possibility of 2 kilograms, for example, will be .8 if its membership in the fuzzy set "large quantity (of Xyleum)" is .8.

Finally, notice that the _match_ between our fuzzy general knowledge and the fuzzy evidence may itself be inexact. Suppose, based on our experience as intelligence analysts, that we believe some relevant fuzzy general rule, such as the following: "Usually, extremely large quantities of Xyleum are used to build components of System ZAP." We can use Zadeh's fuzzy logic to consider what the purchase of a "_large_ quantity of Xyleum" ($E_2$) implies in the light of this rule pertaining to "_extremely_ _large_ quantities", and to draw appropriate conclusions about the chance that Malbridgia is now building ZAP.

_Barbara_: Well, Zara, at least now I can see now where we differ. You seem to be talking about vagueness or ambiguity in language. A Bayesian--like most other analysts, I dare say--would try to eliminate _that_ kind of uncertainty before he even began his analysis! I suspect fuzzy set theory has interest _only_ as a description of the way people sometimes reason, rather than as a normative guide.

_Zara_: That depends on how _successful_ analysts are in eliminating fuzziness, doesn't it, Barbara? Even in highly technical contexts reasoning is often fuzzy (Zadeh, 1983a). For example, a doctor observes that a patient is "seriously burned." By the way, that very expression is incorporated within a rule in MYCIN, a computer system designed to replicate medical reasoning. PROSPECTOR, a system for geological reasoning, refers to "abundant" quartz sulfide "veinlets" with "no apparent alteration halos." Such fuzziness is no easier to eliminate than probabilistic uncertainty. Its origin is not sloppiness of language, but incompleteness of understanding. The more complex the phenomenon, the more fuzzy you must be to say anything relevant about it, even if you are an expert. Rather than eliminate such imprecision, and sacrifice relevance, Zadeh offers a way to incorporate it rigorously within an analytical approach.

_Art_: Zara may have a point, Barbara. The evidence in our own problem appears to be full of fuzzy terminology in addition to the example we just looked at: for example, a "very reliable" source, "immediate" purchase, "vital" to the development of ZAP ($E_2$); a "mid-level" government employee, a "usually dependable" source ($E_3$); "recently" reported, "rather boastfully", "several" of the "novel" subsystems, travel "freely" ($E_4$); "influential", "strongly" back, "several" subsystems ($E_5$). It's hard to imagine converting all of these to crisp, yes-or-no questions.

Zara: Even if you did, Art, how meaningful would your result be? You would have had to turn this into a different problem. Zadeh provides a method for deriving what the possible probabilities are from the evidence as you naturally conceive of it.

Phyllis: If I follow you, Zara, there are some cases where we could use either probability or possibility. Could you clarify for us the relation between these two concepts?

Zara: With pleasure--though I warn you the relationship is not simple. It may be helpful first to identify two contexts in which "possibility" plays a role in Zadeh's work. At the broadest level, possibility theory is a systematic framework for interpreting the meaning of natural language utterances. In this framework, as in the example of "a large quantity of Xyleum" a proposition induces a possibility distribution on related (perhaps implicit) variables, in this case, the weight of Xyleum bought by Malbridgia. In other words, given that a large quantity was bought, we can assign a possibility value between zero and one to each weight. This approach can be applied, in principle, to any proposition whatsoever; and in fact, Zadeh has made impressive progress in showing how natural language statements can be interpreted by means of fuzzy logic, and in showing how rules of inference, based on fuzzy set theory, can be applied to them.

The second context in which "possibility" crops up is as a feature within certain natural language statements. Some propositions involve "possibility-qualification"--i.e., hedges like "very possible," "quite possible," "almost impossible." The analysis of these statements is a special case of the application of possibility theory in the broad sense. For example, this analysis involves a fuzzy set corresponding to the hedge, together with procedures for determining the degree of membership of the hedged proposition in that set.

As you might have guessed, possibility and probability are alternative forms of hedge. The analysis of linguistic forms involving probability-qualification is analogous to that of possibility-qualification. Such hedges, called "fuzzy probabilities," include not only expressions like "quite probable" and "not very likely," but also so-called "fuzzy quantifiers" such as "most," "usually," "several," "few," and "more than half." Each of these is a fuzzy set containing numbers between 0 and 1 to varying degrees. Take the sentence, "usually, extremely large quantities of Xyleum are used for System ZAP". The membership of a particular proportion, say .3, in the fuzzy set denoted by "usually" determines the possibility that .3 is the proportion of the time that extremely large quantities of Xyleum are used for ZAP.

Phyllis: So what you're saying is, at the level of specific linguistic forms, possibility and probability are alternative ways of expressing uncertainty. But both of these have a place in a broad possibilistic framework?

Zara: Exactly, if you'll pardon the expression.

Sy: This is all very interesting, Zara, but you still haven't told us how these concepts help us deal with our original problem: weight of evidence.

Zara: The point I have been leading up to, Sy, is that Zadeh's theory can help in two ways: by means of possibility-qualification and by means of fuzzy probabilities. Each of these can serve as a tool for exploring and representing the knowledge that underlies an ordinary probability assessment.

Possibility, you recall, is degree of compatibility with one's knowledge. Assessments of possibility thus depend on what the evidence fails to exclude: i.e., to what degree the evidence fails to prove that the proposition is false. Assessments of possibility are thus fairly conservative, focused on what we know. In contrast, assessing the probability of an event will often require considerable guessing--about what might really have happened--that goes well beyond what the evidence proves or disproves. Possibilities are thus a less demanding sort of assessment and more frequently reflect the form our knowledge actually takes. For example, even if I know very little about Pink Thursday, I could pretty confidently assess a possibility distribution for its size, based on what I know about the range of sizes of such organizations; but I might find it quite hard to assess a probability distribution. According to Zadeh's (1978) possibility/probability consistency principle, possibilities provide an upper bound for probabilities (and thus help determine the possibility distribution of fuzzy probabilities).

A second approach based on Zadeh's theory is the use of fuzzy probabilities. As we noticed earlier, Bayesians insist that probabilities are single numbers, regardless of the amount (or scarcity) of the underlying data. For Zadeh, however, if the information you have is fuzzy, as it so often is, so are the probabilities based on it.

Sy: Well, these fuzzy probabilities sound to me like second-order probabilities under a different name.

Zara: Not at all, Sy. For one thing, a possibility distribution, unlike a second-order probability distribution, can't be summarized by a single number, i.e., the "average". For example, your vague probability that H (e.g., that Malbridgia will build ZAP given that Muldip supports it) might be expressed as a possibility distribution that assigns a possibility of zero to all probabilities below .1 and above .95, and that peaks at .7. When we use it to decide among actions, this distribution does not collapse to a single number.

Phyllis: Well, what can we do with these fuzzy probabilities then?

Zara: Let me give you just one example. An interesting technique proposed by Zadeh (1982) supports the assessment of probabilities for unique events. For example, how are we to assess the probability that Muldip will support the building of ZAP, given that he said he would? There is no large sample of cases where Muldip said he would support ZAP, in which we can determine the frequency with which he actually did support it! Here again, the traditional probability paradigm is unsatisfying: if empirical frequencies are unavailable (and if there is no natural decomposition in terms of other probabilities) our only recourse is direct subjective assessment. Zadeh's approach by contrast, involves an analysis in terms of other events (e.g., other cases where Muldip, or other politicians, in Malbridgia or elsewhere, said they would take some action); the crucial factors are the degree of similarity of each such event to the unique event at issue, and the extent to which those other events possess the property whose probability is being assessed (e.g., the action was taken).

Sy: So, your claim is that there are two "fuzzy" approaches to the assessment of weight of evidence: direct evaluation of the compatibility of a proposition with your knowledge, i.e., "possibility", and assessment of your degree of confidence in a range of possible probabilities?

Zara: Right. And I would argue they are complementary...

Art: If I can interrupt a moment, I'm quite disturbed. Please tell me how all this is going to make assessment any easier? The standard Bayesian theory was pretty daunting, demanding precise numbers where they seemed entirely out of place. Fuzzy set theory, like second-order Bayesian theory, seems to want to measure the imprecision. That may be terrific from the normative point of view, Sy, but what about the poor fellow who has to provide all those numbers?

Zara: Not to worry, Art. Numerical inputs aren't really required. Decision makers would only need to provide verbal descriptions (e.g., "pretty likely" or "about 35%"). Such descriptions could then be automatically translated into fuzzy set membership functions expressing the "possibility" (or confidence) that a particular number was the probability required. Computations would take place with the underlying membership functions, and the results could be translated back into appropriate verbal descriptions. The trick here is to realize that extreme precision is not needed at the level where we assess possibilities--results will be relatively insensitive to inaccuracies--so an approximate verbal approach is entirely justified.

Barbara: I have a feeling that you may run into trouble with your verbal approach, Zara. People might disagree rather substantially on the possibility distributions corresponding to particular verbal expressions. So you would have to get numerical assessments from each decision maker after all. I doubt if the results will be insensitive enough to errors to let you get away without doing that.

Zara: Well, these are questions for empirical and formal research.

Barbara: In any case, I have to admit that the betting paradigm is not much good as a tool for analyzing fuzzy evidence. But you must admit, Zara, it provides a pretty direct link between probabilities and decision making. I wonder what a decision maker would do with these "fuzzy probabilities" once he had them?

Zara: I think we can help him, Barbara. We simply "fuzzify" the computation and comparison of expected utilities (Watson, Weiss, and Donnell, 1979; Freeling, 1980, 1983). In doing so, we utilize a very general inference rule. It tells us how to compute the fuzzy set membership function for a variable when it is a function of variables whose fuzzy set membership functions are known. So if we know the possibilities for the probabilities and utilities of the various outcomes, we can derive the possibilities for the expected utility (and also for the maximum expected utility).

Barbara: Well, the real question is whether all these possibilities finally yield a recommended action.

Zara: In many cases, where one option "fuzzily dominates" another, there is a clear recommendation. In other cases, there are a number of possible decision rules one could adopt (Freeling, 1980). But perhaps we should not expect unambiguous prescriptions when the data are very fuzzy.

Sy: I think part of what Barbara was getting at, Zara, was the normative basis of fuzzy set theory. Why should we accept it as a recipe for belief or for action? What, for example, is the basis of this "very general rule" you just referred to?

