adopt other approaches as heuristics or as richer representations of the issues involved. It seems that Spiegelhalter's approach has been similar.

Secondly, one could validate an expert system by its comparison with expert performance. One can ask whether the diagnosis achieved by Spiegelhalter's system was better or worse than that achieved by competent diagnosticians. There is of course a debate over whether an expert system should be appraised in this way. Is the goal to reproduce the abilities of an expert, or to improve on the abilities of available human judges? If it is the former, then indeed it is sensible to compare performance with experts, but in this case one wonders why one should not use the experts themselves. This could be answered by observing that very often experts are in short supply. If, on the other hand, our goal is to improve on human inference behavior, then the criterion of conformity with some expert performance is not appropriate. A final measure of the appropriatness of an expert system is user satisfaction. To what extent do the people who interact with the expert system feel that the system is of use to them? In Spiegelhalter's case there are two kinds of people involved, namely the patients and the doctors. As Spiegelhalter observes, it is very important that the doctors are supportive of the endeavor and that they do not feel that their professional competence is in any way being threatened. It is perhaps more important, however, that the patients feel that they are being properly attended to. Spiegelhalter seems to have achieved success on both fronts.

4. SUMMARY

Although the purpose of the conference was to discuss the use of the different theories for the representation of uncertainty in expert systems, the principal speakers, perhaps wisely, devoted their discussion mainly to arguing the cases for the use of their different theories in general. On the basis of the discussions we had at this conference, it seems to me that one can summarize as follows. Probability theory has a strong intellectual support and in principle there is no reason why one should not be satisfied with this theory. Its use does, however, lead to enormous problems of complexity, and as a matter of practice it is necessary to seek for approximations. Fuzzy set theory can be viewed as a heuristic for handling those situations where imprecise inputs and imprecise inferences are required without the need to resort to the greater complexity of probability theory. Belief function theory can be thought of as a way of representing inferences from evidence within the probabilistic framework.

There are yet other alternative approaches to handling uncertain inferences which were not mentioned at the conference, and notable among these is the nonmonotonic logic of Doyle. Recently Cohen (Cohen, Watson and Barrett, 1985) has suggested a combination of Doyle's theory with both Shafer's and Zadeh's which he has referred to as the nonmonotonic probabilist. This seems an exciting possibility for approaching the problem at the heart of this conference.

ADDITIONAL REFERENCES

COHEN, M. S., WATSON, S. R. and BARRETT, E. (1985). Alternative theories of inference in expert systems for image analysis. Technical Report 85-1, Decision Science Consortium, Falls Church, Va.

SCHUM, D. A. (1981). Sorting out the effects of witness sensitivity and response-criterion placement upon the inferential value of testimonial evidence. *Organizational Behavior and Human Performance* **27** 153–196.

Comment

A. P. Dempster and Augustine Kong

The papers by Shafer and Spiegelhalter are valuable summaries by acknowledged leaders in active research fields. There is much food for thought in both papers, and many of the techniques and issues raised by these authors will gradually become better understood as the field of uncertainty assessment in expert systems advances. Our research on models and techniques for belief function analysis (Kong, 1986; Dempster and Kong, 1986) is complementary to that of Shafer and Spiegelhalter. We all seek to provide tools for real applications, based on carefully constructed analyses expressed through mathematically well-articulated principles of uncertain reasoning.

Lindley is on a different track. He rehearses familiar normative arguments for the Bayesian paradigm, evidently seeking to persuade less committed colleagues to abandon their fallacious ways. Unfortunately, he shows no interest in understanding how his

A. P. Dempster is a Professor and Augustine Kong is an Instructor at the Department of Statistics, Harvard University, Cambridge, Massachusetts 02138.

competitors really think, and hence, does not address the issues on which, in our opinion, credible contemporary debate should focus. As illustrated by Shafer's hypothetical "ploxoma" example, his "challenge" is unconvincing, for he casts himself first in the role of challenger, then of umpire, and finally reverts to challenger, proclaiming himself well satisfied with the result.

