Statistical Science
1987, Vol. 2, No. 1, 3–44

Probabilistic Expert Systems in Medicine: Practical Issues in Handling Uncertainty

David J. Spiegelhalter

Abstract. The development of expert systems in medicine has generally been accompanied by a rejection of formal probabilistic methods for handling uncertainty. We argue that a coherent probabilistic approach can, if carefully applied, meet many of the practical demands being made, and briefly illustrate our claim with three current projects.

Key words and phrases: Evidence propagation, knowledge representation, graphical models, imprecise probabilities.

1. INTRODUCTION

The first problem in discussing "uncertainty in expert systems" comes in defining our terms. We shall view "expert systems" as being programs intended to provide judgments or advice to users in a reasonably convincing manner, in which *knowledge*, whether represented as rules, networks, or frames, is generally characterized by local relationships between propositions of varying generality; *data* is obtained on a new case, upon which the "knowledge" is brought to bear by means of a controlling mechanism. Characteristics which are often said to distinguish such programs from standard statistical or mathematical models include the frequent use of subjective judgments for both the qualitative structure and any accompanying quantification, the emphasis on explanation, and the incorporation of both knowledge and data that is fragmentary. Systems are also often intended to enable "learning," in which knowledge is adjusted in the light of data on past cases.

Within this context the term "uncertainty" is used in a very wide sense and this has led to considerable argument about the role appropriate to formal probabilistic reasoning (Cheeseman, 1985; Spiegelhalter, 1986a, 1986b; and papers in Kanal and Lemmer, 1986). Some misunderstanding may have arisen from the common use of expressions of the form

IF conditions X hold, THEN Y with certainty P.

If Y is a random event which is currently unknown, the statistical view is that P represents a kind of "predictive" uncertainty expressed as a probability (see Lindley, page 18). However, in many expert system applications, precisely the same representation is used when Y signifies some *action* or choice, and P essentially corresponds to "procedural" uncertainty, in that doubt is being expressed about the reasonableness of, or the support for, performing an act or making an assumption. Thus, Cohen (1985, page 52) states that "one's certainty in a result should depend on what the result is wanted for," and Van Melle et al. (1981, page 5) say that "certainty factors" in EMYCIN combine subjective probabilities and utilities to measure "importance."

Thus, there is clearly great potential for confusion between the fairly restricted, statistical sense of uncertainty as applied to *facts* and the use of the term in a broader, linguistic sense in describing uncertainty about *acts*. To try to avoid this confusion in this short paper, we shall explicitly restrict attention to uncertainty concerning potentially verifiable, but currently unknown, events.

We shall concentrate on practical, rather than philosophical, issues concerning the way uncertainty is handled in existing programs. We shall not consider in detail either the representation of knowledge or the control of the program. Published examples motivate the search for a methodology that satisfies a number of demands, and three current projects will then be used to illustrate some specific aspects of the attempt to use probabilistic methods in as effective a way as possible. Finally, an attempt is made to bring the argument together into a prospect for future developments.

2. DEMANDS MADE OF A CALCULUS

The particular complexity of many medical problems has challenged the notion of a rigorous unified treatment of uncertainty and, in general, ad hoc quantifications have been used to measure evidence

David J. Spiegelhalter is a statistician at the Medical Research Council, MRC Biostatistics Unit, 5 Shaftesbury Road, Cambridge CB2 2BW, England.

for various possible underlying hypotheses (Szolovits and Pauker, 1978). The complex interrelationships between disease processes and manifestations have led to various systems for propagating degrees of certainty arising from fragmentary data and combining evidence from different sources. PIP (Pauker et al., 1976) and INTERNIST/CADUCEUS (Miller, Pople, and Myers, 1982) both essentially score hypotheses using evidence from current symptoms that support a hypothesis, which is discounted by a score expressing absent symptoms that would be expected, and a score expressing present symptoms that would *not* be expected. MYCIN/EMYCIN use a more modular structure in which certainty factors are attached to propositions, although CASNET/EXPERT (Kulikowski and Weiss, 1982) propagates weights through a causal network. A statistical system such as that of de Dombal et al. (1972) begins with knowledge derived from a data base, but the simplistic independence assumptions made in combining evidence (although effective in discrimination) ensure that the certainty propagated is not expected to be interpretable as a probability—the same holds for the Bayesian updating technique in PROSPECTOR (Duda, Hart, and Nilsson, 1976). Fuzzy reasoning (Adlassnig, 1980; Fieschi et al., 1983) has also been used as a means of capturing the ill-defined nature of many clinical terms.