Zara: Fair enough. I think the normative appeal of this theory derives from several sources. First, some work has been done--as it has for Bayesian theory--to show that fuzzy set theory is entailed by the acceptance of some quite plausible axioms or assumptions (e.g., Bellman and Giertz, 1973; Fung and Fu, 1975). Secondly, I think individual applications of the theory have their own independent plausibility. Results usually conform to our intuitions about what they ought to be. Thirdly, there is additional justification in the special case where the theory is applied to fuzzy probabilities. Fuzzy logic provides the basis for what is, in effect, a sensitivity analysis or a measure of confidence in the original Bayesian probabilities. For example, the Bayesian probability that two independent events will both occur is, of course, the product of their probabilities. By fuzzifying these probabilities, we can derive the interval within which their product lies at a given degree of confidence. In short, in this application fuzzy logic retains the normative appeal of first-order Bayesian probabilities, while simul taneously relaxing the assessment burden and capturing the imprecision in an expert's process of reasoning.

### 3.8 Belief Functions

Shawn: I couldn't agree more with Zara's central point: the Bayesian framework fails to capture the real significance of the evidence. I think there is a simple explanation of this: Bayesians concentrate on strength of belief in the truth of a hypothesis, rather than on the meaning of the evidence itself. I would like to propose a method that attacks this problem quite directly: Shafer's (1976) theory of belief functions. Rather than "fuzzify" Bayesian probabilities, Shafer, who is a statistician at the University of Kansas, (1976) urges the replacement of Bayesian probabilities by a concept of evidential support. The contrast is between the chance that a hypothesis is true, on the one hand, and the chance that the evidence means that the hypothesis is true, on the other. Thus, we shift focus from truth to the interpretation of the evidence.

Sy: But isn't it the truth that we're interested in?

Shawn: Perhaps, but remember: our only way of finding the truth is through the evidence. So a tool that helps us analyze the evidence may be a lot more helpful than one that focuses our attention directly on the truth of the hypothesis.

Barbara: Hold on a minute! Who said Bayesian theory doesn't help analyze evidence? We've already talked about how the impact of evidence can be quantified using likelihood ratios, and how a large number of subtle interactions among evidence items can be captured using techniques of Bayesian hierarchical inference.

Shawn: Of course, Bayesian techniques permit an analysis of evidence. But I have several complaints, Barbara. First, those techniques often require you to make assumptions that go far beyond anything supported by your evidence. Second, there is no way to represent the amount of knowledge, or weight of evidence, underlying an analysis. Finally, the method of analyzing and com bining evidence is often cumbersome and unnatural. I think you'll see what I mean when I explain how Shafer's theory works.

Recall that in Bayesian theory a probability mass of 1 must be allocated completely among the possible hypotheses. It follows, of course, that the probability of a hypothesis H and the probability of its complement, not-H, must sum to 1. Shafer retains this idea of spreading a fixed amount of some quantity over a set of alternatives. The difference is that we spread this mass among possible _meanings_ of the evidence. We ask: what are the possible interpretations of the evidence, and what are _their_ probabilities.

_Sy_: The notion of the "meaning" of a bit of evidence sounds pretty obscure to me.

_Shawn:_ In fact, Shafer gives that notion a very crisp sense. The possible "meanings" are simply the hypotheses themselves plus all _combinations_ of hypotheses. Let's take Art's problem as an example. We have two hypotheses: H = Malbridgia is building ZAP, and not-H = Malbridgia is not building ZAP.

Now we have some inconclusive evidence $E_4$ in favor of H. $E_4$ is inconclusive since our source, L. Melfata, could be lying or mistaken, the technologist she talked to could have been lying or mistaken, we may be wrong in our assumption that the technical advances mentioned in $E_4$ are needed for System ZAP, and so on. Please note that if any of these contingencies is the case, it does not follow that H is false--that Malbridgia is not building System ZAP. What does follow is that $E_4$ tells us nothing at all about whether it is or is not. Under those circumstances, $E_4$ would be consistent with H _or_ not-H. Shafer represents this by saying that $E_4$ could mean H but also could mean the combination of both hypotheses (H,not-H).

The evidential support which $E_4$ lends to H is simply the probability that $E_4$ means H. Shafer defines a "support function" or "basic probability assignment" $m_4$ to reflect $E_4$'s evidential impact. For example, we might subjectively assess $m_4(H) = .4$ and $m_4(H \text{ or not-}H) = .6$. This captures our intuition that $E_4$ lends some support to H, but no support to not-H. $M_4(H \text{ or not-}H)$ is the probability that $E_4$ says no more than that _something_ is true.

The support function m captures the direct support of evidence for a hypothesis (or combination of hypotheses). But our degrees of belief may also reflect indirect support. Shafer defines the degree of belief Bel in a hypothesis or combination of hypotheses as the total support for any hypothesis (or combination of hypotheses) that implies it. Bel(H) captures the extent to which our evidence means H or means something that necessarily includes H.

The key point here is that some of the support or belief can remain uncommitted to any particular hypothesis. The ability to represent uncommitted belief is a major difference between Shafer and the Bayesians. It means that our modeling need go no further than our evidence takes us.

Barbara: All it really means, Shawn, is that your model is incomplete. The real virtue of the Bayesian approach is that it forces you to take into account all the relevant information. For example, recall that in my treatment of this example, I would first have to assess my prior probabilities or odds for H and not-H, in other words,my degree of belief in each hypothesis before receiving the new data about the technologist's boasting. Then I assess the probability of obtaining $E_4$ on the supposition that each of the hypotheses were true. Finally, I would use Bayes' rule to combine my prior belief with the new evidence. So my new beliefs are logically guaranteed to reflect everything I know about this case.

Phyllis: I have no idea what "logically guaranteed" means, Barbara. I think we've already agreed that a Bayesian analysis wouldn't be psychologically guaranteed to tap all my knowledge about a case. And it may require knowledge that I don't have. We need to talk about real, not idealized, decision makers. An analysis which is logically incomplete, but which more closely matches the way I organize my knowledge, might be more plausible, and convey more knowledge, than the Bayesian one.

Shawn: That's right. I think it's important to reject the idea that the Bayesian theory is automatically appropriate for every problem, just because you can always bet on the truth of the conclusion. The key issue is not whether you can formulate choices among bets that would elicit the required probabilities. The real question is, how confident you feel that these choices and the resulting model capture what is going on. In a game of chance, after all, we bet because of what we know about the probabilities; we don't learn about the probabilities by observing our tendency to bet!

Phyllis: This point seems related to your concept of constructing probability arguments.

Shawn: It certainly is. To the extent that Bayesian theory has anything to contribute, it is by establishing a persuasive analogy between your problem and a situation, like poker or a lottery, where the truth is generated by known chances. We construct Bayesian probability models by reference to such comparison cases, or "canonical examples" (Shafer, 1981). Such analogies, however, will usually be imperfect, because in the canonical example we know the rules of the game that determine how the truth is generated (e.g., the composition of the deck and the procedure for dealing). In real problems, there are nearly always many aspects of the situation where comparable rules cannot be given without making numerous assumptions. When these assumptions become very extensive, it may be better to switch to a simpler kind of model, which is more plausible despite not giving a complete picture of how the truth is generated. Such simpler models can be based on canonical examples in which the meaning of the evidence rather than the truth is generated by known chances.

Sy: So where did Barbara go wrong?

Shawn: I think both in her treatment of prior probabilities and likelihoods. To start with, it is often hard to see where prior probabilities should come from in a Bayesian analysis.

Barbara: Perhaps a new example will help. Suppose a prominent politician in an allied country has been kidnapped; we know that a terrorist organization is responsible and that only four such groups, which we designate A, B, C, and D, could have done it. Our hypotheses are $H_A$, $H_B$, $H_C$, and $H_D$ (that A is the group responsible in the kidnapping, that B is responsible, etc.). If I have no reason to suspect one of the four groups more than another, I can set the prior probabilities equal to one another: $P(H_A) = P(H_B) = P(H_C) = P(H_D) = .25$.

Shawn: Thanks, Barbara. Your example makes it clear that the only way to represent ignorance in the Bayesian theory is to allocate probabilities equally among a set of alternatives. But there are some serious objections to this approach. As a very large number of critics have pointed out, how probabilities are in fact allocated will depend on how the alternatives are described or scaled. We do not know whether A, B, C, or D is responsible, so we assign a probability of .25 for the guilt of each one. But suppose group A originates in the Middle East and groups B, C, and D originate in Western Europe. We have no reason to believe the kidnappers are from the Middle East versus Western Europe. So perhaps we should assign a probability of .5 to $H_A$ and .167 to each of $H_B$, $H_C$, and $H_D$. In the end, the assignment of priors based on "ignorance" is quite arbitrary.

Notice how naturally Shafer handles this case. Our prior beliefs consist only in the knowledge that A, B, C, or D is responsible, and this is all we have to say in our model. We represent this (before receiving any evidence) by assigning support equal to 1 to $\{H_A,H_B,H_C,H_D\}$. The allocation of probability mass within this set is simply unspecified.

Barbara: Some Bayesians would argue that there is no such thing as total ignorance. We always have some prior knowledge, however vague, and this should be reflected in the priors we assess. For example, Art has a prior expectation that Malbridgia is not building System ZAP, based on a briefing he heard about U.S. technical capabilities.

Shawn: Where there really is some knowledge, we can and should represent it. But we can do so in terms of a Shaferian analysis of evidence rather than as "prior probabilities." And where there isn't any knowledge, we shouldn't have to make arbitrary choices.

In Art's case, the briefing is evidence, even though indirect. If we assume that Malbridgia's technology tends to equal or lag U.S. technology, and if the American scientist was honest and accurate, then this evidence means not-H (i.e., that Malbridgia is not building ZAP). Otherwise, the evidence means nothing at all, i.e., {H,not-H}. Since there is a fair chance that at least our first assumption is mistaken (Malbridgia's technology could be ahead of ours), we might have $m_0$(not-H) = .7 and $m_0$(H,not-H) = .3.

Barbara: It seems to me that likelihood ratios are a perfectly appropriate way to represent the impact of evidence.

Shawn: I'm afraid I don't agree, Barbara. Some of the same problems arise in the assessment of likelihoods as in the assessment of priors. In order to estimate the probability of the evidence given a hypothesis, we are forced to include cases where the evidence occurs even though it is in fact not connected to the hypothesis. In the case of $E_4$, we mentioned the possibility of deception or error by L. Melfata or by the technologist, or invalidity of our assumption about the relevance of the technology to ZAP. In all these cases, even if Malbridgia did happen to be building ZAP, $E_4$ would not have really been relevant, and so they are not included in our measure of the support lent by $E_4$ to H, $m_4(H)$. We have no evidence bearing on the chance of any of these eventualities. Nonetheless, the Bayesian probability of $E_4$ given H must somehow purport to model them, at least implicitly.

In Shafer's framework, there is no such requirement. We directly assess the probability that the evidence means, or is connected to, the hypothesis. Of course, we leave open the possibility that the evidence occurred by chance: i.e., the hypothesis is true and yet the evidence doesn't mean the hypothesis. This is included in the support assigned to {H,not-H}. Since this support is uncommitted among the two hypotheses, no specific modeling of what happens when the evidence is not linked to the hypothesis is needed.

