Representing Partial Ignorance

Didier Dubois* – Henri Prade* – Philippe Smets**[1]

* Institut de Recherche en Informatique de Toulouse (I.R.I.T.) – C.N.R.S.
Université Paul Sabatier, 118 route de Narbonne, 31062 TOULOUSE Cedex – FRANCE

** IRIDIA, Université Libre de Bruxelles, 50 avenue Franklin Roosevelt – CP 194/6
1050 BRUXELLES – BELGIQUE

Ignorance is precious, for once lost it can never be regained.

Abstract

This paper advocates the use of non-purely probabilistic approaches to higher-order uncertainty. One of the major arguments of Bayesian probability proponents is that representing uncertainty is always decision-driven and as a consequence, uncertainty should be represented by probability. Here we argue that representing partial ignorance is *not always* decision-driven. Other reasoning tasks such as belief revision for instance are more naturally carried out at the purely cognitive level. Conceiving knowledge representation and decision-making as separate concerns opens the way to non-purely probabilistic representations of incomplete knowledge. It is pointed out that within a numerical framework, two numbers are needed to account for partial ignorance about events, because on top of truth and falsity, the state of total ignorance must be encoded independently of the number of underlying alternatives. The paper also points out that it is consistent to accept a Bayesian view of decision-making and a non-Bayesian view of knowledge representation because it is possible to map non-probabilistic degrees of belief to betting probabilities when needed. Conditioning rules in non-Bayesian settings are reviewed, and the difference between focusing on a reference class and revising due to the arrival of new information is pointed out. A comparison of Bayesian and non-Bayesian revision modes is discussed on a classical example.

1 - Introduction

Should degrees of belief be represented by probabilities, and more precisely by point-valued Bayesian probabilities ? This question is recurrently debated in the community of Uncertainty in Artificial Intelligence, and the controversy has sometimes led to the publication of very strong opinions about who is right and who is wrong. In this position paper we suggest

[1] This work has been partially funded by the CEC-ESPRIT III Basic Research Project 6156 (DRUMS II).

that the controversy is due to a basic misunderstanding of each camp about the background and basic motivations of the theory advocated by the other camp. Theories are always developed with certain purposes, and these purposes delimit what is to be a problem and what is not to be a problem. Our aim is thus to try and suggest that the problems addressed by non-additive uncertainty models, that is, the representation of and reasoning with incomplete knowledge and partial belief, differ from the problem addressed by subjective Bayesians, that is, decision-making.

The marriage of Artificial Intelligence (AI) with subjective Bayesian probability is quarrelsome (to quote Garbolino (1988)) because they have been initially concerned with very different purposes. Bayesian probability in its subjective form has been developed by Ramsey, De Finetti and Savage (see, e.g., Kyburg and Smokler, 1964), with a view to explain what is a rational decision under uncertainty. To quote Ramsey "the kind of measurement of belief with which probability is concerned is a measurement of belief *qua* basis of action". Since then all authors contributing to the subjective probability literature have emphasized this role of probability, and the betting situation serves as a prototype of action under uncertainty.

On the contrary, the question of making decisions is almost absent from a large part of the AI literature, although the situation is changing to-date. The claimed purpose of AI is knowledge representation and inference rather than decision. Moreover AI contributors have greatly insisted that knowledge should be represented by symbols, not numbers, hence the role played by classical logic and computer languages like LISP in the development of AI. It is striking to notice that even when numbers, such as certainty factors, have been empirically introduced in pattern-directed inference systems, there has been a deliberate caution not to use probability theory. Besides, it is also striking that an expert system like MYCIN (Buchanan and Shortliffe, 1984) whose aim was both to diagnose diseases and advise on medical prescriptions did not use utility theory for the latter task. More generally, the eighties have witnessed the emergence of so-called decision-support systems, that were developed rather independently from decision theory.

It is only recently that AI seems to have realized that making decisions is a particular human activity that is worth considering as a particular subfield of interest, and that more than 50 years of decision theory might be relevant in that respect (e.g., Langlotz and Shortliffe, 1989). Moreover the reluctance of AI with respect to numbers seems to be alleviated, if not vanishing. This reluctance was apparently due to two causes, at least: first the inadequacy of first computer languages like FORTRAN to represent and process knowledge, because these languages aimed at encoding algorithms that solve numerical problems, second, the ad hoc use of numbers in the expert system literature that may have cast some discredit on the very idea of using numbers in AI. It seems that many AI advocates of the symbolic approach have misinterpreted the role of numbers and their nature. Namely, measurement theory (Krantz et al., 1971) explains that numbers are a useful tool for representing and exploiting qualitative structures such as orderings. Utility theory, after Savage (1972), completely adopts this point of view which reconciles the use of numbers with the symbolic description of preference. The last 10 years have led to the recognition of preference relations as a basic tool for modeling plausible reasoning by means of nonmonotonic logics (Shoham, 1988). It can be imagined that the next step will be to consider numerical representations for these orderings.

Despite its apparent misunderstanding of numbers, AI has promoted the idea of modeling various human activities that differ from decision, and especially various types of reasoning tasks. The problem of knowledge representation is at the core of reasoning models. In this position paper we claim that how knowledge should be represented depends upon the reasoning task under concern. Knowledge is almost never perfect, accurate, complete. It is pervaded with imprecision, vagueness and uncertainty. Hence it is natural to introduce these aspects inside the various paradigms of knowledge representation. However, whether or not we should adhere to Bayesian subjective probability for these many purposes is open to debate. By closely relating probabilistic modeling to decisional tasks, and promoting this view inside AI, Bayesian subjectivists implicitly ignore the existence of other kinds of human mental activities that deserve scientific investigations. Our thesis here, is that such tasks as deductive reasoning, plausible reasoning and belief revision under uncertainty require a framework for the representation of incomplete knowledge that is broader than the one offered by Kolmogorov axioms and their betting behavior interpretations. Far from rejecting the existing results on Bayesian decision theory, we suggest that the acceptance of non-additive representations of belief by Bayesian subjectivists is conditioned on their accepting the existence of knowledge representation problems not directly aiming at decision-making.

Section 2 develops this point of view. Section 3 shows that a single probability distribution cannot model total ignorance. Section 4 discusses various uncertainty models including second-order probability in relation to this thesis and Section 5 shows that they all agree with respect to the decision problem. Section 6 envisages belief revision. It reviews various conditioning rules, trying to lay bare their role in specific belief change problems. Section 7 points out a counterintuitive behavior of Bayes rule when it is used to revise betting rates induced by degrees of belief that express a state of partial ignorance.

This position paper elaborates on ideas already advocated by the authors in particular settings such as possibility theory and belief functions (e.g., Dubois and Prade, 1990, Smets, 1991, 1993a, Smets and Kennes, 1994).

2 - Partial Ignorance Versus Uncertainty about Decision

The basic claim that motivates the relevance of non-additive representations of belief is summarized by the following postulate:

Postulate 1: Modeling incomplete knowledge differs from modeling a state of uncertainty about how to act.

Very early Ramsey (1931) touched upon this issue when he distinguished between belief as "intensity of feeling, …, of conviction" and belief "qua basis of action". Although he then focuses on the latter, his dismissal of the first point of view is rather unconvincing. Keeping this distinction in mind we claim that there are at least two kinds of uncertainty that mirror Ramsey's distinction: uncertainty about what is true or false, and uncertainty about how to act. The first kind pertains to a state of (possible absence of) knowledge, the second kind refers to a situation of risk. Especially, in the latter not only are there degrees of belief in the occurrence of events, but there is also the possibility of losing or gaining something, depending on a choice (as patent in the betting behavior setting).

Deductive inference and plausible reasoning differ from decision-making. Deductive inference aims at deriving what can be said about a certain domain of investigation given what is known. Plausible reasoning aims at deriving what is the normal course of things given what is known. Decision-making aims at knowing what to do in the front of a given situation when the state of the world is unknown. Deductive inference and plausible reasoning are question-answering tasks, and should be at work in information systems such as deductive databases. On the contrary, a decision-support system supplies advice.

Deductive inference tells us whether a given proposition is true, false or whether its truth status is unknown, given the current state of knowledge. Plausible inference tells us if a given proposition is normally expected to be true. If knowledge were complete in a certain area of investigation, answering a question of the form "is a given proposition true (in that area)" would be easy on a pure yes-no basis, provided that the proposition is clearly articulated. In the absence of complete knowledge, not all questions can be answered in this simple way. However beliefs can be entertained about propositions whose truth or falsity cannot be firmly established. Here belief is understood according to Cohen (1993) as "a disposition, when attending to issues raised, or items referred to, by the proposition that p, normally to feel it true that p and false that not-p, *whether or not one is willing to act or argue accordingly*" (italics are ours). In the scope of information systems, beliefs will be encoded as a preference relation (that is quantified or not) among propositions of interest, and this preference relation enhances the capability of the systems when answering questions. We shall assume that beliefs depend upon how much knowledge is available and provide guidance about truth-values in deductive or plausible inference.

On the contrary, decision theory helps prescribe what to do not to lose “money”, and is not concerned with computing truth-values.We do not argue against the expected utility theory (DeGroot, 1970) for decision-making here. Whenever decision must be made, we accept here that the decider must assess the probability of the various states of affairs and the utility of the consequence of each act in each state of affairs. We also take it for granted that the probabilities used in computing expected probabilities are genuine probabilities reflecting a betting behavior and that Savage (1972)'s postulates are plausible.We do not consider here questionable aspects of decision theory that are embodied in the Allais, Ellsberg and Newcomb paradoxes (e.g. Gärdenfors and Sahlin, 1988).

Some Bayesians defend that even deductive inference is a form of decision (Levi, 1967). Accordingly 'to deduce that a proposition is true' is equivalent to 'deciding that a proposition is true'. A decision is unsually understood as the free choice of a course of action. Clearly, the truth of a proposition is not the result of a decision, it is a property of the proposition in a given state of the world, and partial belief is not a voluntary act (Cohen, 1993). In practice, we shall assume that partial belief in the truth of a proposition mirrors a given state of knowledge about the world.

Let us focus on deductive inference in an information system. Such a system contains the description of a state of affairs ("the world"). A proposition p will be called "true" if the system can establish that p holds in the state of affairs under concern. The case when the system is not capable of telling whether a proposition is true or not is due to incomplete information about the world, what can be called *partial ignorance*[1]. Hence there are three extreme situations regarding the truth status of a given proposition p: its truth-value t(p) can be true (T), false (F), or unknown (U). The latter is not a truth-value strictly speaking since it only expresses that the value of t(p) cannot be determined. Ultimately this truth-value cannot be but T or F if p is a classical proposition. Note that if a decision is to be made on the basis of the truth-status of p (e.g., if p is true then choose d_1, if p is false choose d_2) then ignorance about p results in uncertainty about how to act. But we may be far from a lack of information while still being in a state of total uncertainty concerning the choice of an action. For instance, after 1000 000 throws of a new die, we are still uncertain about the outcome of the next throw (hence about how to make a decision whose reward depends on this outcome) but we know much more about the die than before the first throw (we -almost- know the values of the probabilities of the possible outcomes). Before the first throw, we had not only uncertainty on the outcome but also on the probability of these outcomes. After the throws, we are still uncertain about the outcome, but no more on the probability of these outcomes. The clear distinction between these two states of knowledge supports the validity of postulate 1. Namely, distinct states of knowledge may lead to the same probability distribution. In the scope of decision-making, it means that distinguishing between these states of knowledge is irrelevant (one should bet in the same way whether one knows that the die is fair or if nothing is known about the die). But if the aim is to represent knowledge for other purposes (for instance, answer queries as in information systems), it might be useful to adopt a theory which is capable of accounting for such a distinction between randomness and ignorance.