We can identify a number of considerations that have led to the procedures that have been adopted and that are currently being researched. The strongest has been the claim that a single probability of a hypothesis, even if it were based on extensive data, is not sufficient to convince a clinician: the *evidence* on which to base a conclusion must be retrievable, to enable conflicts and doubtful contributions to be identified. A particular case of this demand for justification is the situation where little relevant data is available and there is essentially *ignorance* concerning the possibility of a hypothesis. This arises particularly in medicine due to the hierarchical, taxonomic structure of disease descriptions in which evidence may be available which supports a general disease category but gives no indication of the relative plausibility of the subcategories of disease. Thus, the hierarchical hypothesis structure is viewed as a natural justification for *ranges* of uncertainty, for which a number of schemes exist (see, for example, Quinlan, 1983), although as we shall emphasize later it is not generally made clear whether such ranges are due to inadequate knowledge or inadequate data. The demand that individual contributions of pieces of evidence should be identified, and that evidence should be able to focus on groups of diseases without distinguishing within that group, has led naturally to the study of the possible role of belief functions in medicine (Gordon and Shortliffe, 1984). Much attention is now being paid to solving the accompanying computational problems and making some allowance for dependencies between sources of evidence. The concept of discounting in belief functions could also be seen as a means of allowing for doubt about the precise numbers to be placed on evidential statements.

To summarize: current interest is focussed on schemes that can propagate measures of uncertainty through complex relationships often defined on a hierarchical structure, that can identify conflicting evidence and lack of evidence, and can cope with incoming data that do not follow a predefined order. The reasoning process should be justifiable and fairly intuitive, and allowance for imprecise specification of numerical relationships would be an advantage.

Although the above desiderata appear admirable, we feel there is an important item that has been largely ignored in practice. This concerns the *operational meaning* of the quantities which express uncertainty and which allows a reasonable basis for both assessment of inputs and criticism of outputs of a system. In the following examples we describe attempts to retain meaning while responding to demands and constraints made by the real practical problems of interest. Refer to Pearl (1986a, b) for further discussion on how probabilistic reasoning can be adapted to meet the demands of expert systems.

3. EXAMPLES OF PROBABILISTIC ANALYSIS

GLADYS—the GLAsgow DYSpepsia System

GLADYS is a program designed to interview patients presenting to a clinic with dyspepsia, and provide a reasoned probabilistic diagnosis based on the symptoms alone. It was developed at the Diagnostic Methodology Research Unit at Glasgow, and runs on a microcomputer with a special keyboard to record patient responses. The control of the interview is strictly algorithmic, in that branches to more detailed interrogation are taken depending on the results to trigger questions. The interview has been found to be accurate and acceptable (Lucas et al., 1976). The responses are analyzed according to a scoring system derived from a modified logistic regression technique, described in detail in Spiegelhalter and Knill-Jones (1984), of which certain aspects are relevant to the issues raised in the previous section.

Firstly, there is a real need to deal with *hierarchical* disease structures, in which for example, certain features may discriminate the generic class peptic ulcer (PU) from other diseases, although other items of information are relevant to discriminating duodenal from gastric ulcer (GU) within the peptic ulcer class. This is accomplished by calculating probabilities conditional on the branch in the hierarchy and then multiplying downward to obtain the overall probability: for example, we calculate $p(\mathrm{GU} \mid \mathrm{PU})$ and $p(\mathrm{PU})$ from which $p(\mathrm{GU}) = p(\mathrm{GU} \mid \mathrm{PU})p(\mathrm{PU})$.