Barbara: I'm not as convinced as you seem to be, Shawn, that there is a clear line to be drawn between cases where we do and do not have "knowledge." For example, although we have no direct evidence regarding the possibility of error or deception, we are likely to have some idea, based on our past experience and general information, about these probabilities. The Bayesian framework, by insisting that we come up with the numbers, may draw more information out of us than we knew we had. In terms of Shafer's own constructive theory, the knowledge may not pre-exist in a very appropriate or accessible form, but the assessment task itself can stimulate us to construct a representation that captures it.

Shawn: Well, the proof is in the pudding. If you can use Bayesian theory to come up with convincing analyses, please do. In fact, the Bayesian theory is a special case of Shafer's, where all support is assigned to single hypotheses, and none is left uncommitted.

This issue, however, is related to my second major complaint about the Bayesian framework. It provides no distinction between probabilities which are based on evidence and those that are not. What we need, and don't get, is a way of representing the weight of evidence that underlies an analysis.

As we discussed earlier, probabilities themselves are simply not appropriate measures of the quality or credibility of an inferential argument. An estimate that there is a 90% chance that Malbridgia's building ZAP would not necessarily be better supported than an estimate that puts the chance at 50%. One would not have much confidence in a conclusion (no matter how high the Bayesian probability) if it requires numerous untested assumptions. Conversely, the 50% probability could reflect the outcome of a thorough sifting of evidence bearing in the validity of those assumptions. The credibility of the conclusion depends on the <u>completeness</u> with which relevant and available evidence has been consulted, not on the probabilities assigned to the events in question.

<u>Art</u>: That sounds like a pretty good answer to some of our customers. They seem to believe that a "good" analysis is one that eliminates all uncertainty.

<u>Shawn</u>: It may be more correct to say that a good analysis effectively exploits the available evidence to determine what our uncertainty assessments <u>are</u>.

Compare three cases in our kidnapping example: (1) I set my prior probabilities $P(H_A) = P(H_B) = P(H_C) = P(H_D) = .25$ entirely on grounds of ignorance and symmetry; (2) I set my prior probabilities equal to one another because I discover that groups A, B, C, and D each committed comparable kidnappings during the past 5 years; (3) I set my prior probabilities equal to one another because the manner in which the kidnapping was carried out has definite, but comparable, similarities to the modes of operation of each of the four groups.

Now it seems clear to <u>me</u> that each of these arguments has quite different merit. Yet a Bayesian analysis will treat them the same. By contrast, in Shafer's system we might have three very different support functions:

(1) $m(H_A) = m(H_B) = m(H_C) = m(H_D) = 0;$
$m(H_A, H_B, H_C, H_D) = 1$

(2) $m(H_A) = m(H_B) = m(H_C) = m(H_D) = .1;$
$m(H_A, H_B, H_C, H_D) = .6$

(3) $m(H_A) = m(H_B) = m(H_C) = m(H_D) = .2;$
$m(H_A, H_B, H_C, H_D) = .2.$

In each case (as in the Bayesian analysis) the support assigned to the individual hypotheses is equal. But our degree of confidence that each of these arguments is valid is reflected in the varying degree of support assigned to the universal set, $(H_A, H_B, H_C, H_D)$. The sum of the support for the individual hypotheses and the universal set, in each case, is 1.

Sy: I like that. Your method of representing belief allows you to give greater force to an argument based on empirical frequencies, like (2), or direct evidence, like (3), than one based on ignorance (1).

Shawn: Notice, however, that we need not give full credence to frequency arguments either. Often it happens that although empirical data are available, their exact relevance to the issue at hand is not entirely clear. For example, in regard to (2), A, B, C, and D's respective circumstances and goals may have changed drastically, in quite different ways, since they committed their earlier kidnappings. We may feel uncomfortable with an unqualified extrapolation of the past into the future. More fundamentally, we may feel uncomfortable treating the present kidnapping as if it were drawn randomly from the pool of kidnappings committed by the four groups. The .6 support assigned to $m(H_A, H_B, H_C, H_D)$ is a discounting factor that represents these elements of doubt (Shafer, 1982).

Sy: In a way, then, the belief assigned to the universal set gives you a measure of the incompleteness of your evidence.

Shawn: In this example, it does; but we can be a bit more general. In a sense, Shafer's Bel function is the *lowest* degree to which a hypothesis can be believed, because it is the lowest level of belief forced on us by the evidence. But we have left open the possibility that the hypothesis is true in a way not revealed by our evidence. So another function introduced by Shafer is the degree of plausibility of a hypothesis, Pl(H), corresponding to the maximum possible belief in H. Pl(H) is equal to the total belief assigned to H (i.e., Bel(H)) <u>and</u> to combinations of hypotheses containing H. The latter belief is uncommitted, but new evidence could cause it to go to H. Pl(H) is thus the extent to which the evidence is not incompatible with H. So, Pl(H) = 1-Bel(not-H).

<u>Sy</u>: I think I see. The interval between Bel(H) and Pl(H) gives a range of belief for H. Its size measures the scope for new discriminations among hypotheses to affect our belief in H; hence, in a sense, it measures the present incompleteness of our evidence.

<u>Shawn</u>: Right. But keep in mind our earlier discussion: incompleteness here is the chance that a particular collection of evidence fails to discriminate a hypothesis from other possibilities. So it reflects the reliability of a <u>particular</u> evidential argument. It makes no attempt to measure how much of the total possible evidence we have obtained.

<u>Art</u>: The latter seems pretty impossible anyway.

<u>Sy</u>: That's certainly true. On the other hand, it may be quite natural to assess the reliability or strength of a particular argument. If you think about it, each argument requires us to assume a specific kind of relationship between the hypothesis and the evidence. The evidence establishes the hypothesis only if the relationship presupposed by the argument is in fact the case. For example, when we use frequency data to estimate probabilities, we assume the sample is representative of the population. If we rely on eyewitness testimony, we assume the witness was in a position to see what she says she saw, and that she was motivated to tell the truth. It seems reasonably realistic for us to assess the probability that assumptions like these are met.

<u>Shawn</u>: This notion of the completeness or reliability of an <u>argument</u> is crucial to understanding what the interval between Bel(H) and Pl(H) means. In no sense is this a fixed bound on what our belief in H could eventually be. Ad-

ditional arguments, based on new evidence could come in that conflict with our present beliefs and thus lower Bel(H) or increase Pl(H). Bel and Pl are part of our analysis of our present evidence. They function as a bound only if we assume that all future evidence is consonant with the evidence we already have.

Art: What does "consonant" mean?

Shawn: Essentially, it means that different bits of evidence vary in the precision of their support i.e., their ability to discriminate hypothesis, but they do not conflict. More accurately, consonance means that support goes only to nested subsets of hypotheses. For example, suppose in our kidnapping example we get three new items of evidence. The first, an anonymous statement issued to the press by the kidnappers, is inconsistent, in its wording and content, with previous statements by one of the groups (D). So we represent this by a support function assigning some support to $\{H_A, H_B, H_C\}$. The next item of evidence (e.g., a demand for money by the kidnappers) might give support to $\{H_A, H_B\}$ since groups A and B are known to be in financial straits. Finally, an eyewitness report turns up regarding the types of weapons carried by the kidnappers, and supports $\{H_B\}$. This evidence is consonant, since the arguments based on the different pieces of evidence differ from one another only by being more precise or more general. They could be valid at the same time--if group B is responsible.

Art: Shafer's plausibility measure, Pl, reminds me of Zadeh's notion of possibility.

Shawn: Quite right, Art. It turns out that if we assume consonance of evidence, Pl is a possibility measure (Dubois and Prade, undated). You recall that Zara described the possibility of a proposition as its compatibility with one's knowledge or evidence. So there is a conceptual, as well as a formal, affinity.

Zara: Well, that's one way of looking at the relationship. Another point of view is to think of Shafer's theory as giving a possibility distribution for the probabilities of a proposition, rather than for the proposition itself. In other words, all probability values lying between Bel and Pl are possible; all others are not. This is a special kind of distribution in which all the possibility values are 0 (outside the interval) or 1 (inside the interval). So, from this point of view, Shafer's theory is a special case of Zadeh's, rather than the other way around. Possibility theory, of course, allows gradations of possibility.

Sy: One thing we've been hinting about is how Shafer's theory handles new evidence. For example, Sharon mentioned that the interval between Bel and Pl reflects incompleteness of evidence. So I presume that when new evidence comes in, that interval shrinks. Does Shafer's theory tell us how that happens?

Shawn: It does indeed. So far, Sy, we have been focusing on the representation of evidence in Shafer's theory. But we could just as easily have introduced his theory in terms of how it handles the combination of evidence. As you recall, in Bayesian theory we can divide a problem into simple components, make assessments, and then combine them using Bayes' rule. Now Shafer argues that his framework comes closer to capturing the traditional concept of a "probability argument" than Bayes' rule!

Let's take a simple example. In Art's problem, suppose once again we have only one item of evidence, $E_4$, with a support function $m_4(H) = .4$ and $m_4(H,\text{not-}H) = .6$. Now suppose we receive a second bit of evidence, $E_5$, regarding the assertion by the politician P. F. Muldip that he would support building ZAP. We regard this new evidence, like the first, as inconclusive: the report of Muldip's assertion may be dishonest or mistaken, Muldip may have intended to back ZAP but changed his mind or lost his position, Muldip may have lied, Muldip may have backed ZAP but been overridden in the final decision. We represent the new evidence by a belief function with $m_5(H) = .3$ and $m_5(H,\text{not-}H) = .7$, the latter once again reflecting the chance that this evidence is irrelevant. Now what is our net belief resulting from $E_4$ and $E_5$?

The basic idea is simple. The combined evidence proves that Malbridgia is building ZAP if at least one of the evidential arguments is valid. The probability that both are invalid is:

$$m_{4,5}(H,\text{not-}H) = m_4(H,\text{not-}H) \times m_5(H,\text{not-}H)$$
$$= (.6)(.7) = .42.$$

This is the aggregate support for {H,not-H} based on the combined evidence, and, of course, it is lower than the uncommitted support in either $m_4$ or $m_5$ by itself. The aggregate support for H, then, is:

$$
\begin{aligned}
m_{4,5}(H) &= 1 - m_4(H,\text{not-}H) \times m_5(H,\text{not-}H) \\
&= 1 - .42 = .58.
\end{aligned}
$$

A more general way to look at this same argument is this: The combined evidence $E_4 + E_5$ "means" H when the common element in the meanings of $E_4$ and $E_5$ is H. This happens in three cases: (1) $E_4$ = {H}, $E_5$ = {H}; (2) $E_4$ = {H}, $E_5$ = {H,not-H}; and (3) $E_4$ = {H,not-H}, $E_5$ = {H}. The probability that $E_4$ + $E_5$ means H, then, is just the sum of the probabilities for (1), (2), and (3):

$$
\begin{aligned}
m_{4,5}(H) &= m_4(H) \times m_5(H) + m_4(H) \times m_5(H,\text{not-}H) + m_4(H,\text{not-}H) \times m_5(H) \\
&= m_4(H) + (1-m_4(H)) \times m_5(H) \\
&= .4 + (.6)(.3) = .58.
\end{aligned}
$$

<u>Sy</u>: What if the new evidence conflicts with the old evidence? Suppose we have $E_4$ supporting Malbridgia's building ZAP as before, but now receive $E_1$, suggesting that Malbridgia does not have the technical capacity to build ZAP.

