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 $\{\mu\}_{\text{Bel}}$ 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 $\{\mu\}_{\text{Bel}}$ 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 $\{\mu\}_{\text{Bel}} = .04 \le P$ (slippery) $\le .05$, we require for, say $p = 1/2$, that $(3 + 97r)/(123 - 23r) \le$

$q \leq (23 + 77r)/(118 - 18r)$, which is a fairly narrow band around $q = r$. Thus, far from having no basis for specifying p, q and r, very strong constraints appear to have to be made in order for the decision-theoretic scheme to be coherent, were the probabilities available.

Shafer-Dempster belief intervals are widely interpreted as upper and lower probabilities in the expert-system world, but I had always thought this was an error. Now I admit to being confused.

ADDITIONAL REFERENCE

LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1987). Fast manipulations of probabilities with local representations—with applications to expert systems. Technical Report 87-7, Aalborg University Centre.