3 - Numerical Representations of Total Ignorance

In this section, it is suggested that Bayesian probability is partially inadequate to represent a state of total ignorance. Bayesian probability insists that any epistemic state be represented by a single probability distribution on a suitable set, and that Bayes rule can be systematically relied upon to integrate new information. The suggested failure is directly related to the assumption of a unique probability distribution for representing a state of knowledge.

[1] Here, the situation where one cannot establish the truth or the falsity of p due to limited computational capabilities is ruled out.

How to represent ignorance ? There are two issues. The first one is how to express the contents of a piece of incomplete information. The crudest model is a *disjunctive set* namely a collection of mutually exclusive objects, or situations, or worlds, one of which is the actual one. For instance if we do not know the precise age of the President, we can say that his age x is between 70 and 80 considering the referential set of ages in years. The set {70, 71, …, 80} is disjunctive in the sense that it really means 70 OR 71… OR 80 where an exclusive OR is used, these values are mutually exclusive, since x has a unique value at a given date. Hence incomplete knowledge in its crudest form can be modeled by a disjunction between possible states of the world.

Now the second issue is how to model the truth-status of a proposition p on the basis of incomplete information, by means of some number g(p). Probabilistic tradition suggests the unit interval as a valuation set and the following conventions: g(p) = 0 means that p is certainly false and g(p) = 1 means that p is certainly true. Note that g(p) is not a truth-value but models *certainty of* truth. Indeed we only have that g(p) =0 implies t(p) = F and g(p) =1 implies t(p) = T. Now what about "unknown" ? how to model total ignorance ? Consider a finite set $\mathcal{P}$ of propositions that form a Boolean algebra, built from the language used in the information system. As usually done in logic, we indentify tautologies to T and contradictory propositions to F, such that for any truth-assignment function t, t(T) = T and t(F) = F. In a given state of knowledge the set $\mathcal{P}$ can be partitioned in three subsets $\mathcal{T}$ = {p, t(p) = T}, $\mathcal{F}$ = {p, t(p) = F}, $\mathcal{U}$ gathering the other propositions and being called the uncertainty set. They are respectively the sets of true, false and unknown propositions. The state of knowledge will be called *total ignorance* if and only if $\forall$ p $\neq$ T, F, p lies in $\mathcal{U}$. The certainty function $g_{\emptyset}$ that models this state of knowledge is called a *total ignorance function*. By convention, $g_{\emptyset}$ should attach the same number α to all propositions p in $\mathcal{U}$ so as not to discriminate among propositions in $\mathcal{U}$. Let Ω be the set of possible worlds one of which is the actual world. Any proposition p can be mapped to the subset M(p) of Ω of worlds in which p is true so that M(T) = Ω, M(F) = Ø, and $\mathcal{P}$ is equated to 2^{Ω}. Since total ignorance means $\mathcal{T}$ = {T} it corresponds to viewing Ω as the disjunctive set of possible states of affairs or worlds. By convention the following postulate for the representation of total ignorance looks natural insofar as it expresses that all propositions in $\mathcal{U}$ are equally uncertain:

Postulate 2: The total ignorance function $g_{\emptyset}$ should be such that

$$\forall p, p' \neq F, T, g_{\emptyset}(p) = g_{\emptyset}(p') \neq 0,1.$$

The following consequence of postulate 2 is noticeable:

Corollary: The total ignorance function $g_{\emptyset}$ cannot be represented by a probability measure, as soon as $|\Omega| > 2$.

Proof: Assume $|\Omega| > 2$. Assume there is a probability measure P that is a total ignorance function $g_{\emptyset}$. Let $\omega \in \Omega$, $A \subset B \subset \Omega$ such that $B - A = \{\omega\}$. The existence of distinct subsets such as A and B is ensured by the assumption that $|\Omega| > 2$. Then $P(\{\omega\}) = P(B) - P(A) = \alpha - \alpha = 0$. Hence $\forall \omega \in \Omega$, $P(\{\omega\}) = 0$. Thus P is not a probability measure. Q.E.D.

Actually ignorance can take a much more drastic form, namely one can ignore what are the possible states of the world, i.e., Ω may fail to be exhaustive. Knowledge is expressed by means of a language in which propositions p can be uttered. The language determines the states of the world that are discernible and those which are not. The set of possible worlds where p is true or false need not be precisely known when the truth-status of p is unknown. And Postulate 2 claims that, if nothing is known about the age of somebody, the way to express it should not depend upon whether the considered age scale is [0,100], [0,150] or [0,200] nor on the choice of an epistemically meaningful partition of the scale. Representing ignorance about die throws must include the fact that the number of facets of the die can be unknown as well. Clearly this is hardly the case with a probabilistic representation. In probability theory, one always starts with a set Ω on which an algebra is built, algebra endowed which a probability measure. Probability theory does not consider a permanently changing or atomless algebra.

In Bayesian theory, uninformed priors are usually represented by uniform probability distributions. For instance, in the die case $P(\{i\}) = 1/6$ for $i \in \{1, 2, \ldots, 6\}$. Clearly with this representation $P(\{1\}) < P(\{2, 3, 4, 5, 6\})$ hence it is not genuine ignorance, but a default fair die assumption. However uniform probabilities look natural as representing total uncertainty about how to act with a (6-facet) die in a betting situation. When making a decision we must know what is the set of possible decisions and what is the set of possible situations. As pointed out by Cohen(1993) this is a matter of acceptance, not a requirement that the decision-maker be omniscient. The uniform probability represents the fact that we are not ready, due to ignorance, to put more money on one outcome than on another one, given that the bet is on guessing which facet will be revealed by the throw. Clearly, if the betting situation is described in terms of two alternatives {win, lose} driven by an unknown device producing numbers between 1 and 6, such that you win if the outcome is 1, and loose otherwise, the uninformed prior is P(win) = P(lose) = 1/2. If the decision-maker does not know that this unknown device is a fair die, he has implicitly assumed that $P(\{1\}) = P(not(\{1\}) = P(\{2, 3, 4, 5, 6\}) = 1/2$. In another situation, you might hear that the procedure for getting the numbers is the throw of an octaedron, with the following decision rule for the output: outcome i is reported if the obtained facet is $i < 6$, and is 6 if the obtained facet is 6, 7, 8. The uninformed prior is then $P(\{i\}) = 1/8$ if $i < 6$ and 3/8 if $i \geq 6$. The uninformed prior $P(\{1\}) = 1/2$, $P\{not\{1\}) = 1/2$ is neither consistent with the uninformed prior where it is known that a regular die is used, nor with the case where it is known that the octaedron is used with the above output. This inconsistency makes no difficulty with the betting problems since in a given betting problem the situation is clearly defined. However this inconsistency is not permitted in a knowledge representation problem where several partitions of the world correspond to different points of view that must coexist in an information system. The deduction of the truth of a proposition and the belief allocated to it should not depend on the way worlds are defined. The representation of total ignorance should not depend on the existence of the die or the octaedron, and should leave room for these possibilities as they become known. Bayesian probability was never tailored for taking into account this situation. It can cope with the canceling of possibilities (through renormalization of the probability distribution function) but it has not been made for the case when a new possibility emerges (e.g., suddenly the die breaks), or when the granularity of the representation changes (what was thought as a single outcome is made of several distinct mutually exclusive situations).

4 - Various Models of Partial Ignorance

In this section several non-Bayesian approaches to uncertainty are reviewed so as to assess their capability at representing partial ignorance.

4.1. Two numbers are needed for the quantification of partial ignorance

The previous section has modeled partial ignorance in a very crude way by attaching to any proposition p representing a subset $M(p) \subseteq \Omega$ of possible worlds a tag mentioning "true" (T), "false" (F) or "unknown" (U), and defined the corresponding sets $\mathcal{T}$, $\mathcal{F}$, $\mathcal{U}$ of propositions that partition $\mathcal{P}$. It is easy to check that if $p \in \mathcal{T}$ then $\neg p \notin \mathcal{T}$, otherwise we face an inconsistency, more specifically $t(p) = T \Leftrightarrow t(\neg p) = F$ and $p \in \mathcal{U} \Leftrightarrow \neg p \in \mathcal{U}$. Now $\mathcal{T}$ can be viewed as the set of *certainly* true propositions. Similarly let $\mathcal{F} = \{p \mid t(p) = F\} = \{\neg p \mid p \in \mathcal{T}\}$can be viewed as the set of *certainly* false propositions. The set $\mathcal{T} \cup \mathcal{U}$ contains the propositions which are either certainly true or unknown, i.e., the propositions which are *possibly* true. Let $E \subseteq \Omega$ be the set of worlds where all propositions in $\mathcal{T}$ are true, then partial ignorance is expressed by the disjunction of worlds in E since the actual world lies in E. Complete knowledge is then when E reduces to a singleton of Ω. It corresponds to the case where the set $\mathcal{U}$ is empty (no uncertainty).

For representing the certainty of propositions in the state of partial ignorance, one might try to keep the convention based on mapping T to 1, F to 0, and U to some $\alpha \in [0,1]$, and use a single function g, called a *partial ignorance function*. In order to obtain some nicely behaved function, one would like it to be compositional with respect to classical logic connectives, that is, g should ideally behave as a truth assignment function. This means especially that $g(p \vee q)$, $g(p \wedge q)$ are determined by g(p) and g(q) only, and $g(\neg p)$ is determined by g(p) only. This is maybe too ambitious. Indeed this is not true for probability measures, since while $P(\neg p)$ is indeed determined by P(p) only, $P(p \vee q)$ is the sum of P(p) and P(q) only if $p \wedge q = F$. Hence a partial ignorance function g that satisfies $g(p \vee q) = D(g(p), g(q))$ and $g(p) = n(g(\neg p))$ whenever $p \wedge q = F$ might be searched for the following rules:

Proposition: A partial ignorance function g cannot be compositional with respect to conjunction nor disjunction, but is compositional with respect to negation only. Moreover g is not decomposable with respect to union, i.e., $g(p \vee q)$ is not a function of g(p), g(q) even if $p \wedge q = F$.

Proof: Function n such that $g(p) = n(g(\neg p))$ can be defined by: $g(p) = 0 \Leftrightarrow g(\neg p) = 1$, and $g(p) = \alpha \Leftrightarrow g(\neg p) = \alpha$, i.e., g is compositional for negation. Let p be such that $g(p) = \alpha$ and some operation I be such that $g(p \wedge q) = I(g(p),g(q))$, $\forall p,q$, assuming that g is compositional for conjunction. Then if q = p with g(p) = α we have $I(\alpha,\alpha) = \alpha$ and if q = ¬p with g(p) = α we have $I(\alpha,\alpha) = g(\neg p \wedge p) = 0$. Hence $I(\alpha,\alpha) = 0$ which contradicts the compositionality assumption. A similar reasoning applies to disjunction. Namely, let q = ¬p, then $g(p \vee \neg p) = 1$ even if $g(p) = \alpha = g(\neg p)$, while $g(p \vee p) = \alpha$, hence there is no way to find a function D such that $g(p \vee q) = D(g(p),g(q))$. The last statement of the proposition is also established this way since $p \wedge \neg p = F$. Q.E.D.

A consequence of the lack of compositionality and even decomposability of function g is that it will be very cumbersome to use. Especially the strength of probability theory is that a probability assignment to elements of Ω is enough to represent a probability measure on 2^{Ω}. The above result states that this is not possible for partial ignorance functions that attach a single number to each proposition.