<u>Shawn</u>: The logic is essentially the same. The combined evidence proves that Malbridgia is building ZAP only if $E_4$ is valid and $E_1$ is invalid. In other words, the combined evidence $E_4$ + $E_1$ means H only when $E_4$ = {H} and $E_1$ = {H,not-H}, since this is the only pair of meanings for $E_1$ and $E_4$ that has H as a common element. But notice that <u>both</u> evidential arguments cannot be valid; i.e., Malbridgia cannot both build and not build System ZAP. So the chance the combined evidence means H must be normalized to exclude the impossible case where $E_4$ = {H} and $E_1$ = {not-H}. If we let $m_1$(not-H) = .9 and $m_1$(H,not-H) = .1,

$$
m_{1,4}(H) = \frac{(1-m_1(\text{not-}H)) \times m_4(H)}{1-m_1(\text{not-}H) \times m_4(H)}
$$

$$= \frac{(1-.9)(.4)}{1-(.9)(.4)} = .06$$

and $$m_{1,4}(\text{not-H}) = \frac{(1-.4)(.9)}{1-(.9)(.4)} = .84.$$

Sy: These arguments seem fairly straightforward. But is there a rule in Shafer's system that takes the place of Bayes' rule?

Shawn: The rule of combination for Shafer's system is called Dempster's rule, and it is essentially just a generalization of the kind of intuitively appealing "probability arguments" I have just described. Dempster's rule is more general in that it can be used to combine support functions that make use of the full representational capability of Shafer's system: i.e., where support can be provided by a bit of evidence for any number of hypotheses or combinations by hypotheses.

I wouldn't really say it "takes the place" of Bayes' rule though, Sy. Bayesians sometimes talk as if the main use of Bayes' rule was to update beliefs automatically as new evidence comes in. That presupposes that we have anticipated the possible evidence ahead of time and assessed the relevant likelihoods. But I think it's very rare that we can do so. For example, in Art's problem, the probability of obtaining a report about a technologist boasting ($E_4$) was very small indeed before it was actually obtained.

Phyllis: This point reminds me of your notion of constructing a probability analysis, rather than eliciting it. Only ideal decision makers, not real people, come fully prepared with "true" probabilities for all possible contingencies.

Shawn: Exactly. And there's a stronger point. Even if we had anticipated the actual evidence, the real-life Bayesian may not be wise to abide by the results of his automatic updating. Other new knowledge, not previously anticipated, may well have been acquired along with the evidence, which changes his assessments of the likelihoods used in updating. In other words, it's impossible to anticipate all the ways our _interpretation_ of the impact of a given bit of evidence could change.

_Sy_: But if your revised beliefs don't conform to Bayes' rule, won't you be incoherent, and subject to de Finetti's Dutch book?

_Shawn_: That's a common misundertanding. Coherence only requires that your beliefs at a given time be consistent with one another. It does not require that a current probability cohere with _previous_ priors and likelihoods (Shafer, 1981; Horwich, 1982). In my view, the appropriate use of an inference framework is to organize our knowledge as it exists at a particular time, not to try to control its process of growth and change (Shafer, 1981). The focus of Shafer's system is the _combination_ of available evidence, rather than updating beliefs.

_Barbara_: As you know, Shawn, Bayes' rule can also be used in an analysis _after_ the evidence is obtained.

_Shawn_: Yes, but that brings me to my third major complaint: Bayesian methods are not a natural or efficient way to express our knowledge. In this context they are inefficient because we have to assess probabilities for $E_4$, given the various hypotheses, even though we already know $E_4$ has occurred. They are unnatural because we are asked to imagine a counterfactual situation in which we imagine we do not know that $E_4$ occurred. Now what _are_ we supposed to know in that situation? Certainly not exactly what we really knew before $E_4$ occurred. We may have acquired considerable relevant knowledge since then which can, and should, affect our evaluation of $E_4$. Some of this knowledge may even have been obtained on account of $E_4$. Yet we are asked to imagine this knowledge without knowing that $E_4$ did occur. By contrast, in the assessment of Shaferian support functions, the question is about an actual rather than a counterfactual state of affairs: what is the probability that the evidence in fact means H?

_Barbara_: People do seem able to provide Bayesian likelihood assessments, though I grant that it's easier with a recurring rather than with a unique problem. In any case, as we discussed earlier, there are other forms of Bayesian analyses than the use of Bayes' rule. For example, the probability of an event H can be broken down into its probability given some other event

C, its probability given not-C, and the probability of C. Perhaps analyses such as this are more natural when we are organizing our beliefs at a given time rather than updating them.

Shawn: I entirely agree, Barbara. Let me point out a major advantage of Shafer's theory in this regard, however. In any Bayesian analysis other than the use of Bayes' rule, each assessment is made (in theory, at least) with all the evidence; we are supposed to make use of all of our knowledge in all judgments. The problem is broken down into simpler questions, but the evidence itself is not decomposed.

Phyllis: But we agreed earlier that real decision makers are seldom able to keep all the evidence in mind at once.

Shawn: That's right. In contrast, Shafer's system permits us to focus on parts of the evidence separately. Each support function describes the impact of a distinct collection of evidence. Dempster's rule of combination can then be used to pool the different support functions into a new support function reflecting all the evidence. Shafer's system provides a way of thinking about different parts of the data separately, like Bayes' rule, but does so naturally.

Sy: Shawn, you have argued that Shafer's system supports "intuitively appealing" arguments and that it is a "natural way" of decomposing evidence. In terms of Shafer's concept of a "constructive" theory, perhaps you would claim that these considerations add up to a normative argument in favor of Shafer's system. But it all seems a bit flimsy to me compared to the rigorous axiomatic derivations associated with the Bayesian system. For example, if I follow Dempster's rule rather than Bayes', wouldn't I be subject to a Dutch book?

Shawn: In fact, you won't be. There is a natural interpretation of Bel and of Pl in terms of betting which will lead to "rational" behavior even in the Bayesian sense. An assumption implicit in de Finetti's Dutch book argument is that if 70 cents is the most money you would be willing to pay for a gamble that pays 1 dollar if H is true, then you would be willing to pay 100-70 = 30 cents for a gamble on not-H. If this assumption is rejected, you can avoid a Dutch book without being a strict Bayesian (e.g., Smith, 1961). In terms of

Shafer's theory, we use Bel(H) to determine the stakes at which we would bet on H, and we use 1-Pl(H) = Bel(not-H) to determine our willingness to bet on not-H. To the extent that our knowledge is incomplete (i.e., Pl(H) > Bel(H)), we withhold willingness to bet.

Let me stress though that I don't think avoidance of a Dutch book is the principle rationale for an inference theory. In fact, the structure of Shafer's theory is richer than what would be necessary simply to support a theory of betting. The main goal is to illuminate the evidence by comparing it to well-understood paradigm cases, or canonical examples. For Shafer's theory, the canonical examples concern cases where the meaning of the evidence is gen erated by some chance process, like a lottery. The real justification of Dempster's rule, therefore, is that it is the appropriate rule for this set of examples.

### 3.9 Inductive Probabilities

Colette: L.J. Cohen (1977) has proposed a theory of inference which is simpler than other views both mathematically and in its assessment requirements. It seems to me, moreover, that this theory gets closer to what we really mean by completeness of evidence. Cohen, who is at the University of Oxford in England, suggests a framework based on the factors that could, in principle, prevent a conclusion's being established by your evidence. The "inductive probability" of a hypothesis is defined as the number of such factors which have been tested and ruled out.

Phyllis: It sounds like Cohen is quite close to Shafer. Aren't they both interested in the extent to which an argument can prove the hypothesis, in contrast to the Bayesian's interest in the truth of the conclusion?.

Colette: That's right. But there's an important difference. Cohen is not interested in the probability that the evidence means, or proves, the hypothesis. His central claim is that evidence for a proposition is incomplete as a function of the number of different kinds of ways the analysis could turn out to be invalid.

Art: It sounds like that might simplify the assessment task quite a bit.

Colette: Precisely. We could assess our confidence in a hypothesis simply by counting the number of potentially invalidating factors that have been ruled out. From Cohen's point of view, Bayesians and Shaferians seem to have things backwards. They give abstract formalisms for manipulating probabilities or degrees of support, but they tell us very little about where the numbers should come from. It doesn't help to hear that we should not adopt a conclusion "because" degree of support is too low or "because" we won't bet on it at sufficiently low odds. What we need to know is why: what reasons are there to withhold acceptance? So Cohen suggests that we look directly at the tests that a hypothesis would have to pass before we believed it. We should withhold acceptance of a hypothesis when there is potential evidence we haven't looked into yet which could disconfirm it.

Sy: But if both Cohen and Shafer are concerned with provability of a hypothesis by evidence, I'd expect some relationship between the two theories.

Shawn: Indeed there is, Sy. Cohen's system is a special case of Shafer's--just like Bayesian theory and possibility theory, I might add. It happens to be the same special case associated with Zadeh's theory: consonance of evidence. The idea is that new evidence progressively eliminates possibilities, rendering our beliefs more and more precise; but new evidence does not lend support to conflicting hypotheses.

Colette: In fact, Sharon, two critical advantages of Cohen's theory arise from this restriction: its simplified metric, based on counting, and its incorporation of the notion of accepting a hypothesis. We'll talk about that later.

Art: So how does Cohen's theory work?

Colette: Why don't we take our evaluation of $E_4$ in Art's problem as an example. The first question we want to consider is whether L. Melfata is telling the truth.

Art: So the hypothesis we're examining is that L. Melfata really heard the technologist boasting?