Lindley oversimplifies by identifying Bayes with the use of probability. In fact, numerical probabilities which are both syntactically and semantically very close to Lindley's probabilities are essential to three alternative approaches ("classical statistics," "upper and lower probabilities," and "belief functions") which he criticizes. But surely it is first necessary to understand the various styles of reasoning with probability implied by each of these systems, before either choosing among them or judging which circumstances are appropriate for each. In our view, moreover, the belief function system is very close to Bayes, and indeed includes Bayesian models as special cases, so it is not easily rejected in favor of Bayes except by arguments whose artificiality is painfully obvious from the belief function standpoint.

We make no attempt here to defend classical statistics, upper and lower probability systems, or fuzzy logic, where the last seems fundamentally different from the others, but we can accept that each may have an appropriate place in valid and useful formal analyses. Instead, we comment briefly on the flexibility which belief function theory adds to Bayesian theory in its ability to incorporate evidence. Then we discuss at greater length the connection between belief functions and decision theory.

Lindley repeats verbatim his discussion of the Shafer (1982) "ploxoma" example, as he says, to provoke further discussion. As matters stand, Shafer has not modified his original representation, which contains three belief function components: (a) a range .05 to .15 to describe the prior probability of "virulent ploxoma," (b) a vacuous belief function to describe the inability of the "ordinary ploxoma" experiment to distinguish between x_2 and x_3, and (c) a 25% discounting applied to the experimental data about ordinary ploxoma. Evidently, these features were introduced to illustrate the flexible forms of uncertainty representation encompassed by the belief function paradigm. Lindley's response is to suggest converting Shafer's analysis to Bayesian form by altering (a) to a single prior probability .1, (b) to the indifference prior assigning .5 to each of x_2 and x_3 given that one or the other has occurred, and (c) to $E(\gamma_i \mid \beta_i) = \beta_i$ where $(\gamma_1, \gamma_2, \gamma_3)$ refers to valid chances of (x_1, x_2, x_3) given ordinary ploxoma for the new patient George, although $(\beta_1, \beta_2, \beta_3)$ refers to the questionable experimental results.

So far, neither side has explicitly addressed their differences over (a) or (b). But Lindley does now respond to the Shafer (1982) rebuttal of his altered (c), by allowing that we might have "no confidence at all" in the ordinary ploxoma study, in which case "$E(\gamma \mid \beta)$ would not depend on β." That is, 100% discounting means that the Bayesian constructs $E(\gamma_i)$ from wherever Bayesians construct such priors, thus implicitly introducing other sources of information processed together in the Bayesian's head to produce the prior. Presumably, if the Bayesian were, more realistically, to adopt less drastic discounting, the same prior about $(\gamma_1, \gamma_2, \gamma_3)$ would still be assessed and combined with information from the data via another assessed prior $f(\gamma \mid \beta)$. Thus, Bayesian analysis is not at all simple in execution, if one takes it seriously.

Belief function methodology does introduce more complexity into the class of available representations of uncertainty, although not typically into the task of assessing specific representations. Lindley criticizes the mathematical generalization as lacking "necessity" in the sense of William of Ockham. The important question is whether the added flexibility is necessary in practice to permit satisfactory representation of an analyst's state of uncertainty about the real world. We believe that it is literally impossible to answer the question outside the context of real examples based on attempts to construct formal representations of uncertainty reflecting actual uncertain knowledge of the real world. Because their example is purely hypothetical, neither Shafer nor Lindley is able to discuss in any specific way the construction of their specific models.

