Varieties of ignorance and the need for well-founded theories.

Philippe SMETS[1]
IRIDIA-Université Libre de Bruxelles
50 av Roosevelt, CP194/6, 1050 Brussels, Belgium

1. Introduction.

Recent research in automated reasoning have been oriented toward the implementation of commonsense reasoning in AI. Such an enterprise could only succeed if uncertainty could be mastered by appropriate approximate reasoning models (Lopez de Mantaras, 1990). Probability theory, the obvious candidate for modelling uncertainty, was seen as too limited to express all forms of uncertainty.

New models have been developed based either on numerical methods (possibility theory, theory of evidence, theory of the certainty factor) or non-numerical methods (non-monotonic logic, default logic, autoepistemic logic). Our discussion is focussed on numerical approaches but could be extended to non-numerical approaches mutis mutandis.

The great danger in computer implementation of approximate reasoning is the use of inappropriate, unjustified, ad hoc models. Newcomers in the domain of commonsense reasoning could be overwhelmed by the multitude of models. Their reaction could be either to accept one of them and use it in every context or to use all of them somehow at random. Both attitudes are wrong. The stubborn use of one model is inappropriate as ignorance, uncertainty and vagueness are really different concepts. Random usage usually leads to inappropriate matching.

This paper is a plea for the use of correct models. An understanding of the forms of ignorance and the nature and the foundations of each model are required. Before using a quantified model, we must

1) provide a meaning for the numbers, i.e. provide canonical examples where the origin of the numbers can be justified
2) understand the fundamental axioms of the model and their consequences. The choice of axioms should be justified by "natural" requirements.
3) study the consequence of the derived models in practical contexts to check their validity and appropriateness

A common error consists in accepting a model because it 'worked' nicely in the past. This property is not a proof that the model is correct. Experimental results can only prove that a model is wrong, not that it is correct. They only give hints about its value.

To illustrate our message, we present a survey of certain forms of ignorance and of the mathematical models that have been suggested to quantify ignorance. We first present an example of an inadequate model. We present the survey, and finish with a plea for future studies covering the integration of the models.

---

[1] The following text presents research results of the Belgian National incentive-program for fundamental research in artificial intelligence initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. Scientific responsibility is assumed by its author. Research has partly been supported by the DRUMS (Defeasable reasoning and Uncertainty Management Systems) project funded by EEC grants under the ESPRIT II Basic Research Project 3085.

2. Example

Certainty factors (Shortliffe and Buchanan, 1975) based models were probably the first to be used in an expert system to quantify uncertainty. Their use results from the excellent insight that probability models are too restrictive to model quantified beliefs as they appear in diagnostic contexts. A piece of evidence e could support a hypothesis h without necessarily supporting the complement of that hypothesis. The authors rejected the rule that

$$\text{belief}(h|e) = f(\text{belief}(\neg h|e))$$

where belief(h|e) is the belief that hypothesis h is true given the piece of evidence e. But for lack of alternative models they created an ad hoc model based on measure of belief, disbelief and what is known by now as certainty factors. The aim was great, the result was shaky.

Not only was he meaning of the numbers not supported but what was more, the models did not satisfy some fundamental requirements. What does .7 mean in "my CF is .7"? Why not .6 or .8? Some yardstick is required. For subjective probability theory, urns provide a yardstick such as our belief that the next randomly selected ball is white is equal to the proportion of white balls in the urn - an objective unassailable value. Analogous canonical examples have also been developed for belief functions based models (Shafer and Tversky, 1985).

As an example of the weakness of the model itself, suppose there are two rules IF E_1 THEN H and IF E_2 THEN H with certainty factor CF_1 and CF_2. The parallel combination of CF_1 and CF_2 gives the combined CF_{12} for H given both pieces of evidence:

$$(*)\qquad CF_{12} = \begin{cases} CF_1 + CF_2 - CF_1.CF_2 & \text{if } CF_1 \text{ and } CF_2 \text{ are positive} \\ CF_1 + CF_2 + CF_1.CF_2 & \text{if } CF_1 \text{ and } CF_2 \text{ are negative} \\ \dfrac{CF_1 + CF_2}{1 - \min(CF_1, CF_2)} & \text{if } CF_1.CF_2 \text{ is negative} \end{cases}$$

A natural requirement for any model for parallel combination is that the combination should be associative. But the rules (*) are not associative.

Furthermore (*) should never be applied if E_1 and E_2 are deduced from a common piece of evidence E_0. Suppose we have the following rules

IF E_0 THEN E_1
IF E_0 THEN E_2
IF E_1 THEN H
IF E_2 THEN H

where all certainty factors are 1. Let the certainty factor of E_0 be α, then both E_1 and E_2 have certainty factor equal to α. A blind application of (*) leads to a certainty factor of $2\alpha - \alpha^2$ for H. The result is wrong as the set of four rules can be reduced into IF E_0 THEN H with CF = 1, and therefore the certainty factor of H given E_0 should be α.