Colette: Right. Let's call that hypothesis B. But we now have to distinguish two senses of "hypothesis" and "evidence" in L.J. Cohen. In one sense, our hypothesis is the specific event, B, as you suggested, i.e., Melfata heard the technologist; and the evidence for B is Melfata's testimony together with our other information about Melfata, as summarized in $E_4$. But Cohen believes that in reasoning of this sort, there must be a generalization or uniformity that justifies, to some degree, the inference from $E_4$ to B. In our example, this general hypothesis would concern human truth-telling behavior and would lay out pertinent grounds for believing that what someone like Melfata says is true. Now suppose we formulate such a generalization, tailored to our present example. According to this generalization, anyone with certain traits who says something with certain properties in a context with certain characteristics, is telling the truth. We fill in the blanks here with the information we have about Melfata, about her testimony, and about our overall problem. This generalization might say that anyone who has had the opportunity to observe what he reports, whose report is internally consistent, and whose report is coherent with at least some other evidence, would be telling the truth. Now our evidence for (or against) this hypothesis is to be found in our accumulated knowledge about when, where, how, and why people tell the truth. If this hypothesis were a general hypothesis in a scientific domain, the relevant knowledge might even be a series of rigorous experimental tests.

Sy: I think I see where this is heading. To the degree that the general hypothesis about truth-telling is supported by relevant knowledge, we are justified in inferring B from $E_4$. So the support for our generalization about truth-telling can be construed as the degree of "provability" of B by $E_4$. The generalization in effect establishes the link between our (specific) evidence and our (specific) conclusion.

Colette: Exactly. Cohen introduces two measures corresponding to the two types of hypotheses, and they are related just as you described. The first is "inductive support": This is the extent to which a general hypothesis has passed the tests which could falsify it. The second measure is the "inductive probability" of a specific conclusion conditional on specific evidence. Here's how the connection works.

Imagine a simple syllogistic argument:

(1) x is an R;

(2) All Rs are Qs;

(3) Therefore, x is a Q.

This is, of course, an example of "deductive inference." Cohen's notion of inductive inference generalizes this to the case where we are uncertain of the generalization, all Rs are Qs, hence also uncertain of the conclusion, x is a Q. Now let s[(2),K] be the inductive support afforded to (2), the generalization, by our knowledge K. And let $P_I$[(3),(1)] be the inductive probability of (3), the conclusion, given (1)--i.e., the probability that x is a Q given that x is an R. Then, for any integer i,

$$s[(2),K] \geq i$$

implies that

$$P_I[(3),(1)] \geq i.$$

Shawn: This is very reminiscent of Shafer's ideas. For him, the direct support of the evidence (1) for the conclusion (3) is the probability that (1) means or proves (3); in this example, that is just the probability that the generalization (2) is true. If (2) is false, the argument from (1) to (3) is invalid, and (1) is irrelevant--i.e., it supports {(3),not-(3)}.

Zara: Well, there is definitely a similarity between Shafer and Cohen. Both of them ignore the fact that terms utilized in reasoning (R and Q) are likely to be fuzzy rather than crisp.

Sy: Good point, Zara. A difference between Shafer and Cohen, though, is in how the measures of inductive probability $P_I$ or support m are arrived at.

Colette: That's right. For Shafer, m is assessed by direct judgment, and is interpreted numerically as if it were a real probability. For Cohen, as I said before, $P_I$ is just a count. And he provides a framework in which it can be derived.

<u>Sy</u>: I remember complaining that Bayesian theory gave us little or no guidance in selecting probability values, so I'm curious about what Cohen has to say about this.

<u>Colette</u>: Cohen believes that in every domain where we want to know the truth about something, such as a field of science or human behavior, there is a set of "relevant variables" associated with the general hypotheses in that domain. In a scientific investigation, for example, the relevant variables represent potential explanations of a phenomenon that compete with the hypothesis of interest. So they are factors which must be experimentally controlled before the hypothesis can be accepted. The scientist performs a series of studies in which, hopefully, the phenomenon of interest continues to be observed despite variations in each of those factors.

<u>Sy</u>: So the support for the hypothesis equals the number of tests that are successfully performed?

<u>Colette</u>: Right. Each test eliminates one more way the hypothesis could have been wrong. The tests are ordered in terms of the importance of the variables.

<u>Sy</u>: This may or may not make sense in an experimental science, Colette, but I have a real problem seeing its relevance to a field like intelligence analysis. It's hardly usual or even possible to test general hypotheses about what countries are up to.

<u>Colette</u>: The main difference, I think, is that we have to draw on our informal, implicit experience to determine how each relevant variable affects the truth of a hypothesis, rather than on formal experiments.

<u>Art</u>: Even in the scientific example, though, we had to rely on such experience to determine what the relevant variables <u>were</u>.

<u>Colette</u>: That's right. Let's go back now to our evaluation of $E_4$ in Art's problem and start with a very simple general hypothesis: Everyone who speaks is always telling the truth. Based on our experience, this claim has little or no inductive support. The reason is that it fails to pass the pertinent tests. Consider all the variables that might falsify it: We distrust a

speaker if he has a certain demeanor (shifty eyes, etc.), if what he says is internally inconsistent, if what he says is independently implausible, if it clashes with other evidence, if we have certain facts about his motives or character or his opportunity to obtain knowledge of what he reports, and so on. Let's suppose we have identified 20 such variables. To get a valid generalization, we can qualify the original claim by stipulating specific levels on each of the 20 relevant variables which are advantageous to the truth of the hypothesis: anyone who speaks with a relaxed demeanor, gives a consistent and plausible report, etc., is telling the truth. Based on our general knowledge of people, the resulting very lengthy hypothesis would have maximal inductive support.

Sy: So any hypothesis that involved only some of the required qualifications would have only partial support?

Colette: That's right. If advantageous levels on only i of the 20 relevant variables are mentioned in the hypothesis, its support is i/20. For example, it's likely that Melfata had an opportunity in her travels to hear the technologist boasting; her report is internally consistent; and it coheres with at least some of our other evidence. So our generalization which we use to justify the inference from $E_4$ to B will include three qualifications that are advantageous to its truth. Thus, we might conclude that the inductive probability of B, Melfata's telling the truth, given $E_4$ is 3/20.

Barbara: It's pretty clear that 3/20 isn't a Bayesian probability!

Colette: Certainly not. It measures the completeness or weight of evidence behind the hypothesis, B, that Melfata is telling the truth. It does so simply by counting the number of relevant kinds of evidence that have been covered and which are favorable to B. When the inductive probability of a hypothesis B is low, it doesn't mean that the inductive probability of not-B is high. In fact, the inductive probability of not-B will be zero, since we can't have conflicting support. What a low inductive probability does mean is that there is potential evidence that has not yet been considered.

Shawn: It's like Shafer's support function again, in that you can have positive support for B and zero support for not-B. B falls short of maximal support because some of it goes to {B,not-B}--corresponding, I suppose, to

Cohen's unconsidered evidence. In both theories, new evidence can increase support for B without decreasing support for not-B. I agree with Colette that this accords much better with our intuitions about evidential impact than the Bayesian approach.

Art: Something has been puzzling me, Colette. In our discussions of Bayes and Shafer, we assessed the impact of $E_4$ directly on H, the hypothesis that Malbridgia is building ZAP. Why aren't we doing that in Cohen's case?

Colette: Well, we could do that if there were a pertinent general hypothesis to establish a link between $E_4$ and H. But there are just too many different kinds of considerations involved to expect a single generalization here involving a single domain of knowledge or experience. In Cohen's system, we would probably construct a separate analysis for each inference: from $E_4$ to the truth of what $E_4$ reported, from the technologist's boasting to the truth of what he said, and from that to Malbridgia's building ZAP.

Barbara: That's exactly why in the Bayesian analysis we talked about breaking the assessment down into a hierarchical or cascaded inference. We probably do have more confidence in our assessments when we consider the steps in this chain separately. But, as Art said, we had the option of assessing $E_4$'s impact on H directly. And the result probably succeeds in capturing a fair amount of our relevant knowledge.

Sy: In Bayesian hierarchical inference, you can compute the probability of the ultimate conclusion, H, based on uncertainties in each link of the chain. Is there some similar way to combine the results of the separate inferences in Cohen's system?

Colette: There isn't. And the reason brings out an important philosophical difference between Cohen and the other views we have considered. For Cohen, inference is domain-dependent; each inference process takes place in its own closed universe of relevant variables. Cohen's measures of inductive support and inductive probability are appropriate for comparisons of hypotheses within a domain, but not for comparisons that cross domain boundaries.

Sy: So the guidance we get in assessing inductive probabilities comes at a steep price. We can't do as much with them once we have them.

Colette: Well, don't be too hasty. I'm not sure it's so bad to be forced to be honest about what we know and don't know. Cohen's theory makes us partition the inference process according to natural divisions in our knowledge.

Zara: But I think it's worth pointing out that, at least from my point of view, this feature of Cohen's system is not essential. We could drop its reliance on general hypotheses and the derivation of "inductive probability" from "inductive support." We then have a theory of the "necessity" of a hypothesis that complements Zadeh's theory of "possibility." Both have attractive ordinal, or "counting," properties.

Shawn: That's because both of them are special cases of Shafer's system, where evidence is consonant.

Colette: But there are some things you can do with inductive probabilities that you can't do with Bayesian probabilities or Shaferian belief functions.

Sy: Such as?

Colette: You can draw conclusions, in the ordinary rather than the probabilistic sense of the word. In the hierarchical inference case, if we get sufficient evidence in favor of a hypothesis at a lower level (e.g., that Melfata told the truth), we can accept it. We can then use it as evidence in the inference at the next higher level, and so on.

Barbara: Instead of propagating uncertainty through a series of inferences, it sounds like you're suppressing it: acting "as if" an uncertain proposition is true. That's inviting trouble.

Art: It seems to me that we do that all the time. But isn't there any way to "accept" conclusions in the Bayesian or Shaferian systems?

Colette: Not really, Art. You end up with an assignment of probabilities (or degrees of support) to hypotheses, but not with a set of "acceptable" hypotheses.

Sy: Why can't you just decide to accept all the hypotheses whose probability, or support, exceeds a certain threshold?

Colette: Cohen argues that a reasonable criterion of acceptance cannot be formulated in terms of these measures. For example, let's say I set a threshold x so that if the Bayesian probability of a hypothesis is greater than x, I accept it. But now it's possible that I will accept proposition J and I will accept proposition K, and yet I won't accept the conjunction J&K. The reason, of course, is that the probability of the conjunction is the product of the probabilities of its component events, and so (unless one of the probabilities is unity) will be less than each of them. So a Bayesian who tried to formulate criteria for acceptance would end up not believing some very clear logical consequences of other things he believes. And this seems both irrational and contrary to what most people would actually do.

An even worse problem for both Bayesian and Shaferian theories is that they may provide support for conflicting hypotheses, e.g., for J and for not-J. So if people did believe the logical consequences of their beliefs, we would have support for a contradiction, J&not-J.