Important points about Lindley's "challenge" are, first, that a meaningful test must deal with real examples, and, second, that the umpire must rely on some assessment of help given to third party clients. Bayesian decision analysis has been out in the field for about 30 years, and in our (subjective) assessment has achieved only limited penetration into what might seem to be its natural markets. If this failure to penetrate were simply due to ignorance of the techniques on the part of practitioners, then Lindley's proselytizing might have a point. A more plausible explanation is that practical construction of realistic Bayesian models is typically very difficult. Because belief function analysis does not pretend to the vague and difficult goal of integrating all evidence available to the analyst, but instead attempts only to represent explicit and limited packets of independent evidence, the process of constructing belief function models is inherently simpler, and on this score belief function methods have excellent prospects for success in practice. Of course, computational difficulties are another matter, and here the tradeoffs are less clear, because computationally neither approach has advanced beyond its infancy. All in all, there would seem to be sound practical reasons for seeking to relax Bayesian constraints to obtain more flexible, explicit, and realistic representations of uncertainty.

Since the 1950s, both Bayesian and frequentist decision theorists have agreed that Bayesian expected loss is the appropriate numerical criterion for comparing possible acts. Such a Bayesian analysis introduces a precise ordering among acts, in the sense that the set of all acts is partitioned into subsets where the analyst is indifferent within a subset but has a complete preference order between subsets. Since general belief functions replace expected loss by upper and lower expected losses, moving to a belief function framework implies a tradeoff. On one hand the analyst simplifies inputs to specific sources of evidence, while paying on the other hand by having only partial orderings among acts. The partial orderings arise because the single numerical Bayesian expectation is replaced by numerical upper and lower expectations.

Lindley evidently does not wish to allow partial ordering, but do the standard normative arguments which he presents really prohibit it? For example, the scoring rule argument posits an artificial decision problem, where the acts are possible choices of a numerical measure of uncertainty about some unknown binary state, and the losses are heuristic quality assessments of the numerical measure of uncertainty. The conclusion for this decision problem, as for decision problems generally, is that Bayesian decision rules are the only admissible decision rules. More precisely, the conclusion is that, if there is a rule that selects a single act from the available set of acts, the rule must minimize expected loss under some probability distribution. Note, however, that the condition "if there is a rule" tacitly prejudges the question at issue. For if we choose to report only a partial ordering, we are, in effect, opting to specify no rule, so the admissibility result becomes irrelevant.

Lindley abuses the theory of belief functions by substituting belief into a scoring measure as though it were a simple probability. It may therefore be helpful to sketch what we see as the right way to think about belief functions in a decision-theoretic framework. To illustrate, consider a decision problem with decision space $D = \{d_1, d_2, d_3\}$, outcome space $W = \{w_1, w_2\}$ and loss function given in Table 1. Let $D^* = \{d(p_1, p_2, p_3) \mid p_1, p_2, p_3 \geq 0, \sum_{i=1}^{3} p_i = 1\}$, where

TABLE 1
Loss function

	d_1	d_2	d_3
w_1	0	10	20
w_2	45	20	10

$d(p_1, p_2, p_3)$ denotes the randomized decision that selects decision d_i, $i = 1, 2, 3$, with probability p_i. Following DeGroot (1970),

$$L(w_1, d(p_1, p_2, p_3)) = 10p_2 + 20p_3 \tag{1.1}$$

and

$$L(w_2, d(p_1, p_2, p_3)) = 45p_1 + 20p_2 + 10p_3$$

where $L(w, d)$ denotes loss. From Figure 1 it is clear that a decision $d(p_1, p_2, p_3)$ is admissible in the ordinary decision theory sense if either $p_1 = 0$ or $p_3 = 0$, thus including the pure decisions d_1, d_2, and d_3. The minimax decision is easily computed to be $d(0, \frac{1}{2}, \frac{1}{2})$ where

$$L(w_1, d(0, \tfrac{1}{2}, \tfrac{1}{2})) = L(w_2, d(0, \tfrac{1}{2}, \tfrac{1}{2})) = 15.$$