This example shows the danger of using ad hoc models blindly. The correct way is to build a set of requirements and to build a model that satisfies these requirements as is done in Heckerman (1986).

3. Variety of ignorance.

Ignorance can be subdivided into 3 large categories : incompleteness, imprecision, uncertainty (Bonissone and Tong, 1985).

Incompleteness covers cases where the value of a variable is missing.
Imprecision covers cases where the value of a variable is given but not with the precision required.
Uncertainty covers cases where an agent can construct a personal subjective opinion on a proposition that is not definitively established for him.

The distinctions between these categories are vague and it might be argued that incompleteness is just an extreme form of imprecision (when nothing is known). The following examples illustrate the three forms of ignorance.

Incompleteness. Suppose we have a database that should include marital status and the name of the spouse, and on which the information is error-free. Also suppose the database contains the information "John is married" but the value of the variable "name of spouse" is missing. This information is incomplete since we do not know the name of John's wife but what is available is precise (he is married) and certain (the information is error-free).

Imprecision . Suppose the value of the variable "name of spouse" is "Jill or Joan". This information is complete inasmuch as both marital status and the wife's name are known, but it is imprecise because there is some ambiguity as to whom his wife is. No uncertainty is present here since the information is error-free.

Uncertainty. Suppose the information in the database were provided by some untrustworthy person who said that "John's wife is Jill". The information is complete, precise but uncertain since it might be wrong.

A major difference between these three forms of ignorance is related to the objective or subjective component, as to whether the observer is involved or not. Incompleteness and imprecision are **objective** forms of ignorance. They exist independently of the observer: these properties belong to the data. Uncertainty is a **subjective** form of ignorance. It appears when the observer is taken into account. It is the observer that is not certain about the available information. This information only induces some form of partial knowledge or belief in the observer.

Imprecision and incompleteness are **context dependent**. When I invite guest to my party and I only know "John is married" and his wife is "Jill or Joan", this imprecise information is sufficient if I want to invite only married people, but insufficient if I want to assign seats at the dinner tablein such a way that John will be sitting on the right of his wife. Similarly, if I want to select people whose height is above 150 cm, information like 'Paul's height is >170 cm' or 'Paul is tall' are sufficient, whereas if I want to select only those taller than 175 cm, neither piece of information is sufficient.

Within the three categories of ignorance, one can describe many subcategories. The following table presents the types of model, the types of ignorance and an example for each subcategory.

Incompleteness:		
Combinatory	Existential	John is married, but his wife's name is not given
Combinatory	Universal:	All computer scientists like pizza, but their names are not available.
Imprecision:		
Combinatory	Disjunctive	John's wife is Jill or Joan.
Combinatory	Negation	Jill is not John's wife.
Interval theory	Interval valued information	Paul's height is between 170 and 180.
Fuzzy sets	Fuzzy valued information	Paul is tall.
Possibility Theory	Possibility (physical form)	the possibility for Paul's height to be about 175 cm.
Uncertainty:		
Probability Theory Upper-Lower Probabilities	Probability	the chance of it being "heads" when tossing a coin.
Possibility Theory	Possibility (epistemic form)	the possibility that Paul's height is about 175 cm.
Subjective Probabilities Belief functions	Credibility	my degree of belief that cancer X is due to a virus.

Combinatorial models are not considered here. In practice, they cannot be solved through brute force solutions because of the combinatorial explosion. Default logics and other non monotonic logics have been proposed to solve these problems.

4. IMPRECISION-UNCERTAINTY

Among the numerical models proposed to cope with the various forms of ignorance, the most used are the fuzzy sets theory, the probability theory, the upper and lower probability theory, the possibility theory and the theory of evidence.

4.1. Fuzzy Sets Theory

Fuzziness is the property related to the use of vague predicates like in 'John is tall'. The predicates are vague, fuzzy because the words used to define them are themselves ill defined, vague, fuzzy. The idea is that belonging to a set admits a degree that is not necessarily just 0 or 1 as is the case in classical set theory. Intermediate values are accepted in order to cope with borderline cases. It is different from probability and randomness. Randomness talks about the certainty of whether a given element belongs or not to a well-defined set. Fuzziness talks about the imprecision derived from the partial membership of a given and well defined element to a set whose boundaries are not sharply defined. (Zadeh 1965, 1975, Dubois and Prade 1980)

4.2. Probability theory.

Probability theory is used to quantify the chance that an event might occur or the belief that a proposition is true. Events occur or do not occur, propositions are true or false. No vagueness is involved. Probability theory provides a metalanguage that quantifies the chance that some events might occur or some propositions are true. Its adequacy for random processes has been known for centuries. Its role for decision under risk is well established (Degroot 1970). Bayesians postulate that models like this should be used to quantify one's beliefs, but it is still an open question (Fine 1973).