For the sake of numerically representing partial ignorance in a more efficient way one might be tempted to use one numerical scale for certainty of truth (testing whether $p \in \mathcal{T}$ or not) and another one for possibility of truth (testing whether $p \in \mathcal{T} \cup \mathcal{U}$ or not).The idea is that since we have three distinct alternatives, we need two bits to encode them. Let us consider the following convention, involving a function SN describing the amount of necessary support and a function SΠ describing potential support, and mapping $\mathcal{P} = 2^{\Omega}$ to the set {0,1} and defined by

SN(p) = 1 if $p \in \mathcal{T}$ and 0 otherwise
SΠ(p) = 1 if $p \in \mathcal{T} \cup \mathcal{U}$ and 0 otherwise[1].

SN(p) = 1 means that p is fully supported by the available information, i.e., p is necessarily or certainly true. SΠ(p) = 1 only means that p might prove to become true if further information is obtained (but it might turn out to be false as well). SN(p) and SΠ(p) are special cases of degrees of necessity and possibility respectively and denoted N and Π in possibility theory (Dubois and Prade, 1988a). It is easy to check that SN(p) = 1 ⇒ SΠ(p) = 1, and that $p \in \mathcal{U}$ if and only if SN(p) = 0 and SΠ(p) = 1. Hence the truth-status about p is expressed by the pair (SN(p),SΠ(p)), (1,1) represents (certainty of) truth, (0,0) (certainty of) falsity and (0,1) ignorance. Clearly these conventions do not depend on the number of elements in Ω.

We can carry the compositionality of truth assignments over to the pair (SN,SΠ) *as a whole* in the presence of incomplete information. Indeed it is easy to check that:

$$SN(p \wedge q) = \min(SN(p),SN(q)) \;,\; S\Pi(p \vee q) = \max(S\Pi(p),S\Pi(q)) \;,\; SN(p) = 1 - S\Pi(\neg p). \quad (1)$$

---

[1] By convention, the notation SN and SΠ is used for the binary concepts of necessary and potential support as well as their generalization. So SN (resp.: SΠ) will denote necessity (possibility) in modal logic, necessity (resp.: possibility) measures, lower (resp.: upper) probabilities functions, belief (resp.: plausibility) functions... depending on the context. We keep the same notation in order to stress the coherence between the various concepts underlying the particular measures.

Note that SN is not compositional for the disjunction nor is SΠ compositional for the conjunction. The disjunctive encoding of the state of incomplete knowledge by means of the set $E = \{\omega \mid p \text{ is true in } \omega, \forall p \in \mathcal{T}\}$ is sufficient to derive both SΠ and SN functions since:

$$S\Pi(p) = 1 \Leftrightarrow M(p) \cap E \neq \emptyset$$
$$SN(p) = 1 \Leftrightarrow E \subseteq M(p).$$

where M(p) is the set of worlds of Ω in which p is true.

The membership function of the disjunctive set E is called a possibility distribution. The above encoding of partial ignorance is in agreement with Zadeh's possibility theory (Zadeh, 1978, Dubois and Prade, 1988a). It is easy to show that SN = SΠ if and only if E reduces to a singleton of Ω and that in this situation (where the uncertainty set is empty), SN coincides with a truth assignment. This is the state of complete knowledge.

Note that one way of recovering a partial ignorance function g is to let $g(p) = \frac{SN(p) + S\Pi(p)}{2}$ (then $\alpha = 1/2$). The certainty factor of MYCIN is also a partial ignorance function on [–1,+1] such that F maps to –1, T to 1, and U to 0. Such a certainty factor is obtained if we let $CF(p) = S\Pi(p) - S\Pi(\neg p)$ (see Dubois and Prade, 1988b).

4.2 Ordinal representations of partial ignorance.

The above model of partial ignorance, based on (SN,SΠ) functions is not very expressive compared to probability theory since the ordering of propositions is crude. The model can be enriched first by distinguishing propositions which are more or less certain and more or less possible. This can be done by introducing intermediary levels between 0 and 1, the extreme case being to turn {0,1} into [0,1]. In that case we recover, insofar as we keep equations (1), Zadeh's possibility theory. The membership function π of the set E becomes the one of a fuzzy set, albeit a disjunctive one since π(ω) expresses to what extent it is possible that ω be the actual world, π(ω) = 0 corresponding to complete impossibility.

At that point the model is not genuinely numerical since (1) only requires a linearly ordered scale $\{\pi_0 < \pi_1 < \ldots < \pi_n\}$ as the range of SΠ (with $\pi_0 = 0$, $\pi_n = 1$) and another one $\{\eta_0, \ldots, \eta_n\}$ as the range of SN, the duality between possibility and certainty being expressed by $SN(p) = \eta_i \Leftrightarrow S\Pi(\neg p) = \pi_{n-i}$. [0,1] is thus used as an ordinal scale. Possibility theory is the basic tool for uncertainty modeling based on ordinal preference on what is possible and what is not. Starting with a set $E \subseteq \Omega$ containing the actual world, and a partitioning $E_1 \cup \ldots \cup E_n$ of E such that $\forall \omega,\omega' \in \Omega$, $\omega \in E_i$, $\omega' \in E_j$, $i > j$ means that ω is considered more possible than ω', we can build a possibility distribution π on Ω such that $\pi(\omega) = \pi_i$ if $\omega \in E_i$, $\pi(\omega) = \pi_0$ if $\omega \notin E$ and build possibility and necessity functions SΠ and SN describing our knowledge about any proposition p as follows:

$$S\Pi(p) = \max\{\pi(\omega), \omega \in M(p)\}. \quad (2)$$

for instance we may refine our knowledge about the age of the President by rank-ordering the set {70, 71, …, 80} in terms of values that are unsurprizing and values that are more surprizing (using the terminology of Shackle (1961)), thus determining a possibility distribution π. Note that this ordinal setting is not only the one of possibility theory but also belongs to a class of nonmonotonic logics that are called "preferential" by Shoham (1988) and "rational" by Lehmann and Magidor (1992), who carefully study the notion of "ranked models". It is also closely related to so-called well ordered partitions considered by Spohn (1988) and to the theory of potential surprize of Shackle(1961).

4.3 Belief functions

A higher level of sophistication is obtained when the propositions are quantitatively ordered by means of numerical (SN, SΠ) pairs, for which the concepts of addition and multiplication are meaningful, but only the duality property $SN(p) = 1 - S\Pi(\neg p)$ holds in equations (1). For instance, we may consider a probabilistic structure on top of a disjunctive representation of partial ignorance, like in the so-called Dempster-Shafer theory (Dempster, 1967). Suppose we know that the available knowledge can be described as a disjunctive set E which is itself unknown. E can be any one of the family $\{E_1, \ldots, E_k\}$ of non-empty subsets of Ω. Let $m(E_i) = m_i$ be the probability that indeed our partial ignorance is represented by E_i, so that $\sum_{i=1,k} m_i = 1$. We leave it open whether $m(E_i)$ is a frequency-driven probability, a subjective probability or whatever. We just assume that a probability distribution on $\{E_1, \ldots, E_k\}$ exists. Computation of functions of potential and necessary support can then be based on expectations (Zadeh, 1979), namely:

$$S\Pi(p) = \mathrm{Exp}(S\Pi_E(p)) = \sum_{i=1,k} m_i . S\Pi_i(p) \quad (3)$$

$$SN(p) = \mathrm{Exp}(SN_E(p)) = \sum_{i=1,k} m_i . SN_i(p) \quad (4)$$

where $S\Pi_E$ and SN_E are the random possibility and necessity functions that depend on E and $S\Pi_i$, SN_i are the {0,1}-valued possibility and certainty functions based on E_i. It is easy to verify that SΠ and SN are exactly plausibility and belief functions (Shafer, 1976) respectively, and that they obey the duality axiom $SN(p) = 1 - S\Pi(\neg p)$. SN is usually denoted Bel and SP is denoted Pl by Shafer and his followers. These set-functions satisfy all the properties in (1) if and only if the family $\{E_1, \ldots, E_k\}$ of focal elements is nested (consonance case). Moreover if we rule out partial ignorance, i.e., the sets $E_1, \ldots, E_k$ never contain more than one possible world, the probability structure is located right on Ω and $SN(p) = S\Pi(p) = P(p), \forall p$. Potential and necessary support evaluations collapse to probability functions in that case. The mathematical property stating that the sets $E_1, \ldots, E_k$ never contain more than one possible world in a probabilistic model, gives a precise meaning to the claim that probability measures do not account for partial ignorance, since each E_i then expresses complete knowledge.

Other interpretations of the Dempster-Shafer model exist. For instance, the transferable belief model (Smets and Kennes, 1994) corresponds to assuming that degrees of certainty SN are quantified right away by Shafer(1976)'s belief functions. Bayesians claim that beliefs should be additive, their arguments being essentially based on rational decision-making behavior. The transferable belief model exists without reference to any decision and therefore does not postulate the additivity axiom. Owing to its flexibility, it copes easily with partial ignorance. However it does not rule out the case of purely probabilistic frequency-based beliefs.

4.4 Upper and lower probabilities and beyond.

It is interesting to consider the converse situation where partial ignorance is put on top of a probability structure, namely a partially unknown probability structure. Let P be a probability measure which is partially unknown, i.e., P is one of the probability measures in a set $\mathbb{P}$ of mutually exclusive ones. Hence we have an all-or-nothing possibility distribution π on $\mathbb{P}$ such that $\pi(P) = 1$ if $P \in \mathbb{P}$ and 0 if $P \notin \mathbb{P}$. Now P(p) is usually ill-known and can be any number in $\{P'(p), P' \in \mathbb{P}\}$, and we can define upper and lower probabilities $S\Pi(p) = \sup_{P' \in \mathbb{P}} P'(p)$, $SN(p) = \inf_{P' \in \mathbb{P}} P'(p)$ which are set functions more general than plausibility and belief functions (e.g., Kyburg, 1987).

We may also introduce some preferential ordering, encoded by a fuzzy set membership function, each time we have incomplete information expressed by means of a subset. Thus, we can extend Dempster-Shafer theory to a probability distribution on disjunctive *fuzzy* sets (see Yager, 1982, Yen, 1990). Another worth mentioning extension is when we have some preference ordering among the probability measures in the set $\mathbb{P}$ introduced above. A fuzzy set $\tilde{\mathbb{P}}$ can naturally be induced by considering the family of probability measures obeying some elastic constraints such that, "the more or less possible values of P(p) are represented by the fuzzy subset M of the real line", i.e., $\mu_M(r)$ is the possibility that $P(p) = r$. This captures the idea of ill-known probability values (Zadeh, 1984).

A still more elaborate uncertainty model can be obtained by considering the existence of a probability measure on $\mathbb{P}$, therefore a probability of probability, what leads to higher order probabilities. Such meta-probability theory suffers nevertheless from some weaknesses when $\mathbb{P}$ is a family of objective probability functions.[1] The subjective probability P(A) of an event A is usually defined as the *price* an agent is willing to pay to play a game where he/she receives $1 from the banker if the event A occurs and nothing if the event A does not occur. Furthermore the bet is *fair* if the agent is indifferent between being the player or the banker once the price has been fixed. Such a procedure leads to the construction of probabilities that satisfy Kolmogoroff's axioms for probability measures and provides a nice semantics to the concept of subjective probability. This definition is based on the decidability of the underlying bet. When it comes to meta-subjective-probabilities over objective probabilities, such decidability cannot be achieved if the objective probability is defined as the limit of a proportion. Thus the semantics based on fair bets cannot be used to define meta-probabilities, a concept that therefore becomes

[1] De Finetti solved that problem by claiming that objective probabilities do not exist.

'undefined'. Other authors suggest that meta-probabilities are nothing but conditional probabilities (see Kyburg (1989) for a detailed discussion), so that in some sense the higher level disappears.