Suppose our knowledge about the outcome is represented by the belief function Bel over W. Let $\{\mu\}_{\text{Bel}}$ be the collection of probability measures μ over W that satisfy

$$\text{Bel}(A) \leq \mu(A) \leq Pl(A)$$

for all $A \subset W$. For $d, d' \in D^*$, we say d is *uniformly dominated* by d' with respect to Bel if

$$E\{L(w, d) \mid \mu\} \geq E\{L(w, d') \mid \mu\}$$

for all $\mu \in \{\mu\}_{\text{Bel}}$ and there exists $\mu^* \in \{\mu\}_{\text{Bel}}$ such that

$$E\{L(w, d) \mid \mu^*\} > E\{L(w, d') \mid \mu^*\},$$

where $E\{\cdot \mid \mu\}$ denotes expectation computed based on μ. We call a decision *permissible* against Bel if it is not uniformly dominated by another decision in D^* with respect to Bel. Hence, a decision is admissible if it is permissible against the vacuous belief function.

Suppose Bel is

$$\begin{aligned} m(\{w_1\}) &= .6, \\ m(\{w_2\}) &= .2, \\ m(\{w_1, w_2\}) &= .2. \end{aligned} \tag{1.2}$$

It follows that $\{\mu\}_{\text{Bel}} = \{\mu_t \mid .6 \leq t \leq .8\}$, where μ_t denotes the probability measure that assigns probability t to w_1 and probability $1 - t$ to w_2. For the pure decisions,

$$\begin{aligned} E\{L(w, d_1) \mid \mu_t\} &= 45(1 - t), \\ E\{L(w, d_2) \mid \mu_t\} &= 10t + 20(1 - t) = 20 - 10t, \\ E\{L(w, d_3) \mid \mu_t\} &= 20t + 10(1 - t) = 10 - 20t. \end{aligned}$$

Figure 2 plots $E\{L(w, d_1) \mid \mu_t\}$, $E\{L(w, d_2) \mid \mu_t\}$, and $E\{L(w, d_3) \mid \mu_t\}$ for $.6 \leq t \leq .8$. It shows that d_3 is uniformly dominated by d_2 with respect to Bel. Since $E\{L(w, d(p_1, p_2, p_3)) \mid \mu\} = \sum_{i=1}^{3} p_i E\{L(w, d_i) \mid \mu\}$, it is straightforward to prove that $d(p_1, p_2, p_3)$, $p_3 > 0$, is uniformly dominated by $d(p_1, p_2 + p_3, 0)$ with respect