Secondly, the scoring system allows *explanation* of the final probability in terms of the contributing pieces of evidence. For example, a patient described in Spiegelhalter and Knill-Jones (1984) provided the following evidence relevant to a diagnosis of gallstones:

Evidence FOR gallstones		Evidence AGAINST gallstones	
History less than 6 months	77	Pain not severe enough to warrant emergency call to doctor	−43
Pain comes in "attacks"	177	Pain does not radiate	−38
Can enumerate attacks	63		
Attacks produce restlessness	31		
Pain in right hypochondrium	77		
Total	425		−81
Balance of evidence	+344	(Total evidence = 425 + 81 = 506; conflict ratio = 506/344 = 1.5)	
Initial score	−300	(Corresponding to prevalence of 4.7%)	
Final score	44	= 61% chance of gallstones	

Some explanation of the above "explanation" is necessary. The scores given to findings are 100 $\log_e$(likelihood ratios) adjusted, roughly speaking, for correlations between items of information. Thus, the initial score of $S = -300$ is transformed to a prior probability $p = 1/\{1 + \exp(-S/100)\} = .047$, which is simply the inverse of $S = 100 \log_e\{p/(1 - p)\}$. The "conflict ratio" (= total evidence/ | balance of evidence |) is a rough measure of how much the total evidence obtained contradicts itself: a high ratio, say above around 2.5, suggests the clinician should check some of the important questions. The initial score is based on a prevalence in an urban clinic and could be altered depending on circumstances. The scores come from analysis of a data base of 1200 cases and the statistical modelling means the final probabilities are reasonably well calibrated, in that of patients presenting as above, around 60% should turn out to have gallstones as a major cause of their symptoms. This is a very popular characteristic of the system. There is, however, no reason why the scores should not be subjectively assessed provided one monitors whether the predictions have similar properties of calibration.

Thirdly, *imprecision* of the quantification could be incorporated by placing standard errors on the predictions. The above example has a standard error of 42 on the final score corresponding to a rough 95% interval of (.40, .78) on the predictive probability. Finally, *ignorance* may be viewed retrospectively in terms of the total evidence received either for or against a proposition. However, as suggested in Spiegelhalter and Knill-Jones (1984), we may also quantify prospective ignorance in terms of the results that may occur when the data of which we are currently ignorant becomes available. This concept translates into calculating the predictive distribution of the possible final probabilities that may be ascribed to a disease. Tukey (1984) recommended that such distributions should be included as part of the explanation facilities. Thus before an interview, a patient has a fairly *precise* probability of gallstones (95% interval .03, .07), but one based on an ignorance reflected in the wide distribution of feasible probabilities that could be taken on when data become available; whereas at the end of the interview, there is a relatively *imprecise* probability with a 95% interval of (.40, .78), but no remaining ignorance within the bounded context of the system.

We would not normally consider GLADYS as an expert system since it does not use knowledge representation techniques derived from AI, it is not based on expert opinion and it does not operate interactively. However, many of our aims match those of classic expert systems, except that we are determined to remain, as far as possible, within a probabilistic framework.

A Diagnostic System for Chest Diseases

A group at the Chest Clinic at Westminster Hospital are developing a system for probabilistic diagnosis of patients presenting with a normal chest x-ray. The system uses simple independent Bayes updating assuming mutually exclusive disease categories, and our only concern here is with the subjective probability assessments on which the system is initially based. The consultant physician has been required to assess prior probabilities for each of the diseases conditional on the age group of the patient and the main presenting symptoms, as well as the probabilities of the secondary symptoms conditional on each of the diseases. Around each probability he was required to place an interval reflecting his confidence in the point probability. By viewing this range as an approximate 90% interval around a binomial probability one can derive a rough implicit sample size on which his judgment of each probability has been based. These measures of imprecision are currently not propagated through the consultation, although Rauch (1984) suggests ad hoc methods of doing this while allowing for correlated judgments. However, the implicit sample sizes allow the probabilities to be stored as a fraction r/n, and where a confirmed case with the relevant symptom is found the probability may be updated to $(r + 1)/(n + 1)$. This emphasises that probabilistic systems may be based on subjective opinion, and yet a rational means of allowing that opinion to learn from experience is easily available.

IMMEDIATE—A System for General Practice

In contrast to GLADYS, IMMEDIATE is a rule-based AI system written in PROLOG which is being developed by a group centered at the Medical Computation Unit at the University of Manchester. It is designed to support certain activities of general practitioners and its control philosophy is described elsewhere (Dodson and Rector, 1985).

