In general, a combination of approaches is needed to provide a user with timely understandable support when a decision must be made.

Glenn Shafer and Prakash Shenoy (University of Kansas): The ideas on which Lauritzen and Spiegelhalter are working are of great importance not only for expert systems but also for our general understanding of probability.

A central lesson of the paper is that we can exploit for combining evidence the same conditional independence structure that we use in thinking about causation. This lesson is widely relevant. It applies not only to situations where sensible full probability distributions are available, but also to situations where the data are more fragmentary and we must make do with partial or qualitative judgements.

The conditional independence structure is more important than the probability numbers or the rules of calculation. We can replace the numbers by a verbal scale (likely, very likely, etc.), and we can replace Bayes's theorem by other rules (as we do in the theory of belief functions), but we cannot dispense with a structure that tells us what evidence bears on what propositions.

It is disturbing to note that a student must study probability for a long time before he or she will see this importance of structure. Markov trees and fields are topics for an advanced course in probability, and path analysis will usually appear only in a fourth or fifth course in statistical inference. Lauritzen and Spiegelhalter's paper suggests that we must change this if we want, as statisticians, to retain intellectual leadership in the area of probability. We must emphasise the importance of structure in our elementary teaching. One way of doing this is to recognise the need for structure to justify the rule of conditioning (Shafer, 1985).

Would the authors clarify further the relation between their work and the work of Pearl (1986a)? It is obvious from Fig. 2 that we can successively calculate the clique marginals for the cliques $\{\alpha, \tau\}$, $\{\lambda, \beta, \sigma\}$, $\{\varepsilon, \tau, \lambda\}$, $\{\varepsilon, \beta, \lambda\}$, $\{\delta, \varepsilon, \beta\}$ and $\{\varepsilon, \xi\}$ in one just pass through the graph. This natural ordering of the cliques can also be used for propagating new evidence efficiently, as Pearl has shown. The authors show us a more flexible approach, one which can use any triangulation and any ordering with the running intersection property. But what is gained by this flexibility?

Professor Philippe Smets (Université Libre de Bruxelles): Although the authors did not intend to compare the pros and cons of the various models they propose to handle uncertainty in a coherent manner, it is questionable whether their proposed Bayesian model is as meaningful as they claim. Once applicable, probabilities are perfect, but are they really applicable for medical diagnosis? For instance, can one defend the p value in $P(\text{MU.loss} = \text{other} \mid \text{disease} = \text{other}) = p$?

To simplify, suppose $B \equiv$ bronchitis and $D \equiv$ dyspnoea. I know what $P(D \mid B)$ is but what about $P(D \mid \bar{B})$? The difference is due to the 'well-defined' nature of the set B and the 'ill-defined' nature of the set $\bar{B}$. The causal relation between diseases and symptoms is usually clear and its translation into a conditional probability can be defended, but what about $P(D \mid \bar{B})$? Let the set of diseases be defined by the family of mutually exclusive and exhaustive diseases $B_1 B_2 \ldots B_n$ $(B = B_1)$. For each B_i, I can define $P(D \mid B_i)$ but to define $P(D \mid \bar{B}_i)$ I need the *a priori* distribution of the B_is. If I have it, as in a well-defined population, then $P(D \mid \bar{B})$ is well defined. But medical diagnosis is not performed in a well-defined context.

In reality, the population is poorly defined. $P(D \mid B_i)$ can be used as essentially it does not depend on time and space, but the prior repartition $P(B_i)$ is not so constant. It varies in time and space. $\bar{B}$ is an every-varying hotchpotch of people. It is in practice unknown; therefore $P(D \mid \bar{B})$ is also unknown.

If $P(D \mid \bar{B})$ cannot be estimated the algorithm breaks down. We could accept the unavailability of $P(D \mid \bar{B})$, compute the limits between which $P(D \mid \bar{B})$ must be... and derive upper and lower probabilities, or we could use the transferable belief model, which allocates parts of a total unitary belief to some subsets (as with probabilities) without requiring additivity (Smets, 1988). The part of belief assigned to a subset quantifies the amount of belief that supports that subset and does not support any strict subset due to lack of relevant information. If new relevant pieces of evidence become available, that amount of belief could be allocated to some subsets. This model implies implicitly the use of belief functions and Dempster's rules of conditioning and combination. This transferable belief model looks similar to Dempster and Shafer's model (Shafer, 1976) except that it does not require any underlying concept of probability and Dempster's rules are not arbitrary. In that model, $\text{bel}(D \mid B)$ is defined as in the Bayesian model but $\text{bel}(D \mid \bar{B})$ can be described by a vacuous belief function that describes adequately a state of total ignorance (which can hardly be done within the Bayesian framework). The use of the transferable belief model requires fewer conditional probabilities than the Bayesian model and the computation will be even easier than that needed within the Bayesian framework.

Discussion of "Local computation with probabilities on graphical structures and their application to expert systems," by S.L. Lauritzen and D.J. Spiegelhalter. Journal of the Royal Statistical Society, Series B, 50. 1988