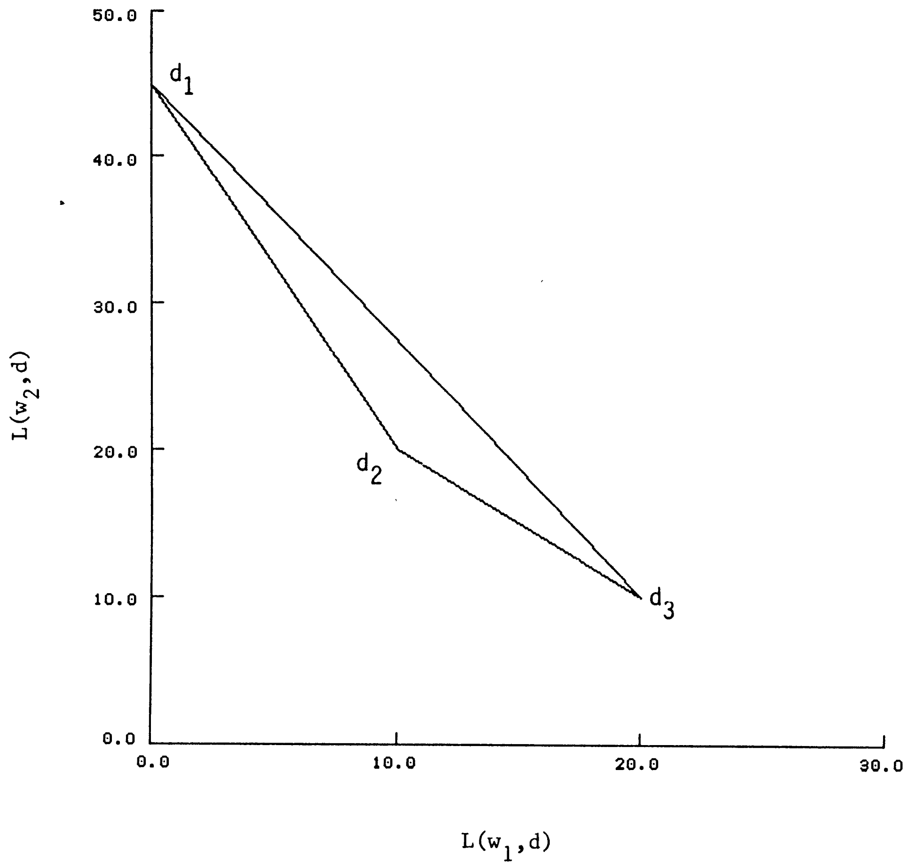

FIG. 1. $L(w_1, d)$ *versus* $L(w_2, d)$.

