Comment

Glenn Shafer

I found it a pleasure to read the articles by Dennis V. Lindley and David Spiegelhalter. They present an elegant case for the use of Bayesian (i.e., conditional probability) methods in expert systems. Lindley provides a concise summary of arguments he and others have developed over the last several decades in support of the claim that rationality demands the use of conditional probability. Spiegelhalter supplements this with an account of what is actually being accomplished using conditional probability in diagnostic systems, and he also contributes some interesting new arguments.

In response, let me first express my admiration for the practical work Spiegelhalter reports on. The GLADYS system is especially attractive, because it brings close to reality the kind of probability calculation philosophers have always considered ideal—the calculation of conditional probabilities on the basis of massive and directly relevant frequency data. I share Spiegelhalter's excitement with the prospect that widespread use of microcomputers will enable us to attain this ideal much more often than we have in the past.

WHY GENERALIZE PROBABILITY?

Spiegelhalter emphasizes capabilities of the Bayesian language that are often overlooked. Weights of conflict can be used to monitor Bayesian analyses, and weights of evidence can be used to explain the results. There are Bayesian definitions of imprecision and ignorance. We do not need to generalize from Bayes to belief functions in order to formalize these concepts.

The desire to generalize Bayes does not spring, however, from dissatisfaction with the ideal of conditional probability. It springs from the realization that this ideal is sometimes unattainable. Directly relevant frequencies are often unattainable. Sometimes we can make decent conditional probability arguments even without such frequencies, but sometimes we cannot. Sometimes we simply lack evidence for some of the probability judgments that a given conditional probability analysis requires.

The only satisfactory description of uncertainty, Lindley tells us, is probability. He is no less correct than the man who believes that the only satisfactory household is one with a dozen servants. It's wonderful if you can afford it.

STANDARDS OF RATIONALITY

What should we say about the claim that rationality demands we make Bayesian analyses regardless of the availability of the ingredients? For my own part, I find that every argument for this claim boils down to another appreciation of the beauty of the Bayesian ideal.

Lindley believes that Savage's axioms are so self-evident that their violation would look ridiculous. But in fact these axioms derive their appeal from the Bayesian ideal rather than vice versa. If we did not have the picture of conditional probability and expected utility in mind, we would not even be able to understand most of Savage's axioms (Shafer, 1986b).

The idea of a scoring rule also derives from the Bayesian ideal rather than vice versa. It has relatively little force in abstraction from that ideal. If we intend to assign a number to each of two complementary events and accept a penalty for each event based on the number's distance from one if the event happens and its distance from zero if it fails, then we should make the two numbers add to one. But how would we explain this game to a naive listener? We would say that the numbers are supposed to be like probabilities—close to one for events that are expected to happen and close to zero for events that are expected to fail. The game fits the picture of additive, or frequency-like, probability, and it is incomprehensible outside that picture. It does not fit the theory of belief functions, where a degree of belief close to zero indicates inadequate evidence for the event, not assurance that the event will fail.

Another argument for Bayes is based on the relatively sharp preferences given by expected utility calculations. We can calculate upper and lower expectations from belief functions, but these will not give a definite preference between two alternatives as often as the Bayesian calculation will. But would we expect such sharp preferences were it not for our fascination with the Bayesian ideal? Would we really expect an analysis of our evidence and pre-existent preferences to tell us always exactly what to do, leaving no occasion for caprice? In fact, human beings, unlike Buridan's ass, are capable of choosing without sufficient reason, and they often use that capability. Building a similar capability into a computer is one of the easier tasks of artificial intelligence.

CONSTRUCTIVE PROBABILITY

In my contribution to this symposium, I say that Bayesian analyses use games of chance as canonical examples to which to compare actual evidence. Lindley says such games provide a standard by which to measure belief. There are commonalities here, but also important differences. It is difficult to use the verb "measure" without pretending that there is a well-defined property to be measured. Talk about canonical examples encourages a more constructive attitude.

One aspect of the constructive nature of Bayesian probability judgment, emphasized by Shafer and Tversky (1985), is the fact that we must construct our starting point. We must construct a probability distribution before we can condition it or multiply it by likelihoods. Bayesian theorists often assert categorically that every new experience must be treated in terms of its likelihood. Lindley, for example, declares that "an AI system faced with uncertainty about A_2 and experiencing A_1 has to update its uncertainty by considering how probable what it has experienced is, both on the supposition that A_2 is true, and that A_2 is false." But since a person may get around to constructing "initial" probabilities only after experiencing A_1, he or she has the option of treating A_1 as part of the evidence for those initial probabilities. Consider Lindley's investigator, who has discovered evidence that a criminal is left-handed. Instead of treating this evidence in terms of its likelihood, the investigator uses it directly in constructing a probability distribution.

There are problems, of course, where the construction can all be done in advance and then applied to many cases. GLADYS deals with this kind of problem; the same framework is applied to one patient after another. If I understand Spiegelhalter correctly, he believes that the bounded nature of expert systems means that this is the only kind of problem with which they can deal.

A finite system that permits construction can, however, deal with an unbounded range of situations. This is one of the fundamental points of the generative theory of grammar. The constructive nature of human reasoning makes us capable of exploring ever new realms of experience, and the ambition of AI is to duplicate this capability. Rule-based expert systems are one attempt to do so. These systems do not handle probabilistic reasoning very well, and many AI theorists would conclude from this that probabilistic reasoning has little role in genuine intelligence. In order to prove them wrong, we must do more than retreat to bounded systems like GLADYS. We must take the problem of automating construction seriously.

ADDITIONAL REFERENCE

SHAFER, G. (1986b). Savage revisited (with discussion). *Statist. Sci.* **1** 463–501.

Comment: A Tale of Two Wells

Dennis V. Lindley

The main issue is whether uncertainty should be described by probability, belief functions, or fuzzy logic; not just in artificial intelligence and expert systems, but generally. Are we to be probabilists, believers, or fuzzifiers? Or do we need some mixture of all three disciplines? To me, the important distinction between the methods rests on the rules of combination of uncertainty statements. Do we operate with the calculus of probability, the rules of belief functions, or with those of fuzzy logic? In my paper the challenge was made "that anything that can be done by these methods (belief functions and fuzzy logic) can better be done with probability." This reply will address one such challenge and I hope to show that Dempster's rule for belief functions does not behave as well as Bayes rule. My discussion is therefore chiefly addressed to Shafer and Zadeh. The omission of any discussion of Spiegelhalter's contribution arises because I agree substantially with it, and highly regard it. I wish that his program for dyspepsia had been more Bayesian and that he had recognized that uncertainty about a probability is usually a reference to the desirability of obtaining more data, so that his conflict ratio should really reflect this. To return to the challenge.

In 1685 the then Bishop of Bath and Wells wrote a paper in which the following problem was discussed. Two witnesses separately report that an event is true. Both are known to be unreliable to the extent that they only tell the truth with probabilities p_1 and p_2 respectively. What reliability can we then place, in the light of the witnesses' testimonies on the truth of the event? The Bishop's answer was $1 - (1 - p_1)(1 - p_2)$. The following is a precis of his argument. If the event is false, both witnesses must have lied, an event of probability $(1 - p_1)(1 - p_2)$. Consequently one minus this is the required reliability.

The result retains its interest today because the