Two aspects of its development are of interest here. Firstly, although the knowledge structure and uncertainty propagation bears some resemblance to that of PROSPECTOR, a deliberate aim is that the probabilities should be made to cohere: thus initial probability judgments should form a valid joint distribution, and, as data arrives, uncertainty be propagated in a way that retains its interpretation as subjective probability. Secondly, part of the control mechanism is based on a range of ignorance or evidence availability that is an explicit calculation of the maximum and minimum probabilities of a proposition that could be achieved when further information becomes available. This may be seen as a summary measure of the predictive distributions of final probabilities described under GLADYS. Explicitly calculating the range of potential probabilities of a proposition helps toward an assessment of the importance of establishing relevant patient characteristics, which in turn ensures that the clinician is informed as to the most telling questions to ask.

4. DISCUSSION

The preceding section is an inadequate glimpse of some work currently being carried out in probabilistic systems, and we have only been able to mention aspects according to their capacity to illustrate the practical implementation of important issues in the handling of uncertainty. In this section, we attempt to summarize these issues with the aid of examples drawn from the systems introduced above.

Status of Propositions

It is clearly preferable that all propositions in a system are crisply defined and, at least theoretically, verifiable at some point in the future, as required by Smith (1961) or de Finetti (1974). Nevertheless, the inevitable imprecision of statements (e.g., "the pain is relieved by food") makes it tempting to allow degrees of truth of propositions and adapt a fuzzy calculus. It should, however, be emphasized that it is not the true state of the world to which the system has access, but the *assertion* of the state of the world (The patient has replied YES to the question "Is the pain relieved by food?"), and this is necessarily made crisp by the restricted means one has to put information into the system (e.g., just a YES/NO button). An expert system can therefore force the user to be categorical in his assertions, although we acknowledge that user demand for qualifications of degree may create the need for an alternate calculus to deal with partly true propositions.

A statistician may tend to view a knowledge base as a set of related nodes, each corresponding to a random variable which may take on a number of mutually exclusive and exhaustive values. The rules attempt to define a distribution on the variables. For control purposes, however, it may be necessary to have action nodes which correspond to conclusions on which further analysis is conditioned. These may well not be strictly verifiable propositions; for example, in a system designed for statistical analysis, there may be assertions of normal errors or linear relationship. Strictly speaking a decision-theoretic argument should be used for any interim decision made in the control of a consultation, but this is not usually practicable. As suggested in the "Introduction," the justification for probability is not so clear in these cases, instead it could be reasonable to adopt a calculus of compatibility or degree of support for a hypothesis or conclusion for which a probability is not well defined.

Knowledge Representation and Explanation

We feel that probabilistic methods can handle hierarchical taxonomic structures without extending into belief function methodology (Pearl, 1986b). There is, however, a great need for further work in coherent assessment and propagation of probabilities through the network structures arising from rule-based systems. The graphical representations of certain log linear models described by, for example, Wermuth and Lauritzen (1983) are crucial, with propagation schemes extended from those of Kim and Pearl (1983); Spiegelhalter (1986b) describes efficient propagation schemes allowing for imprecise probabilities and automatic tuning of the subjective assessments.

Subjective judgments may be deliberately *overspecified* to allow for identification of incoherence due to poor assessments or weak modelling, or *underspecified* and padded out using, for example, the maximum entropy methods of Cheeseman (1983). By using such a structure and explanation facilities similar to GLADYS, one should be able to fulfill the aim, described by Dempster (1985), of justifying quantified judgment explicitly in terms of the sources of evidence.

Intervals and Probabilities

As we emphasised in discussing GLADYS, two types of range of probability must be distinguished. The first, due to inadequacies in the knowledge base, concerns the *imprecision* in the quantifications. This may be represented by a standard error or even a fuzzy qualifier, but in either case the range represents a type of automatic sensitivity analysis conditional on the data already obtained. This interval will generally tend to widen as more data come in.

This should be contrasted with an interval based on *ignorance* concering the current case, and one way in which this can be defined is in terms of the probabilities that could be taken on when the unknown data, denoted X, becomes available. If D represents a disease with current probability $p(D)$, then the predictive distribution of the eventual probability $p(D \mid X)$ may either be fully calculated as in GLADYS or summarized by its range as in IMMEDIATE. We note that by conditional expectation, $E\{p(D \mid X)\} = p(D)$. Hence our current probability may simply be thought of as the mean of the distribution of possible final probabilities. This distribution narrows as the consultation proceeds.