Barbara: Why bother with acceptance at all? If I know the probabilities I assign to every relevant hypothesis, I know all I need to know for a decision: I can maximize expected utilities. By "accepting" hypotheses, you're not only suppressing uncertainty; you're increasing the chance of inappropriate action.

Colette: In fact I think it is the Bayesian who is suppressing uncertainty, at least in the sense of incompleteness of evidence. But I see I have to clarify what I mean by acceptance; to me, it is primarily a device for representing degree of uncertainty, not sweeping it under the rug. The uncertainty in a hypothesis (even a probabilistic one) is, for me, the risk I would be taking if I accepted it: i.e., it is the amount of potential evidence out there that I haven't looked at yet. Now it's up to me to decide how much of this sort of risk I will take; in other words, how much of the relevant evidence I will require before I accept a hypothesis. As we noted earlier, Bayesian probabilities do not help me in the least in dealing with this kind of uncertainty. Moreover, I believe the Bayesian is wrong in supposing that this kind of uncertainty is irrelevant to action.

Art: I expect there's a positive side of acceptance, too, Colette. I suspect it will prove to be very difficult to communicate a complex Bayesian probabilistic analysis, in which every relevant hypothesis is assigned a probability.

Colette: That's right, Art. By "communicate" I assume you mean more than simply providing the inputs for a decision analysis in which we maximize expected utility. Action may be rather far down the road, in science and even in intelligence analysis. In the meantime, others must be able to understand (in some intuitive sense of that word) what the possible accounts of the situation are. Discussion in terms of acceptance enables us to present alternative full, cogent models for consideration. Each model will be, as noted above in the discussion of conjunction, a logically closed story: i.e., anything logically implied by the model is also part of the story. Such a model or models, together with some measure of the completeness of their evidential support, is--I submit--what the product of an intelligence analysis ought to be.

Phyllis: Shawn said that Cohen's theory fits a special case of Shafer's system, where evidence couldn't support conflicting hypotheses. Cohen is clearly relying on that feature in his concept of acceptance, where you don't want to have support for a contradiction. But I don't see how the possibility of conflicting conclusions has been ruled out.

Colette: In Cohen's system, if the evidence lends positive inductive support to conflicting hypotheses J and not-J, it follows that something is wrong with the evidence. Support for a contradiction should lead us to reevaluate the method by which we arrived at that support. Apparent contradictions prompt us to reconsider our understanding of the problem and to generate new hypotheses to explain them away. New variables may need to be added, or the ordering of the variables may need to be changed. Under the revised list of relevant variables, the evidence should become consistent. Cohen (1977) compares this process to a "scientific revolution," in which our basic presuppositions and methods shift in the face of recalcitrant data.

Shawn: Well, this looks like wishful thinking to me. Sometimes we simply won't be able to resolve the inconsistencies. Perhaps this should worry us, but we should still be able to represent all the relevant evidence, even when it appears to point in different directions. I'm afraid Cohen's theory puts us in a straightjacket.

<u>Colette</u>: The problem is, how can we use the evidence at all if it's inconsistent? We certainly couldn't use it to draw conclusions.

<u>Art</u>: I guess the problem is that a lot of the real work of inductive inference, i.e., establishing and revising the list of variables, is left outside of the theory proper. Cohen's theory doesn't support the assessment process as much as I had hoped.

<u>Phyllis</u>: In that respect, it's no different from the Bayesian theory with its reliance on the "art" of problem structuring, or Shafer's notion of constructing a new probability argument each time we receive new evidence. Perhaps we have here an inescapable feature of theories that purport to explain or guide inference.

# 4.0 THE BRIEFING

As a result of his conversations with Phyllis, Sy, Barbara, Zara, Shawn, and Colette, Art comes well prepared to the briefing. He has mustered all the relevant evidence he can find, but has been unable to discover any evidence which is conclusive. Nonetheless, with the assistance of his colleagues, he is prepared to answer questions about the strength and plausibility of his conclusions.

Art's strategy in the briefing is to begin with a more traditional, qualitative approach and to introduce concepts or analyses based on the various theories of inference when they become appropriate in response to questions or comments.

New participants in this dialogue include several unidentified members of the audience, and Cus, the customer for whom the briefing is intended.

## 4.1 Verbal Hedging Based Upon a Marshalling of Evidence

Art begins his briefing by stating the two hypotheses suggested by the customer's requirements:

$H_1$: Malbridgia is now building a prototype ZAP system,

$H_2$: Malbridgia is not building a prototype ZAP system.

Art acknowledges a prior expectation that Malbridgia is not yet ready to develop a prototype ZAP system, and its basis in a recent briefing Art heard by a nationally-recognized American scientist.

Art then summarizes the specific evidence he and his colleagues had obtained. His briefing chart looks like this:

| | $H_1$ | | $H_2$ |
|---|---|---|---|
| $E_2$ | (Summarized) | $E_1$ | (Summarized) |
| $E_3$ | * | | |
| $E_4$ | * | | |
| $E_5$ | * | | |

**Viewgraph 1.0**

In summarizing the evidence, Art points out that $E_1$ seems to favor $H_2$ but all the rest of the evidence seems to favor $H_1$ to varying degrees. In this discussion Art tells how he considered the credibility of the sources and how, in certain instances, such as $E_4$, he was forced to assign relatively small weight to the evidence because of doubts about the credibility of the source. In addition, Art discusses how each item of evidence is inconclusive or, to some degree, consistent with the truth of either hypothesis. Art concludes by saying that, in his opinion, the "preponderance of the evidence" seems to favor $H_1$ over $H_2$. His concluding statement is, "I believe it more likely that A is now building a Z system than that A is not." Following this statement, Art asks if there are questions; several hands are raised. The first person recognized is the person who requested the briefing.

<u>Cus</u>: Your conclusion is actually quite tepid and certainly not very specific, even though you did inform us about its evidentiary basis. Incidently, we thank you for telling us what your initial bias was. I am afraid I must ask you to be a bit more specific about how strongly you believe that the evidence favors $H_1$ over $H_2$. I asked for this briefing because I must recommend possible choices for our own weapon planners. In order to do this, I must have some sense of HOW MUCH MORE LIKELY is $H_1$ than $H_2$. All you have told me is that $H_1$ "is more likely" than $H_2$; this is not enough. If $H_1$ is just slightly more likely than $H_2$, I might recommend one thing; but if $H_1$ is, say, ten times more likely, I might recommend another.

I have a feeling also that by quantifying your reasoning, you will give us a better understanding of how you reached your conclusion. You have not really told us <u>why</u> you think $E_1$ is outweighed by the other evidence.

### 4.2 A Point Probability Analysis

Art: I am prepared to respond directly to your question. In fact, I can be very specific about how much more likely I believe $H_1$ is than $H_2$. I was not more specific in my opening statement because some persons have an aversion to the kind of specific numerical assessment that I will now provide for you. My next chart will show how I used a well-known probabilistic rule, called "Bayes' rule," for revising my opinion about the relative likeliness of $H_1$ and $H_2$, based on the evidence I was able to find. I believe this rule is appropriate to my task since it allows me to show you how my initial opinion about the likeliness of $H_1$ relative to that of $H_2$ changed as a result of incorporating each item of evidence I considered. I ask you to consider my second briefing chart: Line 1 of this viewgraph shows the "odds-likelihood ratio" form of Bayes' rule. It instructs us to take the product of the prior odds of $H_1$ to $H_2$ and the likelihood ratios for each item of evidence in order to determine the posterior odds of $H_1$ to $H_2$.

Prior odds expresses the extent of initial biases or expectations about the relative likeliness of the hypotheses before the evidence is evaluated. The likelihood ratios for each evidence item show which hypothesis each item favors inferentially and by how much. For any evidence item $E_i$, if the likelihood ratio $L_{Ei}$ is $> 1.0$, the item favors $H_1$. If $L_{Ei}$ is $< 1.0$, the item favors $H_2$. Thus, the posterior odds of $H_1$ to $H_2$ show the relative likeliness of $H_1$ and $H_2$, *after* the evidence has been incorporated. This is a simplified version of Bayes' rule. Proper use of this expression requires that consideration be made about whether or not the evidence items are independent, conditional upon $H_1$ or upon $H_2$.

**Posterior Odds:**

$$\Omega_1 = \frac{P(H_1|\text{Evidence})}{P(H_2|\text{Evidence})} = \Omega_0 \times L_{E_1} \times L_{E_2} \times L_{E_3} \times L_{E_4} \times L_{E_5} \quad (1)$$

where $\Omega_0 = P(H_1)/P(H_2)$ = Prior odds,

$L_{E_i} = \frac{P(E_i|H_1)}{P(E_i|H_2)}$ = Likelihood ratio for any item $E_i$

$$\Omega_1 = \frac{1}{5} \times [\frac{1}{2} \times \frac{3}{1} \times \frac{4}{1} \times \frac{1.5}{1} \times \frac{1.5}{1}] \quad (2)$$

$$= 2.7$$

$$P(H_1|\text{Evidence}) = \frac{\Omega_1}{1+\Omega_1} = \frac{2.7}{3.7} = 0.73 \quad (3)$$

$$P(H_2|\text{Evidence}) = 0.27$$

**Viewgraph 2.0**

In line 2 of this viewgraph are the specific values I used in my calculations. Initially, I thought $H_2$ was about 5 times more likely than $H_1$. This reflects the prior bias I told you I had. I evaluated the evidence items as follows: I thought $E_1$ more probable assuming $H_2$ than assuming $H_1$, in the ratio 1/2. The other four items I thought all favored $H_1$ over $H_2$ in the ratios shown in line 2. As you see, the result of my applying Bayes' rule leads me to say that it is about 2.7 times more likely that A is building the prototype than that A is not. I can go one step farther and determine the posterior probability of $H_1$ and $H_2$ in light of my evidence. Line 3 shows how you convert posterior odds to posterior probability. If you ask me how probable $H_1$ is, given my evidence, I will say that this probability is 0.73. Under the rules of probability in which Bayes' rule is appropriate, I must say that the posterior probability of $H_2$, given the evidence, is 0.27. I believe I have given you the specific answer you requested.

Hmmm. I see several hands in the air. My attempt to be specific has not pleased everyone!

Member of Audience: I must say that I don't have any idea where you got these numbers you call odds, likelihood ratios, etc. I thought probability was supposed to apply to well-defined sampling operations in which the probability of an event is estimated by the number of times the event occurred in the sample compared with the size of the sample. What is the sampling operation here? Have you collected data from a series of extremely similar cases? I doubt it! Your evidence seems to involve rather unique or one-of-a-kind events, such as $E_3$. In addition, your hypothesis $H_1$ is a unique event and not subject to a sampling operation of any kind I can think of. Your whole approach sounds very subjective to me and I must say that your use of mathematics seems like an attempt to make respectable a judgmental process that you could make come out any way you chose to.