Assumption	Type of construction	Theory
incomplete information on the set of situations	disjunctive sets of situations error intervals	propositional calculus
ordinal knowledge on the set of situations	disjunctive sets of situations equipped with a complete preorder	possibility theory preferential logics
numerical betting rates or frequencies on the set of situations	a probability distribution over the set of situations	probability theory
probabilities on a set which is related with the set of situations	a probability distribution on disjunctive sets	Dempster-Shafer theory
numerical non-additive degrees of belief on the set of situations	a belief function	transferable belief model
imprecisely known probability distribution on the set of situations	convex set of probabilities	upper and lower probabilities
numerical additive degrees of belief on the set of possible probabilities	probability over a set of probabilities	higher order probabilities

Table 1. Partial ignorance models.

Metaprobability might also be suggested to quantify introspective probabilities about one’s own subjective probabilities. It would quantify our doubts about our own beliefs. Classical modal logics for beliefs (the KD45 model, see Chellas(1980))reject such metabeliefs. In these theories, for instance Hintikka(1962)'s epistemic logic, the functions SN and SP are represented by modal symbols for necessity and possibility. The existence of introspective metaprobabilities about one’s own probabilities can be questionned. It surely contradicts the idea that an agent has always a belief about every event.

The above review, also pictured in Table 1, covers most of the non-additive probability models available to-date. Especially, second order probabilities appear to formally contrast with these models since they are probability structures on top of probability structures (see Table 1).

5 Decision Under Partial Ignorance

How to make a decision in the face of partial ignorance ? If we follow the Bayesian approach, and we start from a set of possible worlds Ω, we must

A) Partition Ω into a family of subsets $\theta_1, \ldots, \theta_n$ that forms a frame of discernment for the decision problem (the betting frame), i.e., one accepts to bet on some element of $\Theta = \{\theta_1, \ldots, \theta_n\}$.

B) Define a probability function P on Θ that agrees with our representation of partial ignorance.

C) Compute the expected utility of any act a depending on Θ as $U(a) = \sum P(\theta_i)u_i(a)$ where $u_i(a)$ is the utility of the act a when the actual world is in θ_i.

This procedure corresponds to the expected utility theory for decision-making and is applicable whenever the frame, the probability P and the utility function u are given. Alternative schemes that refrain from selecting a probability function P but use the Choquet integral applied to the lower probability function SN (Schmeidler, 1989) or generalize the De Finetti consistency argument to belief functions (e.g., Jaffray, 1988) will not be discussed here insofar as they violate Savage postulates of decision. The above selection procedure of a betting probability that agrees with a given state of partial ignorance does not apply to qualitative, ordinal representations such as possibility theory where only a plausibility ordering of worlds is available. Nevertheless, Boutilier (1993) recently suggested to put probabilities on top of possibilities. Namely given the available knowledge, select the most plausible worlds according to our (non-probabilistic) beliefs, and build a betting probability on this set.

Suppose now that we do not know exactly the values of $P(\theta_i)$, but that we accept to describe our knowledge by a second order probability on Θ. Let PP denotes this second order probability. It is well-known (e.g., Kyburg, 1989) that the resulting utility function is:

$$U(a) = \sum_i E(P(\theta_i))\, u_i(a) \tag{5}$$

where $E(P(\theta_i))$ is the expected probability of $P(\theta_i)$ with respect to the second order probability PP. This result is well accepted among Bayesians.

Now assume we start with partial ignorance, and we have chosen the betting frame Θ. In the most general case we have partial ignorance about the probability distribution. We only know that the actual probability measure P belongs to a set $\mathbb{P}$ of possible probabilities. The requirement that the chosen probability measure agrees with our representation of ignorance means that we should select P in $\mathbb{P}$ so that $P(p) \leq S\Pi(p)$, which is here equal to the upper probability. This is a version of Zadeh's consistency principle that claims that what is probable must be possible (e.g., Dubois, Prade and Sandri, 1993).

Starting with a set $\mathbb{P}$ and no special knowledge on which elements of $\mathbb{P}$ corresponds to the actual probability measure, the Bayesian approach usually leads to considering a uniform second order probability distribution over $\mathbb{P}$ and to applying the second order utility formula (5). In other words, $\mathbb{P}$ is represented by its center of gravity $\bar{P}$.

Note that the maximum entropy method for selecting $\bar{P}$ in $\mathbb{P}$ is not suitable here. For instance consider a coin problem where {head, tail} is the referential set and assume that it is known that SN(head) = 0.2, SN(tail) = 0.4. It indicates that "tail" is rather preferred to "head" as the most plausible outcome. The set of probabilities corresponding to this information can be represented by $x = P(\text{head}) \in [0.2,0.6]$ and $1 - x = P(\text{tail}) \in [0.4,0.8]$. The probability measure which has maximal entropy in this set is the uniform probability on {head, tail}, which throws away the information on the preference for tail. However the center of gravity $\bar{P}(\text{head}) = 0.4$, $\bar{P}(\text{tail}) = 0.6$ preserves this information.

In the case of belief functions, we have a probability distribution over pieces of incomplete evidence E_i . They induce upper and lower probabilities on Θ. Smets (1990) has advocated and justified by some rationality requirements the use of a "pignistic transformation", a transformation that had already been intuitively suggested by Williams (1982), and Dubois and Prade (1982). Namely if the belief function is represented by $\{(E_j,m_j), j = 1,k\}$ where the E_j's are the focal elements and the m_j's are the basic belief masses (basic probability masses), the selected probability P', called the pignistic probability, is such that:

$$\forall i,\ P'(\theta_i) = \Sigma_{\theta_i \in F_j} \frac{m_j}{|F_j|}.$$

If SΠ is the plausibility (expected potential support) function induced by $\{(E_j,m_j), j = 1,k\}$ and defining $\mathbb{P} = \{P: P \leq S\Pi\}$ then it can be proved (see, e.g., Kruse et al., 1991, Dubois et al., 1993) that the pignistic probability coincides with the center of gravity of $\mathbb{P}$. Note that the pignistic transformation, even though justified by rationality requirements outside any probabilistic context, comes down to being mathematically equivalent to changing each piece of partial ignorance E_i into a uniform probability P_i on E_i representing a Bayesian uninformed prior, and P' is then the convex mixture of the P_i's with weights m_i. Hence the pignistic transformation approach to decision with belief functions is consistent with the higher-order probability approach, viewing the weights m_i as meta-probabilities. It is thus a generalized indifference Laplacean principle. Note that the same approach still applies when the belief function is consonant, and satisfies the decomposability property (1). Hence numerical versions of possibility theory are liable of the same treatment of the decision problem, although the axiomatic justification of the pignistic transformation proposed by Smets and Kennes (1994) does not apply to the subclass of consonant belief functions alone.

This property should not be misunderstood when applied to the transferable belief model. In that case, the initial concept is the belief function that quantifies our beliefs over Ω. When a decision must be made on the betting frame Θ, one computes the pignistic probability P' and uses it to compute expected utilities. It turns out that $SN(A) = Bel(A) \leq P'(A) \leq Pl(A) =$

$S\Pi(A)$ $\forall A$ subset of Θ. If one then considers all possible betting frames Θ_i derived from Ω and the set of the pignistic probabilities P'_i derived on each frame Θ_i, one has (Wilson, 1993):

$$SN(p) = \min_i P'_i(p) \quad \text{and } S\Pi(p) = \max_i P'_i(p) \qquad \forall p \in \mathcal{P}$$

In the transferable belief model, SN = Bel is not the lower bound of a family of probability measures on Ω that satisfy some particular constraints (as in Dempster-Shafer theory). SN directly quantifies the degrees of belief and turns out to be the lower bound of the pignistic probabilities that can be built on the betting frames related to Ω. The lower bound concept is a derived property, not an essential one as in Dempster-Shafer theory.

Nonadditive theories of uncertainty are not in total disagreement with probability theory. Owing to the pignistic transformation, they can even remain coherent with Bayesian decision theory. It is important to stress that the probability measure at the pignistic level does not represent our state of partial ignorance, it only *results* from it. It is the representation of uncertainty needed for rational decision-making and justified by Savage's axioms. More specifically the use of an additive measure at the pignistic (betting) level protects the user from a synchronic Dutch Book. The avoidance of diachronic Dutch Book (Teller, 1973, Jeffrey, 1988) is also obtained with the pignistic probability generated in the transferable belief model, but the procedure is more tricky than the one described here, since the information about a relevant forthcoming but yet unknown experimental result must be included in the model (Smets, 1993b).

The above results also indicate that all numerical representations of partial ignorance are completely compatible with the Bayesian second order probability approach in the decision problem. This fact may be used as an argument against non-Bayesian formalisms, as being useless, if it is the case that there is no uncertainty management problem but decision. But the Bayesian representation of ignorance leads to results that differ from those obtained by other, less biased representations, for problems such as belief revision, as seen in the next sections.

Additional remarks

1) The pignistic transformation can also be generalized to the case when the E_j's are disjunctive fuzzy sets, that is, to belief functions with fuzzy focal elements. Then we turn each numerical possibility distribution $\pi_j = \mu_{E_j}$ into a probability measure P_j such that p_{ij} is the probability of θ_i induced by distribution π_j, and the pignistic transformation is then generalized to fuzzy focal elements using

$$\forall i,\ P'(\theta_i) = \Sigma_{\theta_j \in \text{Support}(E_j)}\, p_{ij} \cdot m_j.$$

3) In the case of fuzzily-known or interval-valued probabilities we can also compute the possibility distribution restricting the possible values of an expected utility given the fuzzy set of possible values of the $P(\theta_i)$'s and/or of the $u(a_i)$'s (e.g., Dubois and Prade, 1988c). In this case, the acts a_i cannot be linearly ordered in general (because we only have interval or fuzzy

interval values for the U(a_i)'s) and we cannot always decide what is the best act to perform, but it enables the user to see if the partial ignorance pervading the probability and/or utility values prevents or not from safely isolating a best act.

6. Focusing versus Revision.

There are two kinds of information that may convey some form of partial ignorance: factual evidence and generic knowledge. Factual evidence consists of information gathered on a case at hand, or the description of the actual world in a given situation. This information can be more or less precise and more or less reliable. Generic knowledge pertains to a class of situations considered as a whole, but does not refer to a particular case. Sometimes this class of situations is well defined (it is a population in the sense of statistics) and the generic knowledge is of a frequentist nature. Sometimes the relevant class of population is much more vaguely described (as in the famous "birds fly" example), and the generic knowledge describes rules of plausible inference of the form "if all I know is A then plausibly conclude B". The levels of confidence expressed by the rules can be numerically modelled via a conditional probability, or be handled in a purely ordinal setting (as in default reasoning). The difference between generic knowledge and factual evidence can be illustrated by a diagnosis problem. The generic knowledge of a clinician consists in his knowledge about the links between the diseases and the symptoms and the distribution of the diseases in the population (in practice, the likelihoods and the prior probabilities). The factual evidence consists in the symptoms collected from the patient under consideration.

Focusing and revison are two distinct forms of conditioning. The difference between focusing and revision is most clear in settings where generic knowledge is represented and processed as distinct from factual evidence. Such a distinction is not relevant in propositional logic for instance, since every piece of information takes the form of a propositional sentence. Focusing consists in conditioning the generic knowledge by the factual evidence, i.e., in the example, changing the reference class in order to focus on those cases that share the same symptoms as those of the patient. In probability theory, focusing is achieved by applying the Bayes rule of conditioning.