4.3. Upper and Lower Probability

Among the models proposed to describe degrees of belief, Smith (1961, 1965) and Good (1950, 1983) have postulated that one can often only claim that the probability function that describes our degrees of belief belongs to a convex set P of probability functions. This set can be characterized by the so called upper and lower probabilities, that is the maximal and minimal probability given to each proposition, where the extremes are taken on the probability functions belonging to the convex set P.

A special case of upper and lower probabilities has been described by Dempster (1967, 1968). He assumes the existence of a probability function on a space X and a one to many mapping M from X to Y. Then the lower probability of A in Y is equal to the probability of the largest subset of X such that its image under M is included in A. The upper probability of A in Y is the probability of the largest subset of X such that the images under M of all its elements have a non empty intersection with A.

4.4. Possibility-necessity

Incomplete information such as "John's height is above 170" implies that any height h above 170 is possible and any height equal or below 170 is impossible. This can be represented by a possibility function defined on the height domain whose value is 0 if h < 170 and 1 if h is ≥ 170 (with 0 = impossible and 1 = possible). Ignorance results from the lack of precision, of specificity of the information "above 170" . Its fundamental axiom is that the possibility of the disjunction of two propositions is the maximum of the possibility of the individual propositions. (Zadeh 1978, Dubois and Prade, 1985).

When the predicate is vague like in "John is tall', possibility can admit degrees, the largest the degree, the largest the possibility. But even though possibility is often associated with fuzziness, the fact that non fuzzy (crisp) events can admit different degrees of possibility is shown in the following example. Suppose there is a box in which you try to put tennis balls. You can say: it is possible to put 20 balls in it, impossible to put 30 balls, quite possible to put 24 balls, but not so possible to put 26 balls...These degrees of possibility are degrees of realizability and totally unrelated to any supposedly underlying random process.

Two forms of (continuous valued) possibility have been described: the physical and the epistemic. These 2 forms of possibility can be recognized by their different linguistic uses: it is **possible that** and it is **possible for** (Hacking 1975). When I say it is possible that Paul's height is 170, it means that for all what I know, Paul's height may be 170. When I say it is possible for Paul's height to be 170, it means that physically, Paul's height may be 170. The first form, 'possible that', is related to our state of knowledge and is called epistemic. The second form, 'possible for', deals with actual abilities independently of our knowledge about them. It is a degree of realizability. The distinction is not unrelated to the one between the epistemic concept of probability (called here the credibility) and the aleatory one (called here chance). These forms of possibilities are evidently not independent concepts, but the exact structure of their interrelations is not yet clearly established.

Necessity is defined by the dual property**:** the necessity of a proposition A is the complement (to 1) of the possibility of not-A.

4.5. Credibility: the transferable belief model.

Information can induce in us some subjective, personal credibility (hereafter called belief) that a proposition is true. Its origin lies either in the **random nature** of the underlying event or in the **partial reliability** that we give to the source of information.

In the first case, one ends up with a probability function if one accepts the **frequency principle** (Hacking 1965) that, given the chance that a random event X might occur is p, our degree of belief that it will occur is p.

IF chance(X)=p THEN belief(X) = p

This is fundamental for the **classical bayesian model**, as it relates chance and belief.

When randomness is not involved, there is no necessity that **credal states** (the psychological level where beliefs are entertained) have to be quantified by probability functions. (Levi 1984). The coherence principle advanced by the bayesians to justify probability functions is adequate in a context of decision (Degroot 1970), but it cannot be used when all one wants to describe is a cognitive process. Beliefs can be entertained outside any decision context. In the **transferable belief model** (Smets 1988) we assume that beliefs at the credal level are quantified by belief functions (Shafer 1976). When decisions must be made, our belief at the credal level induces a probability function at the so-called **'pignistic' level** (the level at which decisions are made). This bona fide probability function will be used in order to make decisions using expected utilities theory. Relations between belief functions and pignistic probabilities are given in Smets (1989).

5. Combining models of ignorance.

The various forms of ignorance can be encountered simultaneously and it is necessary that we be able to integrate them. In common-sense reasoning, two forms of ignorance, sometime three, are often encountered in the same statement. Just to give an idea of the problem, consider the following example of generalized modus ponens.

