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 would help people to do this. If these all fail, then it will be necessary to think afresh. But forensic scientists, finding it necessary to think about probabilities for clothing stains (for example) have been able to assess them.

There remains the contribution by Dempster and Kong. They really throw the book at me and I am at a loss how to react. Certainly no response within the limits that the editor is likely to impose on me could be adequate. It is therefore perhaps best to remain silent except for one remark that touches on a point raised by others. One reason that I reject belief functions is that, at every stage, they are more complicated than probability—and that is hard enough, as Watson points out in connection with Schum's work. They involve more assessments and harder calculations. Furthermore, in my experience it is never necessary to extend the probabilistic argument in the way the theory of belief functions suggest. For example, if imprecision about a probability is relevant, then probability theory will require its assessment within its own calculus. Dempster and Kong reinforce this point when they take several paragraphs to solve the simple decision problem in their Table 1.

In conclusion may I thank those responsible for arranging the conference that led to these papers, and the editors for encouraging them to appear. I hope that readers will feel that the issues we address are important, both in theory and practice. If any readers feel they can meet the challenge it would be interesting to hear from them.

Rejoinder

David J. Spiegelhalter

By concentrating on applications, I appear to have escaped lightly in the discussion. Dr. Watson pointed out the multitude of criteria that could be used for evaluation of aids to clinical decision-making. Some order can be introduced by classifying all criteria according to whether they concern the system as *decision-maker* or as *aid*, and whether they are measures of *process* or *outcome*. Thus "internal coherence" is a process measure of the system as decision-maker, "comparison with experts" is an outcome measure as a decision-maker, "user satisfaction" is a process measure as an aid, and "effect on patients' health" is an outcome measure as an aid.

Professor Lindley was concerned about my interpretation of "uncertainty about a probability." Perhaps this phrase should not be used, since it does not differentiate between doubt in one's current beliefs due to *imprecision* in the probability assessments on which that belief is based, and sensitivity in that belief due to *ignorance* of potential future evidence. As evidence accumulates, the imprecision will generally increase as one gets into an increasingly narrow area of experience, but ignorance will be reduced. One's "point" current belief can therefore be thought of as the mean of two second-order distributions, representing what that belief might be now, and what it may become in the future.

Professor Shafer offers a vision of creative systems that can generate arguments in novel situations. He is correct that I, and my clinical colleagues, view expert systems in a much more limited sense, often having very little to do with the tenets of artificial intelligence, although exploiting their programming environments. I remain confident that probability is the appropriate tool in this area, and recent developments in strict probabilistic reasoning using local computations in general causal networks (Lauritzen and Spiegelhalter, 1987) overcome many technical problems. The parallels raised by Professor Shafer between probability/belief-function and expert-system/artificial intelligence contrasts are intriguing.

Both Professors Shafer and Dempster mention upper and lower expected losses from belief functions, which I find rather confusing. Are belief intervals to be interpreted as upper and lower probabilities or not? Suppose we adopt Dempster's decision theoretic structure after hearing "Slippery Fred's" evidence. Then $\{\mu\}_{\mathrm{Bel}}$ obey $.8 \le P$ (slippery) ≤ 1.0, which—from Shafer's original equation (3)—can easily be shown to impose the constraint $q \ge \max\{0, 4(1-2p)/(4-3p)\}$. If $p \ge 1/2$, then $\{\mu\}_{\mathrm{Bel}}$ is equivalent to $0 \le q \le 1$, which does not appear too unreasonable. However, the implicit constraints become much stronger after a crank of the rule-of-combination having seen the thermometer. Let us denote by r the probability the thermometer is right even if it is not working properly. To obtain coherently $\{\mu\}_{\mathrm{Bel}} = .04 \le P$ (slippery) $\le .05$, we require for, say $p = 1/2$, that $(3 + 97r)/(123 - 23r) \le$