Professor Shafer's historical perspective puts the current discussion in an appropriate context, and emphasizes that many of the issues raised in expert system research are by no means novel. The interest in belief function methodology is understandable, as it appears to provide a means of avoiding full subjective assessment of a joint probability distribution, and—by formulating "uncertainty" in terms of reliability of evidence—it seems to attach uncertainty directly to the *rule* rather than the consequences of the rule. All this is very attractive, but users of the methodology also have to take on board a rule of combination that can lead to somewhat unintuitive results (Zadeh, 1986), problems in providing an operational interpretation of the numerical inputs and outputs, and a considerable computational burden.

Shafer does show how computationally efficient schemes are available on simple trees, but this is an extremely restrictive class of model, excluding both multiple causes of the same event, and an element being a member of two taxonomic hierarchies (for example, "gallstones" may also be part of a "dyspepsia" taxonomy). In contrast, efficient probabilistic schemes are now being devised for general graphical structures.

This still leaves the ability of belief functions to deal with "unknown" or "unknowable" probabilities. From a historical point of view, it would be easy to slip into the "likelihood versus Bayesian" debate at this point. But I believe the objective of constructing expert systems enables us to avoid such arguments. In such technological applications, there is real understanding of the problem to be exploited, and from a purely pragmatic point of view, unknown probabilities just do not occur—an assessment can always be obtained by careful questioning. Of course, the subject may not feel too confident in his assessment, and will not be able to list a set of independent sources of evidence for his opinion. But the opinion is there and can be used, although, as Professor Lindley emphasizes, in certain circumstances the imprecision may be relevant. As Professor Shafer points out, explanation of a system's conclusions may be provided at many levels, and probability judgments that have not been "constructed" on specified evidence can, if necessary, be identified. Provided a system's predictive performance is being monitored by scoring rules, it seems quite reasonable in a medical area to exploit "informed guesses" rather than rely on a legalistic paradigm that models unreliable "witnesses."

ADDITIONAL REFERENCE

ZADEH, L. (1986). A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combinations. *Artificial Intelligence Mag.* 85–90.

Rejoinder

Glenn Shafer

Watson and Dempster and Kong underline the point that belief functions are a form of probability. I can only say that I agree wholeheartedly.

I still have some bones to pick, on the other hand, with Spiegelhalter and Lindley.

Spiegelhalter's comments on the computational situation are misleading. He suggests that computationally efficient schemes for belief functions are available only for a very restrictive class of models, whereas efficient Bayesian schemes "are now being devised" for very general models. In fact, most Bayesian computational schemes have belief-function generalizations. It is true that the Bayesian special cases usually require less computation; Bayesian models require more complicated inputs than belief-function models, and there is less need for computation when you begin with more information. But the trade-off between complexity of input and complexity of computation differs from case to case, and belief-function computations are manageable in a greater variety of situations than Spiegelhalter suggests.

In my article, I discussed Judea Pearl's work on propagating Bayesian belief functions in trees, and I noted that Pearl's Bayesian scheme is a special case of a general scheme for propagating belief functions in trees. This general scheme has now been described in some detail by Shafer, Shenoy, and Mellouli (1986). In recent unpublished work, Pearl and Spiegelhalter have made progress in dealing with Bayesian networks that are not trees. Similar work is also underway for belief functions, with the most important contribution so far being Augustine Kong's dissertation at Harvard (Kong, 1986). In the last chapter of this dissertation, Kong shows how the belief-function scheme of Shafer and Logan (1985) can be adapted to handle multiple diseases with no additional computational cost.

Spiegelhalter cites Zadeh in support of the view that Dempster's rule of combination can lead to unintuitive results. For a reply to Zadeh's arguments, see Shafer (1986a).

The Bishop of Bath and Wells whose work on probability Lindley discusses was named George Hooper. Hooper actually became a bishop only in 1703, long after his work on probability was published. Details about Hooper's life and work are given by Grier (1981). Hooper gave two rules for combining testimony, a rule for concurrent testimony and a rule for successive testimony. I have discussed these rules and their Bayesian counterparts elsewhere (Shafer, 1978, 1986c).

Hooper's rules were widely admired in the 18th century; they appear, for example, in Diderot's *Encyclopedie*. The Bayesian analysis that Lindley reviews, together with a corresponding analysis for the case of successive testimony, displaced Hooper's rules in the early 19th century (see Shafer, 1978). But this Bayesian account of "the probability of testimony" quickly became a laughingstock. It was roundly and justly denounced both by logicians critical of probability, such as John Stuart Mill, and by probabilists who preferred a frequentist interpretation, such as Antoine-Augustin Cournot.

The theory of belief functions does not require us to go back to Hooper's rules. Instead it provides a framework that includes both Hooper's analyses and the Bayesian analyses as special cases, along with many intermediate possibilities. The virtue of this flexibility is that we can tailor our analysis to our actual evidence. If we have significant prior evidence, we can use it. If we have evidence for causal dependence between the witnesses, we can use it. If we have instead evidence for dependence in our uncertainties about the witnesses, we can use it. By relating the numbers we offer to actual evidence in this way, we can hope to escape the ridicule that so wounded subjective probability in the 19th century.

ADDITIONAL REFERENCES

GRIER, B. (1981). George Hooper and the early theory of testimony. Dept. Psychology, Northern Illinois Univ.

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

SHAFER, G. (1978). Nonadditive probabilities in the work of Bernoulli and Lambert. *Arch. Hist. Exact Sci.* **19** 309–370.

SHAFER, G. (1986c). The combination of evidence. *Internat. J. Intelligent Systems* **1** 155–179.

SHAFER, G., SHENOY, P. and MELLOULI, K. (1986). Propagating belief functions in qualitative Markov trees. Working paper no. 186, School of Business, Univ. Kansas.

Rejoinder

Dennis V. Lindley

I find myself in general agreement with the contributions of Watson and Spiegelhalter. Watson is right when he says we do not have to accept Savage's axioms. But it is desirable to have an axiom system to support one's calculations and the lack of them must count against the alternatives to probability. Spiegelhalter is right when he says that ultimately it's the appeal of probability that matters: people will see that it makes good sense. Just as with Euclidean geometry, it is the operational aspect that counts, rather than Euclid. Watson queries the existence of the Great Scorer. I do not think it matters because one would wish to behave in such a way that one could not be exposed by his or her arrival. I would regard it as a serious proposal to pay meteorologists, or even medical doctors, according to their scores.

Whilst I find myself in dispute with Shafer, his arguments command respect and are not easily refuted. He contends that the axioms depend on conditional probability and expected utility, rather than that these depend on the axioms. While it is true that historically the concepts pre-date any axiom system, Savage introduced the axioms in order to justify a system, classical statistics, that denies conditional probability (of a hypothesis) and does not admit expected utility (with an expectation over unknowns); and he was much surprised when the axioms destroyed that system.

The scoring-rule argument works for almost every rule and does not depend on 0 or 1 as Shafer suggests. The preferences in Bayesian decision analysis are not necessarily sharp. If d_1 has expected utility 10.927 and d_2 10.926, then d_1 is preferred only slightly to d_2. The analysis is designed to select *an* act because only one act is typically possible.

Shafer also raises the issue of constructive probability. It is difficult, having experienced A_1, to think of probabilities for A_1 if only because probability describes uncertainty and A_1 is no longer uncertain. My response is that we should try to develop methods that