In this way *ignorance* is explicitly defined in terms of the X that we do not yet know. In real life, X is unbounded and so such a calculation is unreasonable, but it is important to note that an expert system is *bounded* and so can always explicitly state what information is missing, provided a suitably efficient search routine is available.

Operational Meaning

Our practical experience has strongly influenced us toward establishing operational meaning to our expression of uncertainty. This has three stages: firstly, the *inputs*, based on either real or imaginary past data, must have sufficient interpretation to allow informed argument. Clinicians often disagree strongly about subjective probabilities, but we have found the resulting discussions illuminating and constructive. The problems of agreeing on numbers with no verifiable interpretation is vividly illustrated in the fascinating transcript of an argument concerning certainty factors contained in the book on the MYCIN projects (Buchanan and Shortliffe, 1984). Secondly, preserving operational meaning in the propagation of uncertainty requires attention to the coherence of the assessments when placed in a large, complex knowledge base. Finally, the *outputs* need to have an externally verifiable interpretation in terms of their calibration against experience. Such calibration is not part of the axioms of subjective probability, but we have found an enthusiastic response from clinical colleagues when they find the predictions provide reasonable betting odds. Of course, a system may process information solely with the aim of providing a, possibly ranked, set of alternatives with some attached measure of evidential support. However, if a system is to be used to guide the *choice* of an option, or its outputs are to be used as inputs to another system, this seems to be inadequate. In fact, a subjectivist statistician may view a diagnostic expert system as a coherence machine, which takes in relevant information, and throws out acceptable betting odds on future events.

Finally, perhaps the most important reason for interpretable quantification is the need for *learning*. As we have illustrated with the chest disease system, updating of subjective probabilities is feasible and should provide a convergence of opinion that may overcome local biases which may otherwise render a system unacceptable.

REFERENCES

ADLASSNIG, K. P. (1980). A fuzzy logical model of computer-assisted medical diagnosis. *Methods Inf. Med.* **9** 141–148.

BUCHANAN, B. G. and SHORTLIFFE, E. H., eds. (1984). *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, Mass.

CHEESEMAN, P. (1983). A method of computing generalised Bayesian probability values for expert systems. *Proc. of the Eighth International Joint Conference on Artificial Intelligence.* 198–202. William Kaufmann, Los Altos, Calif.

CHEESEMAN, P. (1985). In defense of probability. *Proc. of the Ninth International Joint Conference on Artificial Intelligence.* 1002–1009.

COHEN, P. R. (1985). *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach.* Pitman, Boston.

DE DOMBAL, F. T., LEAPER, D. J., STANILAND, J. R., MCCANN, A. P. and HORROCKS, J. C. (1972). Computer-aided diagnosis of acute abdominal pain. *British Med. J.* **2** 9–13.

DE FINETTI, B. (1974). *Theory of Probability* **1**. Wiley, New York.

DEMPSTER, A. P. (1985). Probability, evidence and judgment. In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.) **2** 119–131. North-Holland, Amsterdam.

DODSON, D. C. and RECTOR, A. L. (1985). Importance-driven distributed control of diagnostic inference. In *Research and Development in Expert Systems* (M. A. Bramer, ed). Cambridge Univ. Press, Cambridge.

DUDA, R. O., HART, P. E. and NILSSON, N. J. (1976). Subjective Bayesian methods for rule-based inference systems. *Proc. AFIPS Natl. Compt. Conf.* **47** 1075–1082.

FIESCHI, M., JOUBERT, M., FIESCHI, D., BOTTI, G. and ROUX, M. (1983). A program for expert diagnosis and therapeutic decision. *Med. Informatics* **8** 127–135.

GORDON, J. and SHORTLIFFE, E. H. (1984). The Dempster-Shafer theory of evidence. In Buchanan and Shortliffe (1984), 272–292.

KANAL, L. N. and LEMMER, J., eds. (1986). *Uncertainty in Artificial Intelligence.* North Holland, Amsterdam.

KIM, J. H. and PEARL, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. *Proc. of the Eighth International Joint Conference on Artificial Intelligence.* 190–193. William Kaufmann, Los Altos, Calif.