I strongly believe that it is somehow possible that 'If a father is tall, then his son is usually quite tall'.
I believe that it is more or less true that 'Paul is quite tall'.
What can I say about the height of his son.

This example is evidently too complex to be encountered in practice, but it includes most forms of ignorance..

To deal with problems like this, beliefs, possibilities, fuzziness need to be combined, and **a set of metalanguages** must be constructed. Care must be given however to what are the domains of each operator. For instance, probability deals with two domains, the set of propositions (as are usually mentioned) and the truth domain (that is usually disregarded as it contains only two elements, but must be considered once fuzzy propositions are accepted).

The first problem is to see the **connections** between the probability theory in its frequency approach and the physical possibility theory. The next problem is to see the connections between subjective probability functions, belief functions and epistemic possibility functions. Finally, one must establish the connections between the physical properties and the epistemic properties. There is the further the problem of extending all these theories when the propositions involved are fuzzy.

Almost no work has yet been done in this area. But its importance for **datafusion** is obvious: when several sensors provide information, how do we recognize the nature of the ignorance involved and select the appropriate model, how do we collapse them into more compact forms, how do we combine them, how do we take into consideration the redundancies, the correlations and the contradictions. All these problems must be studied and the implementation of potential solutions tested.

Understanding of the meaning of statements and their translation into appropriate models is delicate, if not hazardous. For example, how do we translate "usually bald men are old". Which of P(bald|old) or P(old|bald) is somehow large? "When x shaves himself, usually x does not die". Which conditioning is appropriate: Pl(dead|shaving) or Pl(shaving|dead)? Is it a problem of plausibility or possibility? These examples are just illustrative of the kind of problems that must be addressed.

Bibliography.

BONISSONE P.P. and TONG R.M. (1985) Editorial: reasoning with uncertainty in expert systems. I.J.Man Machine Studies 22:241-250.
DEGROOT M. (1970) *Optimal statistical decisions.* McGraw Hill, New York.
DEMPSTER A.P. (1967) Upper and lower probabilities induced by a multiplevalued mapping. Ann. Math. Statistics 38:325-339.
DEMPSTER A.P. (1968) A generalization of Bayesian inference. J. Roy. Statist. Soc. B.30:205-247.
DUBOIS D. and PRADE H. (1980) *Fuzzy sets and systems: theory and applications.* Academic Press, New York.
DUBOIS D. and PRADE H. (1985) *Theorie des possibilités.* Masson, Paris .
FINE T. (1973) *Theories of probability*. Academic Press, New York.
GOOD I.J. (1950) *Probability and the weighting of evidence*. Hafner.
GOOD I.J. (1983) *Good thinking: the foundations of probability and its applications.* Univ. Minnesota Press, Minneapolis.
HACKING I. (1965) *Logic of statistical inference*. Cambridge UNiversity Press, Cambridge.
HACKING I.(1975) *The emergence of probability.* Cambridge UNiversity Press, Cambridge.
HECKERMAN D. (1986) Probabilistic interpretation for MYCIN's certainty factors, in KANAL L.N. and LEMMER J.F. (eds) Uncertainty in Artificial Intelligence, North Holland, Amsterdam, pp.167-196.
LEVI I. (1984) *Decisions and revisions: philosophical essays on knowledge and value.* Cambridge University Press, Cambridge.
LOPEZ DE MANTARAS R. (1990) Approximate reasoning models. Ellis Horwood, Chichester.
SHAFER G. (1976) *A mathematical theory of evidence*. Princeton Univ. Press, Princeton, NJ.
SHAFER G. and TVERSKY A. (1985) Languages and designs for probability. Cognitive Sc. 9:309-339.
SHORTLIFFE E.H. and BUCHANAN B.G. (1975) A model for inexact reasoning in medicine. Mathematical Biosciences 23:351-379.
SMETS P. (1988) Belief functions. in Non standard logics for automated reasoning, edited by SMETS P., MAMDANI A., DUBOIS D. AND PRADE H., Academic Press, London, pp. 253-286.
SMETS P. (1989) Constructing the pignistic probability function in a context of uncertainty. Proc. Fifth Workshop on Uncertainty in AI, Windsor, Canada, 319-326
SMITH C.A.B. (1961) Consistency in statistical inference and decision. J. Roy. Statist. Soc. B23:1-37.
SMITH C.A.B. (1965) Personal probability and statistical analysis. J. Roy. Statist. Soc. A128, 469-499.
ZADEH L.A. (1965) Fuzzy sets. Inform.Control. 8:338-353.
ZADEH L.A. (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems. 1:3-28.

Acknowledgemeents:
The author is indebted to Yen-Teh Hsia, Robert Kennes and Aleesandro Saffiotti for their help in completing this work.