Art: You are quite right about one thing: the probabilities, odds, and likelihood ratios I used did not arise as a result of any empirical sampling operation. Indeed, they are simply measures of the intensity of my beliefs about the various propositions at issue. In short, I readily admit that these assorted ingredients in line 2 of my viewgraph are subjective judgments. As far as your allegation that I used mathematics here to make this respectable, I have only this to say. I made admittedly subjective judgments about logically necessary components or parts of my inference task. Bayes' rule simply shows me how to combine these judgments in a logically consistent manner. As it turns out, if I tried to combine them or use them in a different way than I, in fact, did, my inconsistency could easily be exploited by anyone who knows how to construct a "Dutch book" against me. A "Dutch book" is a combination of wagers, based upon my incoherently expressed beliefs, which insures that I will lose regardless of whether $H_1$ or $H_2$ is actually true. I simply desire my beliefs and the process of revising these beliefs to be consistent or coherent. I have honestly expressed the extent of my uncertainties here, and I wish to combine these uncertainties in a rational way.

Member of Audience: I'm certain that you are not simply trying to impress us by putting a mathematical or scientific cloak on your subjective judgments. I appreciate the fact that you have attempted to combine your uncertainties in a consistent manner. I have another reason for being less than impressed with your analysis. You did, in fact, answer the first question by giving a precise estimate. My difficulty is that your analysis suggests you have far more precision in your estimations than any half-way serious person would believe you actually possess. For example, in your second viewgraph, on line 2, you tell us that $E_2$ is exactly 3 times more probable assuming $H_1$ than assuming $H_2$. How can you justify such precision; could this value not have been 5.0, 4.63, or 2.94? How do you know it is precisely 3.0? I must say that your analysis assumes an estimative precision which I, for one, do not believe you, or anyone else, possesses.

### 4.3 A Second-Order Probability Analysis

Art: The numbers I estimated and showed you in my last viewgraph are estimates of my prior uncertainty and of various other uncertainties associated with the evidence I have. Your essential question seems to be: how uncertain am I about my uncertainty? The general issue of estimative precision is one that I did give some thought to. My next viewgraph is designed to give you an estimate of what I will call my "second-order" uncertainty.

| | Upper 95% Certainty Bound: | Point Estimate: | Lower 95% Certainty Bound: |
|---|---|---|---|
| $\Omega_0$: | 1/3.2 | 1/5 | 1/7.9 |
| $L_{E_1}$: | 1/1.2 | 1/2 | 1/3.5 |
| $L_{E_2}$: | 5.2/1 | 3/1 | 1.7/1 |
| $L_{E_3}$: | 9.8/1 | 4/1 | 1.6/1 |
| $L_{E_4}$: | 2.0/1 | 1.5/1 | 1.1/1 |
| $L_{E_5}$: | 2.5/1 | 1.5/1 | 1/1.1 |
| $\Omega_1$ | 10.8/1 | 2.7/1 | 1/1.5 |
| $P(H_1 \mid \text{Evidence})$: | 0.92 | 0.73 | .47 |

**Viewgraph 3.0**

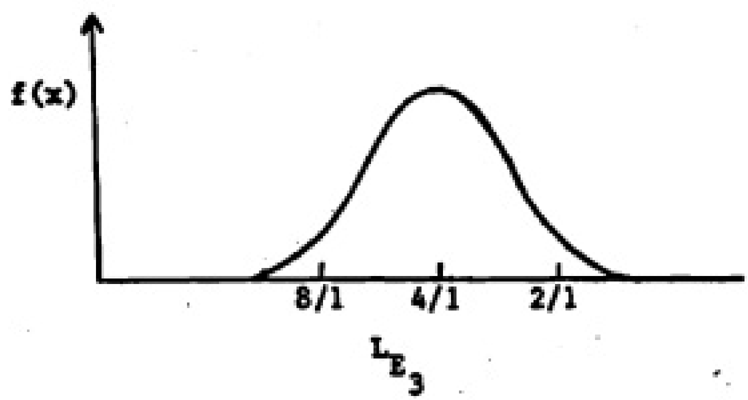
Here I show you intervals which I am 95% certain will contain my "true" prior odds and likelihood ratios. The largest values I call my "upper 95% certainty bounds," and the smallest values I call my "lower 95% certainty bounds." For example, someone asked about $E_2$; I would estimate a 95% chance that the likelihood ratio for this datum is no larger than 5.2/1 favoring $H_1$, and no lower than 1.7/1 favoring $H_1$. By the way, the middle column in this viewgraph contains the "point" estimates I gave you in my second viewgraph. The bottom two rows show the result of applying the probability calculus, together with a few assumptions, to the 95% certainty intervals. The next to last row gives a 95% interval for the posterior odds of $H_1$ to $H_2$. My upper 95% certainty bound corresponds to posterior odds of $H_1$ to $H_2$ of 10.8/1; my lower 95% certainty bound corresponds to posterior odds of $H_1$ to $H_2$ of 1/1.5 (favoring $H_2$).

Member of Audience: I gather that this interval aggregates together all the various uncertainties in your analysis?

Art: That's right. The imprecision in my assessment of prior odds and likelihoods "propagates" into the conclusion. This propagation can be computed by means of a formula in probability theory that expresses the variance of a function of random variables in terms of the variances of the random variables. The last row, by the way, shows what the corresponding posterior probabilities of $H_1$ are for my upper 95% and lower 95% bounds. This gives you some idea of how imprecise I believe my estimate is. Stated another way, this is a view of how uncertain I am about my uncertainty.

Member of Audience: I am very curious about why your "point" estimates when aggregated do not produce a posterior calculation which falls at the midpoint of the range of posterior odds; 2.7 is not halfway between 10.8 and 1/1.5.

Art: Take any one of my estimate intervals, say the one for $E_3$; here my upper estimate is 9.8 and my lower estimate is 1.6, the "point" estimate I gave you was 4. I do not wish you to assume that my uncertainty is spread uniformly across this interval. If it were, I would have reported the midpoint between 9.8 and 1.6, which is 5.7. For each estimate I actually determined a distribution which shows how my uncertainty is spread across the interval between my lower and upper estimates. To simplify this task, I assumed that these distributions were normal on a logarithmic scale. Thus, for example, my uncertainty about $L_E$ follows a log-normal distribution, so that on a logarithmic scale, the graph of my uncertainty in $L_E$ would look like this:




Member of Audience: So this is one of the "assumptions" you said you needed in arriving at your result?

Art: That's right. It seems appropriate to use log-normal distributions when my uncertainty pertains to ratios. For example, it implies that I am equally confident that the true value of $L_E$ falls in the interval between 4/1 and 2/1, as I am that it falls in the interval between 8/1 and 4/1. In each case, I am off by a factor of 2. So, in terms of ratios, or logarithms, these second-order distributions are symmetrical. But on a linear scale, they are asymmetrical or "skewed."

Member of Audience: Now hold on. This doesn't seem right at all. You have been trying to quantify your belief concerning $H_1$ and $H_2$, based on the available evidence. The problem that led to these second-order probabilities was that you were implying too much precision in your assessments. But now, you say that you are 95% certain that your belief lies in the stated range. This sort of "second-order belief" seems to require even more precision than the first-order assessments. If I could say things like I'm 95% certain about ranges of my own belief, I think I could be more precise about my belief. But I can't. What you meant by giving ranges, I think, was that you couldn't interpret the evidence well enough to do better than say your belief was within a particular interval. You certainly couldn't say your uncertainty about your uncertainty was log-normally distributed! You couldn't even say that about your uncertainty. I think you're assuming too much in this exercise about the structure of your beliefs.

Member of Audience: I guess if Art were unsure about the log-normality of his second-order probabilities, he could just introduce third-order probabilities! That does seem like madness!

### 4.4 A Fuzzy Probability Analysis

Art: Well, let me confess that I think there is some justice in those remarks. In fact, I anticipated them. Although I think it can be illuminating and even necessary in building a model to make assumptions like log-normality, it can also be useful to see how far you can get without them. We don't need to stick with the idea that everything we don't know for certain should be modeled with probability theory. Some ideas by Zadeh may help us out here. Instead of assessing probabilities for our first-order probabilities, we can "fuzzify" them, by simply stating what values are possible. I happen to have a viewgraph where I do just that.

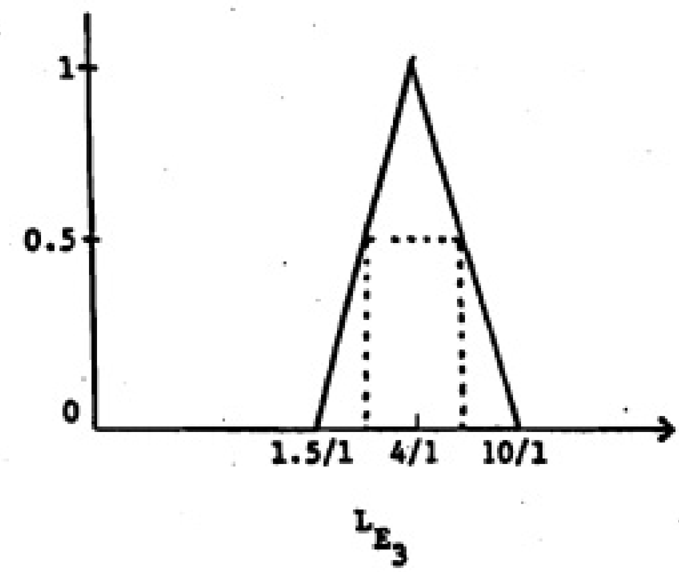
| | Upper Estimate: | Point Estimate: | Lower Estimate: |
|---|---|---|---|
| $\Omega_0$: | 1/3 | 1/5 | 1/9 |
| $L_{E_1}$: | 4/5 | 1/2 | 1/4 |
| $L_{E_2}$: | 5/1 | 3/1 | 1.5/1 |
| $L_{E_3}$: | 10/1 | 4/1 | 1.5/1 |
| $L_{E_4}$: | 2.5/1 | 1.5/1 | 1.25/1 |
| $L_{E_5}$: | 3/1 | 1.5/1 | 1.1/1 |
| $\Omega_1$ | 100/1 | 2.7/1 | 1/11.6 |
| $P(H_1 \mid \text{Evidence})$: | 0.99 | 0.73 | 0.08 |

Viewgraph 4.0

What this means is that $L_E$ , for example, is "about 1/2," but is possibly as high as 4/5 and possibly as low as 1/4. The last two lines give the range of possible values for the posterior odds and posterior probabilities. The upper estimate for the posterior odds, 100/1, is just the result of applying Bayes' rule to the upper estimates for the prior odds and likelihoods. This makes sense, since if the largest possible values for all the priors and likelihoods happened to be true, the value of the posterior odds would be 100/1; so that is the largest possible value of the posterior odds. The same is true, of course, for the lowest possible value of the posterior odds; to compute it, we just multiply the lowest possible values of the priors and likelihoods.