Revision consists in either conditioning the generic knowledge by another piece of generic knowledge (*G-revision*, in the following), or the factual evidence by another piece of factual evidence (*F-revision*). G-revision consists in revising the clinician knowledge about the relation on the disease-symptom space by learning new medical knowledge. F-revision consists in conditioning the clinician’s beliefs about which symptoms prevail for the patient by learning new information about the symptoms.

In the setting of Bayesian probability the distinction is conceptually meaningful but focusing and revising are both expressed by the same conditioning rule, that is Bayes rule. Hence some disputes among probabilists as to the meaning of conditioning. The most widely found view is that Bayes rule operates a change of reference class, namely going from a prior to a posterior probability is not a revision process. Posterior probabilities are precalculated and the input information just prompts the selection of a particular posterior. With this view, the prior probability together with the likelihoods determine a unique joint probability over a space of interest, construed as a body of generic knowledge (the clinician's experience on a certain disease), and conditioning means integrating factual evidence so as to configurate the generic knowledge properly with respect to the reference class of the object on which this factual evidence bears (test results on the patient). This point of view, which is shared by the expert system literature, culminates with the advent of Bayesian networks.

The revision view of conditioning is typically advocated by people working in probability kinematics. Philosophers like Jeffrey (1983), Domotor (1985), Williams (1980), and Gärdenfors (1988), understand Bayes rule as a revision process, by which a prior probability P is changed into a new one P' due to the input information A. This latter view is supported by maximal cross-entropy arguments whereby it is established that Bayes rule obeys a minimal change requirement. This school of thought is called probability kinematics: the input information A is understood as the discovery that $P'(A) = 1$ while the prior is $P(A) < 1$, and must be modified. The possibility that the input information is itself probabilistic is commonly envisaged in probability kinematics: the input information enforces the probability of some event to take on a value which differs from the prior value. In this view the input information is at the same level as the prior information: both are (possibly uncertain) factual evidence or pieces of generic knowledge. For instance the "observation by candlelight" example proposed by Jeffrey involves prior factual beliefs about a particular piece of cloth which are changed upon observing again the cloth with a candle, i.e., a new piece of uncertain factual evidence, this is F-revision. Revision of generic knowledge would correspond to the following case: a probability distribution over sizes of adults in a given country is available, and some input information comes in that nobody in the population is more than 6ft tall.

7 Implementing Revision and Focusing under Partial Ignorance

Suppose we happen to learn that the variable x the actual value of which is partially ignored in Ω has been observed to belong to some subset A. How do we change our belief about x ?

7.1 Propositional and ordinal settings

If the state of partial ignorance is described by a disjunctive set $E \subseteq \Omega$ containing x then the resulting state of knowledge is simply $x \in A \cap E$. Note that this revision does not work if $A \cap E = \emptyset$, which should not happen insofar as the piece of information $x \in E$ is absolutely sure. The language used here, namely the one of propositional logic, is not rich enough to express the concept of focusing as distinct from revision. Namely the treatment necessary to handle a query pertaining to x, to a database containing the piece of information $x \in E$, given that $x \in A$, is the same as when the statement $x \in A$ is to be added to the database: one must compute $A \cap E$ when not empty. This is what Gardenförs(1988) calls "*expansion*".

The revision problem in the ordinal approach to possibility theory is treated in Dubois and Prade (1992b). Namely, each subset B of Ω is changed into $B \cap A$, and the level of possibility $\Pi(B)$ is carried over to $B \cap A$ if $B \cap A \neq \emptyset$. Finally, all subsets $C \subseteq \Omega$ such that $S\Pi(C) = \max\{S\Pi(B \cap A), B \subseteq \Omega\}$ are assigned a maximal level of possibility 1. It corresponds to a notion of conditional possibility $S\Pi(B \mid A)$ defined in (Dubois and Prade, 1988a) as:

$$S\Pi(B \mid A) = \sup \{a \in [0,1], S\Pi(A \cap B) = \min(a, S\Pi(A))\}.$$

This type of ordinal conditioning has been also recently studied by Williams (1994) under the name "adjustment". Recalling that in the ordinal case, the partial ignorance function pair $(SN, S\Pi)$ is equivalent to a subset E of possible worlds equipped with a complete ordering relation expressing plausibility, revision by subset A comes down to restricting (without changing) this ordering to $A \cap E$. Just as in the case of Bayes conditioning the ordering of elements in $A \cap E$ is left unchanged. This revision satisfies all of Gärdenfors (1988) revision postulates, and the SN function coincides with what Gärdenfors calls an "*epistemic entrenchment*" (see Dubois and Prade, 1991). Again, as in the case of propositional logic, the language of possibility theory is not capable, when a single possibility distribution is used and fixed, of accounting for the notion of focusing as distinct from revision.

The question of representing generic knowledge in possibility theory can be addressed if possibility theory is related to conditional theories of default reasoning (Kraus, Lehmann, Magidor, 1990, Lehmann and Magidor, 1992).A body of generic knowledge Δ is encoded as a set of default rules $p \rightarrow q$, each being interpreted as a nonmonotonic inference rule " if p is true then it is normal that q be true". The arrow $\rightarrow$ is nonclassical. Δ is also called a conditional knowledge base. As proved in Benferhat et al. (1992), it is natural to interpret a default rule $p \rightarrow q$, as a constraint on possibility measures of the form $S\Pi(p \wedge q) > S\Pi(p \wedge \neg q)$, i.e., such a default rule can be viewed as expressing that $p \wedge q$ is more normal than $p \wedge \neg q$. A set of default rules $\Delta = \{p_i \rightarrow q_i, i = 1,n\}$ with noncontradictory condition parts can be viewed as a family of constraints

$$C(\Delta) = \{S\Pi(p_i \wedge q_i) > S\Pi(p_i \wedge \neg q_i), i = 1,n\}$$

restricting a family $\Pi(\Delta) = \{S\Pi, S\Pi \in C(\Delta), S\Pi(p_i) > 0\}$ for all i of possibility distributions over the interpretations of a language. Given a piece of evidence encoded as a propositional sentence p and the conditional knowledge base Δ, the sentence q is a plausible consequence of (Δ,p) iff the rule $p \rightarrow q$ can be deduced from Δ. This inference has a precise meaning in the framework of possibility theory namely the (classical) inference of the srict inequality $S\Pi(p \wedge q) > S\Pi(p \wedge \neg q)$ for all $S\Pi$ in $\Pi(\Delta)$ derived from the set of constraints $C(\Delta)$ (Dubois and Prade, 1994b). This inference also perfectly fits with the so-called "preferential inference system" of Kraus et al.(1990) and with the logic of conditional objects (Dubois and Prade, 1993, 1994b). Another more productive type of inference is the socalled "rational closure"(Lehmann and Magidor, 1992). It can be captured in the possibilistic setting by selecting a particular possibility measure SP* in $\Pi(\Delta)$ (the least informed one, see Benferhat et al, 1992) and checking the condition $S\Pi^*(p \wedge q) > S\Pi^*(p \wedge \neg q)$. These methods of inference of a default rule $p \rightarrow q$ from a body of generic knowledge Δ can be viewed as a focusing on the reference class pointed at by p, for the purpose of plausibly inferring q. On the contrary a revision process consists in modifying the set of rules Δ by adding or deleting rules, of by resticting the set of possible worlds.

7.2. Numerical settings.

If the state of partial ignorance is represented by a probability distribution P then, in the scope of revision, P is changed into $P(\cdot \mid A)$ such that:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \quad \text{(Bayes rule).}$$

The assumption behind Bayes rule is that the resulting probabilities of subsets of A should not change in relative value (see Gärdenfors, 1988). The same difficulty as above occurs if A is such that $P(A) = 0$. This case is usually ruled out by assumption. Bayesians often object to interpreting conditioning as the result of a revision of P by the new information $P'(A) = 1$ into $P' = P(\cdot \mid A)$. They rather interpret the probability distribution as describing generic knowledge and $P(B \mid A)$ as the result of changing the reference class of P from Ω to A, a process that we have called "focusing" (Dubois and Prade, 1992a), in which we compute what can be deduced from P for elements in A. Furthermore the Bayesian justification of probabilistic conditioning depends on the temporal coherence principle that is arguable (Smets and Kennes, 1994, Walley, 1991, pp.546-547, Smets, 1993b).

We now consider the case of belief functions and upper and lower probabilities. It can be pointed out that while revision and focusing do coincide in probability theory, they no longer coincide in more general settings (Dubois and Prade, 1992a).

Revision in belief function theory is defined by Dempster rule of conditioning that combines the conjunctive revision mode of the crude partial ignorance model, with Bayes rule normalization underlying a stability of the degrees of uncertainty in relative value. Indeed Dempster rule of conditioning can be described as follows:

- Given the new piece of information A, turn each focal element E_i into $E_i \cap A$, and attach the mass m_i to it, adding the masses m_i and m_j if $E_i \cap A = E_j \cap A$.
- Renormalize the masses allocated to non-empty subsets $E_i \cap A \neq \emptyset$ as with Bayes rule, so as to reallocate the masses m_i such that $E_i \cap A = \emptyset$ proportionally (note that the renormalization is not required in the transferable belief model in which case the mass given to $\emptyset$ represents the incoherence underlying the pieces of evidence that induce the beliefs, see Smets(1992).

This is the most classical form of conditioning, but other cases can be described that reflect other conditioning events (Smets 1991, 1993a). It is well-known (Shafer, 1976) that it comes down to compute the expected degree of potential support:

$$S\Pi(B \mid A) = \frac{S\Pi(A \cap B)}{S\Pi(A)}$$

Interestingly, Dempster's rule of conditioning can also be derived in a pure upper and lower probability context. If the belief function is viewed as characterizing a set of probabilities $\mathbb{P} = \{P \leq S\Pi\}$, then it can be proved that:

$$S\Pi(B \mid A) = \frac{S\Pi(A \cap B)}{S\Pi(A)} = \sup\left\{\frac{P(A \cap B)}{P(A)}, P \leq S\Pi, P(A) = S\Pi(A)\right\} \qquad (6)$$

This result is due to the fact that the constraint $P(A) = S\Pi(A)$ never forbids $\sup\{P(A \cap B), P \leq S\Pi, P(A) = S\Pi(A)\}$ to be equal to $S\Pi(A \cap B)$, i.e., we can always have $P(A) = S\Pi(A)$ and $P(A \cap B) = S\Pi(A \cap B)$ for the same probability measure P, if $S\Pi$ is a plausibility function. Equation (6) makes it clear the kind of revision at work with Dempster rule in the upper and lower probabilities context: the constraint $P(A) = S\Pi(A)$ corresponds to the selection of a maximum likelihood probability, which is very usual in statistics. As a generalization of this principle, Moral and De Campos (1991) have suggested that the magnitude of P(A) reflects the possibility of accepting the corresponding conditional probability $P(\cdot \mid A)$ in the updated set of probabilities.

Gilboa and Schmeidler (1992) have given a decision-theoretic interpretation of $S\Pi(B| A)$ in the setting of upper and lower probabilities, i.e., when the set $\mathbb{P}$ does not necessarily represent a belief function. Indeed equation (6) does hold when Π satisfies order 2 subadditivity only, that is $S\Pi(A \cup B) \leq S\Pi(A) + S\Pi(B) - S\Pi(A \cap B)$. The reason is that the property stating that we can always have $P(A) = S\Pi(A)$ and $P(B) = S\Pi(B)$ for a subset B of A for the same probability measure P is characteristic of order two subadditivity (see Huber, 1981). However equation (6) does not hold for upper probabilities deriving from any set of probabilities $\mathbb{P}$.