KULIKOWSKI, C. A. and WEISS, A. M. (1982). Representation of expert knowledge for consultation: the CASNET and EXPERT projects. In *Artificial Intelligence in Medicine* (P. Szolovits, ed.) 21–55. Westview Press, Colorado.

LUCAS, R. W., CARD, W. I., KNILL-JONES, R. P., WATKINSON, G. and CREAN, G. P. (1976). Computer interrogation of patients. *British Med. J.* **2** 623–625.

MILLER, R. A., POPLE, H. E., JR. and MYERS, J. D. (1982). INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. *N. Engl. J. Med.* **307** 468–476.

PAUKER, S. G., GORRY, G. A., KASSIRER, J. P. and SCHWARTZ, W. B. (1976). Towards the simulation of clinical cognition: taking a present illness by computer. *Amer. J. Med.* **60** 981–986.

PEARL, J. (1986a). On evidential reasoning in a hierarchy of hypotheses. *Artificial Intelligence* **28** 9–15.

PEARL, J. (1986b). Fusion, propagation and structuring in belief networks. *Artificial Intelligence* **29** 241–288.

QUINLAN, J. R. (1983). Inferno: a cautious approach to uncertain inference. *Comput. J.* **26** 255–269.

RAUCH, H. E. (1984). Probability concepts for an expert system used for data fusion. *Artificial Intelligence Mag.* 55–60.

SMITH, C. A. B. (1961). Consistency in statistical inference and decision (with discussion). *J. Roy. Statist. Soc. Ser. B* **23** 1–37.

SPIEGELHALTER, D. J. (1986a). A statistical view of uncertainty in expert systems. In *Artificial Intelligence and Statistics* (W. Gale, ed.) 17–56. Addison-Wesley, Reading, Mass.

SPIEGELHALTER, D. J. (1986b). Probabilistic reasoning in predictive expert systems. Cited in Kanal and Lemmer (1986).

SPIEGELHALTER, D. J. and KNILL-JONES, R. P. (1984). Statistical and knowledge-based approaches to clinical decision-support systems, with an application in gastroenterology (with discussion). *J. Roy. Statist. Soc. Ser. A* **147** 35–77.

SZOLOVITS, P. and PAUKER, S. G. (1978). Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence* **11** 115–144.

TUKEY, J. W. (1984). Discussion of Spiegelhalter and Knill-Jones. *J. Roy. Statist. Soc. Ser. A* **147** 62–64.

VAN MELLE, W., SCOTT, A. C., BENNETT, J. S. and PEAIRS, M. A. S. (1981). The EMYCIN Manual. Report HPP-81-16, Computer Science Dept., Stanford Univ.

WERMUTH, N. and LAURITZEN, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70** 537–552.

Comment

Stephen R. Watson

1. COMMENTS ON SHAFER'S PAPER

One of the things that makes Shafer's theory interesting is that it can be seen as an alternative to the traditional probability theory. Is this really so, however? Firstly, note that one of the strengths of subjective probability theory is the clear cut nature of the axiomatic support for the theory. Indeed, as Lindley's contribution shows, it is possible to claim that probability theory is the only theory one could possibly use to represent uncertainty. Shafer's theory does not as yet have such a clear cut support. For example, although Shafer recognizes the importance of canonical examples, as yet belief function theory is not provided with as strong an axiomatic support as that which is available for probability theory.

It can be claimed, however, that belief functions are indeed rooted in probability theory. It is just that the probability is associated with a power set rather than a simple set. If this interpretation of belief function theory is accepted, then indeed there is no problem, since the philosophical support for probability theory clearly also will support belief function theory. However, Shafer seems in some of his writings not to be very happy with this interpretation of his theory. And if he rejects this interpretation then the problem of a philosophical foundation for belief function theory remains.

The second point I make here concerns concepts of independence. Shafer touches on this point in his paper, but it is worth saying again that concepts of independence in belief function theory are not yet clear. In the application of Dempster's rule to determine the support for a hypothesis on the basis of two pieces of evidence, there is a rather vague notion that the two pieces of evidence should be independent in

Stephen R. Watson is Peat Marwick Professor of Management Studies in the University of Cambridge. He is also Head of the Management Studies Group in the Department of Engineering and a Fellow of Emmanuel College. His address is Department of Engineering, University of Cambridge, Mill Lane, Cambridge CB2 1RX, England.