Member of Audience: Well, that's certainly a simpler computation than the second-order probability analysis. But am I right in saying that you actually think the probability of $H_1$, given the evidence, is somewhere between 0.99 and 0.08? If so, I don't see what you have told us. I could have made such a judgment without ever considering any evidence at all.

Art: Oh, we can do a lot more than that. For Zadeh, possibility comes in degrees; so far, we have talked as though it were all or none. For each of the priors and likelihoods, I have assessed a possibility distribution. For example, the distribution for $L_E$ shows that values near 4/1 are more possible than those farther away; values less than 1.5/1 or greater than 10/1 are not possible at all. These graphs look a little like the probability distribution I showed you:




Now we can give possibility intervals for $L_E$ just as we gave probability intervals. These intervals are called "level sets," since they contain all values whose possibility exceeds some chosen level. For example, we already know that any value whose possibility is greater than zero falls between 100/1 and 1/11.6. Suppose we want to know what interval includes all values whose possibility is at least .5. To get the upper bound on this interval, we first take the upper values of the priors and likelihoods that have exactly .5 possibility; for example, from the dotted lines in the chart we see that the possibility that $L_E$ is 6/1 is .5. We then multiply these upper values together. Similarly, for the lower 50% possibility bound, we take the lower values that have .5 possibility and multiply them.

Member of Audience: I'm sure that this would produce narrower intervals than we got in viewgraph 4.0. But I guess you have to make some assumptions, about the shape of the possibility distribution, just as you did in the probability analysis, to get them.

Art: That's right. Though some would argue that assessments of possibility are easier; and certainly the computation is simpler.

Member of Audience: I think there may be something more fundamentally wrong with both second-order probabilities and fuzzy probabilities. Your final assessment actually shows more uncertainty than you began with. That is, in the fuzzy probability analysis the range of possible posterior odds is much larger than the range of possible prior odds. The same is true in the Bayesian second-order analysis, where the 95% certainty interval for the posterior odds is larger than for the prior odds. Now, we try to find relevant evidence in order to reduce our uncertainty, not to increase it. You have shown us your analysis based on five pieces of evidence. If you are saying that your evidence has value to us, it seems pretty clear that neither second-order probabilities nor possibilities have correctly captured that value.

Gus: I have a concern that may be related. I am going to assume that your point estimates, which are what I actually asked for, simply represent your best guesses. Your conclusion is that the probability that Malbridgia is building a ZAP system, based on your evidence, is about 0.73. This seems to be a fairly strong probability based, as it is, on just five items of evidence. But you have said nothing about any other evidence that may bear upon this problem which you have not been able to obtain or consider. How does this enter into your assessment?

### 4.5 A Belief Function Analysis

Art: These are important and valid points. They help us realize that what we have been assessing with our various intervals is something akin to "measurement error" or imprecision in my assessment of each prior or likelihood. Imprecision is larger in the conclusion than in the premises simply because each premise adds in some new imprecision of its own to the conclusion. But we have not yet focused on the completeness or weight of our evidence, taken as a whole. In that case, presumably, the more bits of evidence, the more certain we are.

You will not be surprised to hear that I have anticipated this problem as well. Here I think some ideas of Glenn Shafer can help us. The next viewgraph shows how a Shaferian analysis might look. For each item of evidence I have assessed what Shafer calls a "simple support function" - i.e., a function that assigns support to only a single elementary hypothesis, $H_1$ or $H_2$, but not to each. This captures our feeling that each bit of evidence adds credence to one or the other of our hypotheses, but not to both simultaneously. Of course, these simple support functions also assign some support to $\{H_1,H_2\}$, reflecting any doubts we might have about the validity of the evidence.

| | $m(H_1)$ | $m(H_2)$ | $m(H_1,H_2)$ |
|---|---|---|---|
| $E_0$ | 0 | .7 | .3 |
| $E_1$ | 0 | .5 | .5 |
| $E_2$ | .6 | 0 | .4 |
| $E_3$ | .8 | 0 | .2 |
| $E_4$ | .2 | 0 | .8 |
| $E_5$ | .3 | 0 | .7 |
| $E_0$-$E_5$ | .76 | .20 | .04 |

$Bel(H_1)$: .76 $Pl(H_1)$: 1-.20 = .80

$Bel(H_2)$: .20 $Pl(H_2)$: 1-.76 = .24.

**Viewgraph 5.0**

The last few lines show the aggregate support function and belief function obtained when we combine these bits of evidence by Dempster's rule. The easiest way to do this is in two steps. First, I combined separately the concurring evidence $\{E_0,E_1\}$ in support of $H_2$, and the concurring evidence $\{E_2,E_3,E_4,E_5\}$ in support of $H_1$. This gave me two <u>conflicting</u> support functions--with support of .955 for $H_1$ based on $\{E_2,E_3,E_4,E_5\}$ and support of .85 for $H_2$ based on $\{E_0,E_1\}$. So the second step was to combine these two functions, normalizing to eliminate the impossible situation in which both functions were valid.

The important thing to notice is that as we add evidence, the range of permissable belief narrows. The uncommitted support at the end of the analysis is .04, which is far less than the uncommitted support based on any of the individual items of evidence. Thus, the range of belief in $H_1$ consistent with all this evidence is between .76 and .80, and the range for belief in $H_2$ is between .2 and .24. These bounds are much narrower than the ones we got from second-order Bayesian probabilities or fuzzy probabilities. The conclusion, I think, is that these different models have touched on different concepts. Here we have a measure of the completeness or weight of our total collection of evidence, rather than of the "measurement error" in our assessment of a probability.

Member of Audience: Perhaps then each of these analytical approaches has some role to play in increasing our understanding.

Art: Perhaps so.

## REFERENCES

Bellman, R.E., and Giertz, M. On the analytic formalism of the theory of fuzzy sets. Inform. Sci., Vol. 5, 1973, 149-156.

Brown, R.V., and Lindley, D.V. Improving judgment by reconciling incoherence. Theory and Decision, 1982, 14, 113-132.

Cohen, L.J. The probable and the provable. Oxford, England: Clarendon Press, 1977.

Cohen, L.J. Application conditions for eliminative induction. In Cohen, L.J., and Hesse, M. (Eds.), Applications of inductive logic. Oxford, England: Clarendon Press, 1980.

de Finetti, B. Foresight: Its logical laws, its subjective sources. English translation in H.E. Kyburg, Jr., and H.E. Smokler (Eds.), Studies in subjective probability. New York: Wiley, 1964. (Original: 1937).

Dubois, D., and Prade, H. Evidence measures based on fuzzy information, undated manuscript.

Freeling, A.N.S. Fuzzy sets and decision analysis. IEEE Transactions on Systems, Man, and Cybernetics, 1980, SMC-10(7).

Freeling, A.N.S. Possibilities versus fuzzy probabilities--Two alternative decision aids. In H.-J. Zimmerman, and L. A. Zadeh (Eds.), Decision analysis through fuzzy sets. TIMS/ORSA Studies in Management Science, 1983.

Fung, L.W., and Fu, K.S. An axiomatic approach to rational decision-making in a fuzzy environment. In L. Zadeh et al., (Eds.), Fuzzy sets and their application to cognitive and decision processes. New York: Academic, 1975.

Horwich, P. Probability and evidence. Cambridge, England: Cambridge University Press, 1982.

Goldsmith, R.W. Evaluating evidence in criminal cases by means of the evidentiary value model. In Gardenfors, P., Hansson, B., and Sahlin, N.-E. (Eds.), Evidentiary value: Philosophical, judicial and psychological aspects of a theory. Lund, Sweden: CWK Gleerup, 1983.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.). Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, 1982.

Keeney, R.L., and Raiffa, H. Decisions with multiple objectives: Preferences and value tradeoffs. New York: Wiley, 1976.

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.

Ramsey, F.P. Truth and probability. In H.E. Kyburg and H.E. Smokler (Eds.), Studies in subjective probability. New York: Wiley, 1964. (Original: 1926).

Savage, L.J. The foundations of statistics. New York: Wiley, 1954.

Schum, D.A. Current developments in research on cascaded inference processes. In Wallsten, T.S. (Ed.), _Cognitive processes in choice and decision behavior_. Hillsdale, NJ: Lawrence Erlbaum Assoc., 1980, 179-210.

Schum, D.A. Sorting out the effects of witness sensitivity and response criterion placement upon the inferential value of testimonial evidence. _Organizational Behavior and Human Performance_, 1981, _2_.

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. Constructive probability. _Synthese_, 1981, _48_, 1-60.

Shafer, G. Lindley's paradox. _Journal of the American Statistical Association_, June 1983, _77_(378), 325-334.

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

Shimony, A. Scientific inference. In R.G. Colodny (Ed.), _The nature & function of scientific theories_. University of Pittsburgh, 1970.

Smith, C.A.B. Consistency in statistical inference and decision (with discussion). _Journal of the Royal Statistical Society, Series B_, 1961, _23_, 1-25.

Spiegelhalter, D. J. _A statistical view of uncertainty in expert systems_. MRC Biostatistics Unit, MRC Centre, Hills Road Cambridge, 1985.

Tani, S.N. _Modeling and decision analysis_ (EES-DA-75-3). Stanford, CA: Stanford University, Department of Engineering-Economic Systems, June 1975.

Watson, S.R., Brown, R.V., and Lindley, D.V. _Three papers on the valuation of decision analysis_ (Technical Report 77-2). McLean, VA: Decisions and Designs, Inc., May 1977.

Watson, S.R., Weiss, J.J., and Donnell, M.L. Fuzzy decision analysis. _IEEE Transactions on Systems, Man, and Cybernetics_, 1979, _SMC-9_, 1-9.

Zadeh, L.A. Fuzzy sets. _Information and Control_, 1965, _8_, 338-353.

Zadeh, L.A. Fuzzy sets as a basis for a theory of possibility. _Fuzzy Sets and Systems_, 1978, _1_, 3-28.

Zadeh, L.A. Fuzzy probabilities and their role in decision analysis. _IFAC Theory and Application of Digital Control_, 1982, 15-21.

Zadeh, L.A. The role of fuzzy logic in the management of uncertainty in expert systems. _Fuzzy Sets and Systems_, 1983(a), _11_, 199-227.

Zadeh, L.A. A computational approach to fuzzy quantifiers in natural languages. _Comp & Maths. with Appls._, 1983(b), _9_(1), 149-184.