The alternative conditioning rule, already suggested by Dempster (1967), De Campos al. (1990), Fagin and Halpern (1989) and Jaffray (1992) and here referred to as focusing, is defined in the setting of upper and lower probabilities as:

$$S\Pi_A(B) = \sup\left\{\frac{P(A \cap B)}{P(A)}, P \leq S\Pi\right\} = \frac{S\Pi(A \cap B)}{S\Pi(A \cap B) + SN(A \cap \overline{B})}. \qquad (7)$$

The idea of focusing is to compute the probability of B in the state when A is supposed to be true without making any assumption about how the set of probabilities should be revised if A were actually true, especially without considering any probability measure in the set $\{P \leq S\Pi\}$ as impossible except those such that $P(\bar{A}) = 0$. It leads to performing a sensitivity analysis on

Bayes rule, when the probability function ranges over the set $\{P \le S\Pi\}$. Interestingly, if $S\Pi$ is a plausibility function then $S\Pi_A$ is still a plausibility function, the counterpart holds if $S\Pi$ is a possibility measure.

$S\Pi_A$ is generally much less informative than $S\Pi(\cdot \mid A)$ and even sometimes less informative than $S\Pi$ itself. For instance if $A \cap E_i \neq \emptyset$, and $A \not\subseteq E_i$ $\forall i$, then $S\Pi_A(B) = 1$ and $S\Pi_A(\bar{B}) = 1$, $\forall$ $B \neq A$, $B \subseteq A$, i.e., we get a total ignorance function on the referential set A. This is surprising in a learning process where A is a new piece of information and revision should improve our knowledge. The reason is that focusing is not made for revision and achieves no learning. To make it clear, suppose that the set of probabilities $\mathbb{P}$ represents knowledge stored in a database. Then $S\Pi_A(B)$ is part of the response of a query asking for the probability of being in B for an individual in A (focusing rule). One should give $SN_A(B) = 1 - S\Pi_A(\bar{B})$ as well. On the other hand $S\Pi(B \mid A)$ is the result of modifying the database by enforcing $P(A) = 1$ (revision). This distinction is in good agreement with the one arising in the ordinal (and logical) setting in the previous section. Instead of a set of default rules $\Delta = \{p_i \to q_i, i = 1,n\}$, consider a set $\mathbb{P}(\Delta)$ of probabilities induced by conditional probability bounds $\Delta = \{P(B_i \mid A_i) \ge a_i, i = 1,n\}$, where B_i (resp.: A_i) is the set of models of q_i (resp.: p_i) . Then focusing on a reference class A gathering the models of the proposition p leads to compute bounds of $P(B \mid A)$ induced by $\mathbb{P}(\Delta)$ for some B of interest. Let B be the set of models of q. If the coefficients a_i are infinitesimally close to 1, that is, of the form $1- \varepsilon$, finding that $P(B \mid A) \ge 1- \varepsilon$ in $\mathbb{P}(\Delta)$ is strictly equivalent to inferring the rule $p \to q$ from Δ in the possibilistic setting (or equivalently, in the preferential system of (Kraus et al. 1990)). See details in Lehmann and Magidor (1992). This is not surprizing because the rule $p \to q$ can be viewed as a conditional event $B \mid A$, in the sense of De Finetti, such that $P(B \mid A)$ is the probability of $p \to q$, and the logic of conditional events is a model of preferential system of (Kraus et al. 1990)(see Dubois and Prade, 1993).

Focusing can be justified in terms of belief functions only, namely $S\Pi_A(B)$, as obtained in (7), can be viewed as the upper limit of a family of belief functions obtained by transferring for all focal elements such that $E_j \cap A \neq \emptyset$, $E_j \cap \bar{A} \neq \emptyset$ only one part of the mass m_j to the set $E_j \cap A$ (see De Campos et al., 1990). Each possible partial mass transfer from each focal element E_j to $E_j \cap A$ determines a possible way of revising the belief function into another one. The focusing process refrains from choosing among these revisions by considering the lower envelope of all the possible revised belief functions.

However the focusing rule, as conditioning generic knowledge by factual evidence, is not part of the belief function theory as introduced by Shafer(1976) in his book and further developed by Smets (1988, 1990), Smets and Kennes, 1994). If we consider belief functions as such, it must be remembered that Shafer (1976) presents it as a theory of *evidence*, as does Smets in the transferable belief model. All uncertain pieces of information are factual, deal with a particular situation. Generic knowledge valid for a class of situations is never envisaged. In such a state of affairs, Dempster rule of conditioning is justified as a F-revision rule that integrates a new piece of evidence into an already existing body thereof. Dempster rule of combination is essentially the extension of Dempster rule of conditioning to uncertain pieces of new evidence, considered on equal grounds with respect to former evidence and assumed to be distinct of it.

On the contrary Dempster's view of belief functions as he originally produced it corresponds to a probability structure on top of partial ignorance, that produces a particular family of upper and lower probabilities. Dempster-like belief functions can then be viewed as modelling generic knowledge with higher order uncertainty, and the focusing rule really makes sense in order to assess the impact of new evidence on plausible conclusions to be derived on a case at hand. But the very rule Dempster laid bare in his paper is then not warranted, since it is a F-revision rule.

It is somewhat unexpected that, while a belief function is more general that a probability function, and Dempster rule of conditioning mathematically subsumes the Bayesian conditioning, Bayesian probability and belief functions (after Shafer(1976 book and the TBM model of Smets) seem to apply to rather different fields of investigation: Bayesian probability is tailored to focusing generic knowledge on proper reference classes described by pieces of factual evidence, belief function theory is construed as a tool for pooling pieces of uncertain evidence, and captures the notion of F-revision. On the contrary, Bayes rule of conditioning appears debatable in the scope of F-revision, and belief function theory proper (i.e. outside the framework of upper and lower probabilities) has never proposed anything to address the representation of generic knowledge about a population. This state of facts has been the cause of many a misunderstanding between Bayesians and advocates of belief functions.

On the contrary the setting of upper and lower probability generalizes the Bayesian theory to higher order uncertainty of the disjunctive type. Then, focusing and revision correspond to different forms of conditioning. If a set $\mathbb{P}$ of probability distributions encodes generic knowledge involving partial ignorance, accounting for a piece of factual evidence A leads to compute the focusing rule (7) and obtain for some conclusion B of interest the probability bounds $S\Pi_A(B)$ and $SN_A(B) = 1 - S\Pi_A(\text{ not } B)$. This if quite different if a G-revision must be performed by enforcing the constraint P(A) = 1 in $\mathbb{P}$, that is, add a new piece of generic knowledge (it leads to the conditioning rule of equation (6)).

7.3 Example

In order to tell focusing from revision let us consider a case where an assumption h is to be confirmed on the basis of an observation e. If we know the a priori probability P(h) and the conditional probabilities P(e | h), P(e | ¬h), the computation of the posterior probability comes down to the well-known formula:

$$P(h \mid e) = \frac{P(e \mid h) \cdot P(h)}{P(e \mid h) \cdot P(h) + P(e \mid \neg h) \cdot P(\neg h)}.$$

At this point, the computation has an ambiguous meaning: what does the data set {P(h), P(e | h), P(e | ¬h)} really encode? It may express what we know about the presence of a disease h in a population, and the probability of observing symptom e when h is present or absent. Observation e is made on an individual, supposed to be a typical element of class e. P(h | e) is the generic probability of h in class e that is applied to the individual. This is focusing. But nothing prevents us alternatively to imagine a situation where "observing e" means discovering that after all P(¬e) = 0 in all cases so that P(h) must be changed accordingly. Then the above computation is a G-revision.

On the contrary the data set {P(h), P(e | h), P(e | ¬h)}may encode uncertain evidence about a particular patient one is very familar with, P(h) being the probability that disease h is present now, P(e) the probability that symtom e will eventually show up. By assumption, this data set encodes a probability judgment that stems from preliminary investigation on the patient, not by using base rate disease statistics. Observing e is a new piece of evidence whereby P(e) = 1 is enforced. This is again revision, namely F-revision however.

What to do when no prior probability P(h) is known but only total ignorance is present? It is well-known that from $P(e \mid h) = \alpha < 1$, $P(e \mid \neg h) = \overline{\alpha} < 1$ and no prior, from sensitivity analysis on Bayes rule formula, we reach total ignorance, namely

$$\left\{ P(h \mid e) = \frac{x\alpha}{x\alpha + \overline{\alpha}(1-x)},\ x \in [0,1] \right\} = [0,1]. \tag{8}$$

Here we have used a focusing rule and we find $S\Pi_e(h) = 1$ and $S\Pi_e(\neg h) = 1$. It makes sense to use this conditioning rule in the first scenario whereby {P(e | h) = , $P(e \mid \neg h) = \overline{\alpha}$} is generic knowledge and e a piece of evidence. It proves that the knowledge of likelihoods tells nothing about the individuals in class e (except in the degenerated case when one of α or $\overline{\alpha}$ is 0) .

In the two other scenarios (revision), we must enforce P(e) = 1. The set {$P(e \mid h) = \alpha$, $P(e \mid \neg h) = \overline{\alpha}$, P(e) = 1} is considered as new data from which we should compute new bounds for P(h), but here the identity $P(e) = \alpha P(h) + \overline{\alpha}(1 - P(h)) \leq \max(\alpha, \overline{\alpha}) < 1$ shows that the new data set is inconsistent. If we apply a maximum likelihood strategy, maximizing P(e) in (8), we find, if $\alpha > \overline{\alpha}$, x = 1 in (8), and $S\Pi(h \mid e) = 1$ and $S\Pi(\neg h \mid e) = 0$. This very strong result is clearly debatable. It turns out that conditioning on this maximum likelihood set {P, $P(e \mid h) = \alpha$, $P(e \mid \neg h) = \overline{\alpha}$, P(e) maximal} is not equivalent to computing $S\Pi(h \wedge e)/ S\Pi(e)$ where $S\Pi(h \wedge e)$ and $S\Pi(e)$ are upper probabilities computed from {P, $P(e \mid h) = \alpha$, $P(e \mid \neg h) = \overline{\alpha}$}. So, a Dempster-rule-like computation does not apply since equation (6) fails. The maximum likelihood probability $P_e(h) = 1$ is not sensible, and it casts doubts on the maximum likelihood strategy in such a situation.

More generally the use of Bayes rule (which is a particular case of (6) anyway) is debatable for the purpose of revision while it is quite reasonable for focusing. Indeed the postulates for justifying Bayes rule are more difficult to advocate for (F- or G-) revision. Bayes rule is debatable for the purpose of revision especially when, as it is the case here, the input information conflicts with the prior information. It might be interesting to investigate the results that minimal change principles of probability kinematics would produce in the above example. Indeed the maximum likelihood rule is clearly not a minimal change rule here.

8 Revising Degrees of Belief Versus Revising Betting Probabilities

Insofar as we admit the presence of non-purely probabilistic representations of higher order uncertainty and particularly of partial ignorance, we are faced with a dilemma in a decision-making situation where we are provisionally prevented from betting due to the arrival of some new information. How should we assign betting probabilities so as to account for the new information? Indeed we are in a setting where beliefs are entertained at the credal level (encoded in some higher-order uncertainty model) and betting probabilities are derived from degrees of belief for decision purposes. Two strategies offer themselves: we can revise the old betting probabilities with Bayes rule applied to the input information or revise the knowledge at the credal level and recompute new betting probabilities. The problem is that the two strategies do not lead to the same final betting rates generally.