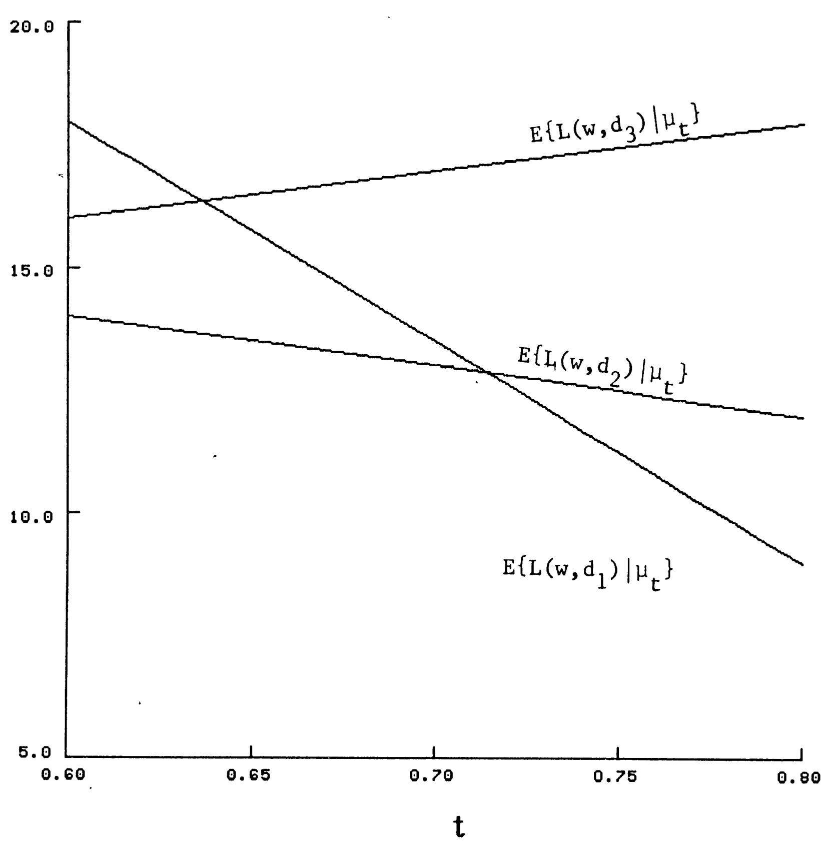

FIG. 2. *Expected loss.*

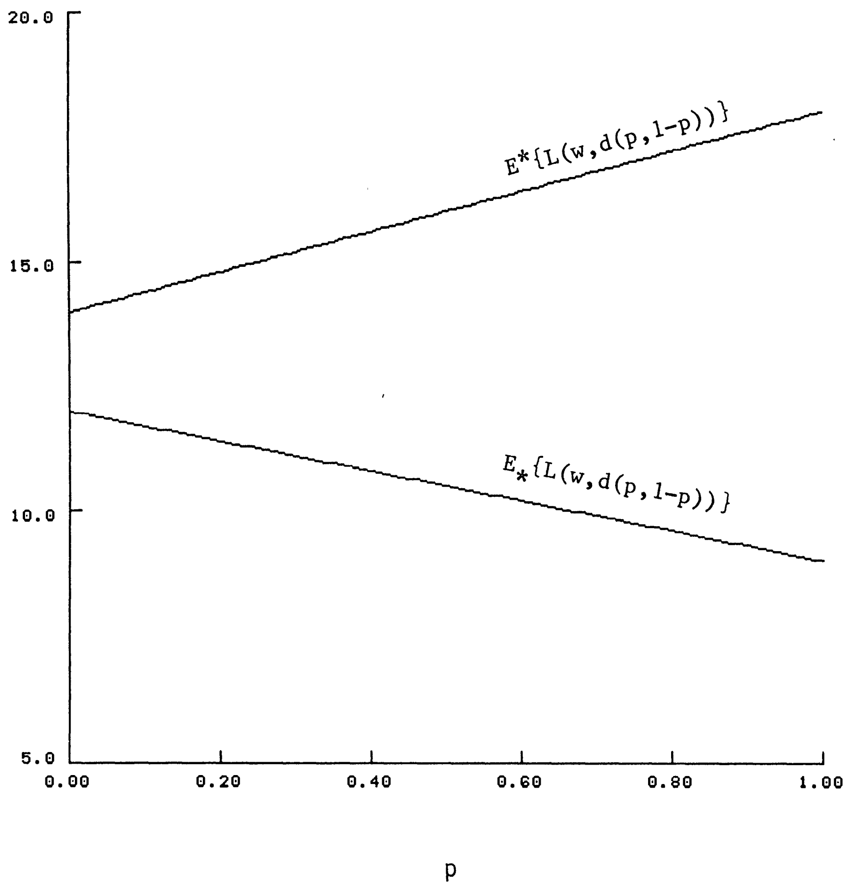

FIG. 3. *Upper and lower expected loss.*

to Bel. The permissible decisions against Bel are $d(p, 1 - p, 0)$, $0 \le p \le 1$.

For a decision $d \in D^*$ define lower and upper expected loss with respect to Bel to be (cf. Dempster, 1967)

$$E_*\{L(w, d)\} = \inf_{\mu \in \{\mu\}_{\text{Bel}}} [E\{L(w, d) \mid \mu\}]$$

and

$$E^*\{L(w, d)\} = \sup_{\mu \in \{\mu\}_{\text{Bel}}} [E\{L(w, d) \mid \mu\}].$$

Figure 3 plots the upper and lower expected loss of the permissible decisions $d(p, 1 - p, 0)$ as functions of p. Since $d(0, 1, 0) = d_2$ has the minimum upper expected loss, we call d_2 the miniupper decision against Bel.

Notice that the miniupper decision against a vacuous belief function is the minimax decision and the miniupper decision against a Bayesian belief function is the corresponding Bayes decision. Hence, the miniupper method is a generalization of minimax and Bayes. Under more general settings where there can be more than two outcomes, it can be shown that the task of finding the miniupper decision can be reformulated as a linear programming problem. Details will be given in a coming technical report.

We are not necessarily endorsing the miniupper decision here. Indeed, in the above example, we have no reason to fault someone who chooses d_1 over d_2. The point is that some guidance toward rational decisions can be made even if uncertainty is represented by belief functions instead of distribution functions.

ACKNOWLEDGMENTS

This work was supported in part by Office of Naval Research Contract 00014-85-K-0496 and Army Research Office Grant DAAL03-86-K-0042.

ADDITIONAL REFERENCES

DEMPSTER, A. P. (1967). Upper and lower probabilities induced by a multivariate mapping. *Ann. Math. Statist.* **38** 325–339.

DEMPSTER, A. P. and KONG, A. (1986). Uncertain evidence and artificial analysis. Research Report S-108, Dept. Statistics, Harvard Univ.

KONG, A. (1986). Multivariate belief functions and graphical models. Ph.D. dissertation, Dept. Statistics, Harvard Univ.