Let us consider a typical example where this phenomenon occurs. Namely, the mode of representation of prior knowledge affects the result of the revision process, and especially the enforced use of uniform betting probabilistic priors lead to debatable results under Bayes rule, for the purpose of revision. We shall reconsider the Peter, Paul and Mary saga (Smets and Kennes, 1994):

Example: The Peter, Paul and Mary saga

Big Boss has decided that Mr. Jones must be murdered by one of the three people present in his waiting room and whose names are Peter, Paul and Mary. Big Boss has decided that the killer on duty will be selected by the throw of a die: if it is an even number, the killer will be female, if it is an odd number, the killer will be male. You, the judge, know that Mr. Jones has been murdered and who was in the waiting room. You know about the die throwing, but You do not know what the outcome was and who was actually selected. *You are also ignorant as to how Big Boss would have decided between Peter and Paul in the case of an odd number being observed.*

Then You learn that if Big Boss had not selected Peter, then Peter would necessarily have gone to the police station at the time of the killing in order to have a perfect alibi. Peter indeed went to the police station, so he is not the killer. Note that the alibi evidence makes 'Peter is not the killer' and 'Peter has a perfect alibi' equivalent. Now the case is summarized by the following items of information:

Item 1: The killer is only Peter, Paul or Mary (one of them)
Item 2: The facts that the killer is a male or a female are equally probable.
Item 3: Peter produces an alibi
Question: Who is most likely to have been the killer ?

If we model this case in the crudest way by means of propositional logic we get that Item 1 is the disjunction Peter $\vee$ Paul $\vee$ Mary, Item 2 is interpreted as total ignorance and Item 3 to $\neg$Peter. The F-revision leads to (Peter $\vee$ Paul $\vee$ Mary) $\wedge$ $\neg$Peter = Paul $\vee$ Mary, i.e., Paul and Mary with equal possibility. Note that all items of information are considered of the same nature here: they are pieces of evidence on the case at hand, and are not generic. They do not pertain to a population.

It can be noticed that the Bayesian analysis based on Item 2 (P(female) = P(male) = 1/2) implies P(Paul $\vee$ Peter) = P(Mary) = 1/2. Then we compute the effect of Item 3 as

$$P(\text{Mary} \mid \neg\text{Peter}) = \frac{P(\neg\text{Peter} \mid \text{Mary}) \,.\, P(\text{Mary})}{P(\neg\text{Peter} \mid \text{Mary}) \,.\, P(\text{Mary}) + P(\neg\text{Peter} \mid \neg\text{Mary}) \,.\, P(\neg\text{Mary})}.$$

One has: P($\neg$Peter | Mary) = 1

but, P($\neg$Peter | $\neg$Mary) = P(Paul | Paul $\vee$ Peter) is unknown.

Hence using a Bayesian postulate (the Principle of Insufficient Reason, an arguable postulate but nevertheless quite reasonable in the present case) one can assume that given that the killer is either Paul or Peter, both are equipossible. Then Paul and Peter are equally probable, i.e., P(Paul | Paul $\vee$ Peter) = 1/2. And it is obtained:

$$P(\text{Mary} \mid \neg\text{Peter}) = \frac{1/2}{1/2 + 1/4} = \frac{2}{3}$$

$$P(\text{Paul} \mid \neg\text{Peter}) = 1/3$$

Hence contrary to the analysis based on propositional logic, a Bayesian policeman discovering Peter's alibi should be tempted to put Mary in jail, or at least to become suspicious towards her. In order to make any sense out of this result, a frequentist interpretation will help: assume Peter, Paul and Mary are prototypes of a group of people say Peter-like and Paul-like, and Mary-like. What the result says is that, given that killers are as often found to be male as female in the union of these three groups (because of the die experiment in the saga, but it could also be obtained by assuming the prior knowledge that there are as many male as female killers), and given that the proportion of killers in Paul-like group is the same as in Peter-like group, then Mary-like killers are found twice more often than Paul-like killers. This result is applied to the particular characters of the saga. But this is focusing, not F-revision since Item 1 and Item 2 are interpreted as two pieces of generic knowledge, namely that killers are as often found to be male as female, and that killers in Paul-like group are as numerous as in Peter-like group, this knowledge is further applied to the particular murder case, considering Item 3 (Peter has an alibi) as factual evidence indeed. The reference class on which the focusing takes place is Paul $\vee$ Mary. This is not exactly our original problem which only involves pieces of uncertain evidence.

The Bayesian result should not be surprising at the mathematical level since using the so-called Bayesian postulate for ignorance representation we can compute a unique distribution on the whole space:

$$P(\text{Paul}) = P(\text{Paul} \mid \text{Paul} \vee \text{Peter}) \cdot P(\text{Paul} \vee \text{Peter}) + P(\text{Paul} \mid \text{Mary}) \cdot P(\text{Mary})$$
$$= 1/2 \cdot 1/2 = 1/4 = P(\text{Peter}) = 1/2 \cdot P(\text{Mary}).$$

In some sense, Bayesian analysis comes down to claim that we know *from the start* that Mary is more likely to be the killer than any of the males. This piece of information is much stronger than what evidence suggests us but derives right away from the generic knowledge about killers (namely : P(Mary) = P(Paul $\vee$ Peter), P(Paul | Paul $\vee$ Peter) = 1/2). However from the available evidence, the dice experiment does not allow us to conclude that 'Mary is the killer' is more probable than that 'Paul is the killer'. We only know that the probability that Mary is the killer is equal to the probability that the killer is 'Peter or Paul'. And nobody told, in the case at hand, that Peter, Paul and Mary were typical representatives of distinct populations relevant to the case (maybe Peter and Paul are typical elements of the same population).

A belief function analysis of the case might consider a stronger interpretation of Item 2 than the one of propositional logic, and admit m(Peter $\vee$ Paul) = 1/2 = m(Mary). The latter is viewed as an uncertain piece of evidence pertaining to the Peter Paul and Mary case. But such an analysis would certainly refrain from splitting forever the weight on Peter and Paul into equal parts (nor in any other way), as the latter is *not at all* suggested by the pieces of evidence. Then Dempster rule of conditioning would simply transfer the mass 1/2 over to Paul, given that Peter has an alibi. Finally one would get P(Paul) = P(Mary) = 1/2 consistently with the propositional logic solution. Here we have performed a genuine F-revision of Item 1 and 2 by Item 3, not a focusing operation.

The present example is typical of the counter-intuitive results one can obtain if Bayesian uninformed priors are used as substitute to partial ignorance functions before a revision process takes place, although the results obtained by Bayesians can be interpreted as sensible if an interpretation is given to the uniform probabilities in terms of generic knowledge (e.g., frequentist). Note that if we have to bet on who is the killer before Peter's alibi is known, we would clearly use P(Mary) = 1/2, P(Paul) = P(Peter) = 1/4 for the bet. But after Peter's alibi is known using P(Mary) = 2/3 sounds strange. This should be a new bet involving only Paul and Mary, with equal chance on male and female, i.e., this is another betting problem.

To further advocate our thesis, notice that the story might have started before it was established that the killer was Peter, Paul or Mary only. Suppose originally, Sue and Deborah were suspected as well but they produced their alibi quite a long time before Peter. Then, the Bayesian method would start with

$$P(\text{Paul}) = P(\text{Peter}) = 1/4 \text{ and } P(\text{Mary}) = P(\text{Sue}) = P(\text{Deborah}) = 1/6$$

(since P(male) = P(female) from Item 2 and using the indifference principle). But then

$$P(\text{Mary} \mid \text{Mary} \vee \text{Paul}) = \frac{1/6}{1/6 + 1/4} = \frac{2}{5} < P(\text{Paul} \mid \text{Mary} \vee \text{Paul}) = \frac{3}{5}$$

Hence the results of Bayesian conditionning depend on where the story starts from. Especially if it starts with n females and m males then

$$P(\text{Mary} \mid \text{Mary} \vee \text{Paul}) = \frac{1/n}{1/n + 1/m} = \frac{m}{m + n}$$

i.e., anything between 0 and 1 according to the n and m. This is reasonable in a frequentist perspective, since it is a focusing process from a population of m + n persons. Indeed, this is a typical piece of generic knowledge. This is much less convincing in a subjectivist belief revision perspective (F-revision), when the end of the story seems to depend on which page the reader starts reading the novel. What we claim here is that a sequence of focusing steps exploiting generic knowledge does not produce the same results as a sequence of F-revision steps modifying an uncertain body of evidence. Indeed, posterior probabilities computed from a previous episode usually do not agree with the prior probabilities which would be given to the same remaining alternative if we ignore this episode. Moreover, although the distribution we start with in the above example seems to correctly encode Item 2 about the equal probability of male and female, it leads to results which are intuitively unexpected, like the following:

$$P(\neg\text{Deborah} \wedge \neg\text{Sue} \mid \neg\text{Mary}) = P(\text{Paul} \vee \text{Peter} \mid \text{Paul} \vee \text{Peter} \vee \text{Deborah} \vee \text{Sue})$$
$$= \frac{1/4 + 1/4}{1/4 + 1/4 + 1/6 + 1/6} = \frac{3}{5}.$$

Item 2 seems to suggest the value 1/2.

Conversely if P(Mary) = 1/6 and P(Paul ∨ Peter|Paul ∨ Peter ∨ Deborah ∨ Sue) = 1/2 are assumed from indifference principles, it would lead to strange priors on the set of 5 persons.

The above example does not question the value of the Bayesian approach that exploits generic knowledge in order to explain a particular situation via focusing on the proper reference class. It just confirms that if all uncertain information on a case is factual, and a F-revision process must take place, a Bayesian representation of partial ignorance can hardly be defended. How to represent total ignorance within classical probability theory is in fact an open question (and maybe an impossible one), often answered by strict Bayesians as being a non-existing problem. As they claim, total ignorance does not exist, one always has a state of knowledge quantifiable by a single probability measure on ANY and EVERY space. Total ignorance is a state where every proposition that is neither a tautology nor a contradiction is believed with the same strength. It cannot be represented in probability theory, as shown in Section 3. Total ignorance is therefore often considered as non-existing, a quite Procustean attitude. Other theories often cope easily with states of partial ignorance, usually because they are tailored for it, like possibility theory, or because of their higher expressive power, like belief functions that may account for any intermediary state between total ignorance, full probabilistic knowledge and complete (plain deterministic) knowledge. On the contrary possibility theory covers a more restricted range of epistemic states between total ignorance and complete (deterministic) knowledge almost without overlap with Bayesian representations.

9 - Conclusion

The main messages that this paper conveys are as follows:

-) The representation of partial ignorance is not necessarily related to the problem of making decisions. The Bayesian probability approach is tailored for decision but not necessarily for other kinds of reasoning (e.g., not always for belief revision).

-) A single number is not enough to quantify belief as expressing partial ignorance. As a consequence, frameworks that explicitly encode higher-order uncertainty must be used for that purpose, such as possibility theory, belief functions or upper and lower probabilities.

-) The extension of conditioning to partial ignorance formalisms reveals that this notion is two-sided. Conditioning can account for the idea of a shift of reference class (that is named focusing here), as well as for belief revision. In formalisms involving higher-order uncertainty, these two tasks are not necessarily implemented in the same way.

Our analysis strongly suggests a distinction between two mental levels where beliefs are involved (Smets and Kennes, 1994): a credal level where beliefs are entertained, combined and revised, and a so-called pignistic level where beliefs are used in a decision process. Deduction, plausible reasoning and belief change are processed at that credal level. The important point is that decision is not explicitly involved in those tasks and that beliefs can be maintained regardless of any subsequent decision. The absence of any forced link to any underlying decision process explains why the probability representation is not required, and other, alternative models can be advocated.

A good coherence between most theories of partial ignorance (possibility theory, belief functions, upper lower probabilities) can be observed. All of them insist on the use of two functions (a certainly-like lower function, a possibility-like upper function) in order to model the state of belief in a proposition. These theories basically differ by their relative power of expressiveness which is dictated by their basic interpretive settings (specified by canonical examples, for instance) or by a modeling assumption, e.g., ordinal levels instead of numerical ones in the case of possibility theory. Qualitative possibility theory represents belief in a purely ordinal setting by ranking propositions in terms of their certainty of being true and their possibility of being true in a separate manner. Possibility theory then comes close to mainstream approaches in nonmonotonic reasoning.The transferable belief model suggests that beliefs should be quantified by belief functions, but other alternatives might be considered, such as probability intervals. The latter approaches differ from Bayesian probability theory because they are more general. All non-purely probabilistic settings have in common, contrary to the Bayesian approach, the idea of leaving room for partial ignorance.

Of course beliefs are governing our acts. So when a decision must be made and insofar as the paradigm of fair bets applies, there must be a transformation between the beliefs held at the credal level into some additive measure to be used at the so-called pignistic level, i.e., the level where beliefs induce some probability measures used to compute the expected utilities needed in order to make rational optimal decisions. At the pignistic level, the probability measure *does not quantify our beliefs* but only our propensity to choose among prescribed acts. It is induced by our beliefs held at the credal level.

The paper has also pointed out that two ways of envisaging conditioning exist, and that they can only be told apart in frameworks that deal with generic knowledge and factual evidence separately. This distinction is especially clear in frameworks accounting for higher order uncertainty. Most of the debates that took place in the past concerning the relevance of Dempster rule of conditioning seem to have their origin in the lack of distinction between belief revision, that involve pieces of information at the same level, and focusing, that consists of projecting generic knowledge on a particular reference class pointed out by factual evidence.

On the whole, it is claimed here that a proper treatment of knowledge-based systems that can accommodate partial ignorance requires new tools for representing uncertainty that somewhat contradict a significant part of the probabilistic tradition. To enforce each and every problem involving partial ignorance into the Bayesian jacket sounds much too restrictive, if not blatantly erroneous.

References

Benferhat S., Dubois D. and Prade H. Representing default rules in possibilistic logic, Proc. of the 3rd Inter. Conf. on Principles of Knowledge Representation and Reasoning (KR'92), Cambridge, Mass., Oct. 26-29, 673-684.

Boutilier C. (1993) The probability of a possibility: Adding uncertainty to default rules. Proc. of the 9th Conf. on Uncertainty in Artificial Intelligence (D. Heckerman, A. Mamdani, eds.), Washington, DC, July 9-11, 1993, Morgan & Kaufmann, San Mateo, CA, 461-468.

Buchanan B.G., Shortliffe E.H. (1984) Rule-Based Expert Systems — The MYCIN Experiment of the Stanford Heuristic Programming Project. Addison-Wesley, Readings, MA.

Chellas B.F.(1980) Modal Logic. An Introduction. Cambridge University Press. Cambridge, U.K.

Cohen L.J. (1993) What has probability to do with strength of belief. In: Philosophy of Probability (J.P. Dubucs, ed.), Kluwer, Dordrecht, 129-143.

De Campos L.M., Lamata M.T, Moral S. (1990) The concept of conditional fuzzy measure, Int. J. Intell. Syst. 5, 237-246

De Finetti B. (1974) Theory of Probability. Wiley, New York, Vol. 2, 307-312.

DeGroot M.H. (1970) Optimal Statistical Decisions. McGraw-Hill, New York.

Dempster A.P. (1967), Upper and lower probabilities induced by a multiple-valued mapping. Ann. Math. Stat. 38, 325-339

Domotor Z. (1985) Probability kinematics, conditionals, and entropy principles. Synthese, 63, 75-114.

Dubois D., Prade H. (1982) On several representations of an uncertain body of evidence. In: Fuzzy Information and Decision Processes (M.M. Gupta, E. Sanchez, eds.), North-Holland, Amsterdam, 167-181.

Dubois D., Prade H.(1988a) Possibility Theory — An Approach to Computerized Processing of Uncertainty. Plenum Press, New York.

Dubois D., Prade H. (1988b) Representation and combination of uncertainty with belief functions and possibility measures. Computational Intelligence, 4, 244-264.

Dubois D., Prade H. (1988c) Decision evaluation methods under uncertainty and imprecision. In: Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision Making (J. Kacprzyk, M. Fedrizzi, eds.), Lecture Notes in Economics and Mathematical Systems, Vol. 310, Springer Verlag, Berlin, 48-65.

Dubois D., Prade H. (1990) Coping with uncertain knowledge —In defense of possibility and evidence theories—. Computers and Artificial Intelligence, 9(2), 115-144.

Dubois D., Prade H. (1991) Epistemic entrenchment and possibilistic logic. Artificial Intelligence, 50, 223-239.

Dubois D., Prade H. (1992a) Evidence, knowledge, and belief functions. Int. J. of Approximate Reasoning, 6, 295-319.

Dubois D., Prade H. (1992b) Belief change and possibility theory. In: Belief revision (P. Gärdenfors, ed.), Cambridge University Press, Cambridge, Mass., 142-182.

Dubois and Prade (1993) Conditional Objects as Nonmonotonic Consequence Relationships Tech Rep. IRIT, To appear in IEEE Trans. On Systems Man and Cybernetics, 1994

Dubois D. Prade H.(1994) Conditional objects and possibility theory. To appear in Conditional Logics (G. Crocco et al.) Oxford Univ. Press, UK

Dubois D., Prade H., Sandri S. (1993) On possibility/probability transformations. In: Fuzzy Logic — State of the Art (R. Lowen, M. Roubens, eds.), Kluwer, Dordrecht, 103-112.

Dubois D., Garbolino P., Kyburg H.E., Prade H., Smets Ph. (1991) Quantified uncertainty. J. of Applied on Non-Classical Logics, 1, 105-197.

Fagin, Halpern (1989) A new approach to updating belief. Research Report RJ 7222, IBM Research Division, San Jose, Cal.

Garbolino P. (1988) Bayesian theory and artificial intelligence: the quarrelsome marriage. In: Cognitive Engineering in Complex Dynamic Worlds (E. Hollnagel, G. Mancini, D.D. Woods, eds.), Academic Press, New York, 279-292.

Gärdenfors P. (1988) Knowdlege in Flux — Modeling the Dynamics of Epistemic States. The MIT Press, Cambridge, Mass.

Gärdenfors P., Sahlin N.E. (1988) Decision, Probability and Utility. Cambridge Univ. Press, Cambridge, G.B.

Gilboa I., Schmeidler D. (1992) Updating ambiguous beliefs. Proc of the 4th Conf. on Theoretical Aspects of Reasoning About Knowledge (TARK'92) (Y. Moses, ed.), Monterey, CA, March 22-25 1992, Morgan & Kaufmann, San Mateo, CA, 143-162.

Hintikka J.(1962) Knowledge and Belief. Cornell University Press Ithaca,

Huber P.J. (1981) Robust Statistics. Wiley, New York.

Jaffray J.Y. (1988) Application of linear utility theory to belief functions. In: Uncertainty and Intelligent Systems (Proc. of the 2nd Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'88), Urbino, Italy, July 1988) (B. Bouchon, L. Saitta, R.R. Yager, eds.), Lecture Notes in Computer Science, Vol. 313, Springer Verlag, Berlin, 1-8.

Jaffray J.Y. (1992) Bayesian updating and belief functions. IEEE Trans. on System, Man and Cybernetics, 22, 1144-1152.

Jeffrey R.C. (1983) The Logic of Decision. 2nd edition, U. Chicago Prc.

Jeffrey R.C. (1988) Conditioning, kinematics, and exchangeability. In: Causation, Chance, and Credence (B. Skyrms, W.L. Harper, eds.), D. Reidel, Dordrecht, Vol.1, 221-255.

Krantz D.H., Luce R.D., Suppes P., Tversky A. (1971) Foundations of Measurement — Vol. 1. Academic Press, New York.

Kraus S., Lehmann D. and Magidor M.(1990) Nonmonotonic reasoning, preferential models and cumulative logics, *Artificial Intelligence*, 44, 167-207, .

Kruse R., Schwecke E., Heinsohn J. (1991) Uncertainty and Vagueness in Knowledge Based Systems — Numerical Methods. Springer Verlag, Berlin.

Kyburg H.E., Jr. (1987) Bayesian and non-Bayesian evidential updating. Artificial Intelligence, 31, 271-293.

Kyburg H.E., Jr. (1989) Higher order probabilities. In: Uncertainty in Artificial Intelligence, Vol. 3 (L.N. Kanal, T.S. Levitt, J.F. Lemmer, eds.), North-Holland, Amsterdam, 15-22.

Kyburg H.E., Jr., Smokler H.E. (Eds.) (1964) Studies in Subjective Probability. Wiley, New York.

Langlotz C.P., Shortliffe E.H. (1989) Logical and decision-theoretic methods for planning under uncertainty. AI Magazine, 10(1), 39-47.

Lehmann D., Magidor M. (1992) What does a conditional knowledge base entail? Artificial Intelligence, 55, 1-60.

Levi I. (1967) Gambling with Truth. An essay in Induction and the Aims of Science. MIT Press, Cambridge, MA.

Moral S. and De Campos L. (1991) Updating uncertain information In: Uncertainty in Knowledge Bases (B. Bouchon-Meunier, R. Yager, L.A. Zadeh, eds.), Lecture Notes in Computer Science, Vol. 521, Springer Verlag, Berlin, 58-67

Ramsey F.P. (1931) Truth and probability. Reprinted in: Studies in Subjective Probability (H.E. Kyburg, H.E. Smokler, eds.), Wiley, New York, 61-92.

Savage L.J. (1972) The Foundations of Statistics. Dover, New York.

Schmeidler D. (1989) Subjective probability and expected utility without additivity. Econometrica, 57, 571-587.

Shackle G.L.S.(1961) Decision, Order and Time in Human Affairs, Cambridge University Press, Cambridge. U.K

Shafer G. (1976) A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ.

Shoham Y. (1988) Reasoning About Change — Time and Causation from the Standpoint of Artificial Intelligence. MIT Press, Cambridge, MA.

Smets P. (1982) Possibilistic inference from statistical data. Proc. of the 2nd World Conf. on Mathematics at the Service of Man (A. Ballester, D. Cardus, E. Trillas, eds.), Universidad Politecnica de Las Palmas, 611-613.

Smets P. (1988) Belief functions versus probability functions. In: Uncertainty and Intelligent Systems (Proc. of the 2nd Inter. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'88), Urbino, Italy, July 1988) (B. Bouchon, L.Saitta, R.R. Yager, eds.), Lecture Notes in Computer Science, Vol. 313, Springer Verlag, Berlin, 17-24.

Smets P. (1990) Constructing the pignistic probability function in a context of uncertainty. In: Uncertainty in Artificial Intelligence 5 (M. Henrion, R.D. Shachter, L.N. Kanal, J.F. Lemmer, eds.), North-Holland, Amsterdam, 29-39.

Smets P. (1991) Varieties of ignorance. Information Sciences, 57/58, 135-144.

Smets P. (1991) About updating. In: Uncertainty in Artificial Intelligence 91 (B. D'Ambrosio, Ph. Smets, P.P. Bonissone, eds), Morgan Kaufmann, San Mateo, CA, 378-385.

Smets P. (1992) The nature of the unnormalized beliefs encountered in the transferable belief model. in Dubois D., Wellman M.P., d'Ambrosio B. and Smets P. Uncertainty in AI 92. Morgan Kaufmann, San Mateo, Ca, USA, 1992, pg.292-297.

Smets P. (1993a) Jeffrey's rule of conditioning generalized to belief functions. Proc. of the 9th Conf. on Uncertainty in Artificial Intelligence (D. Heckerman, A. Mamdani, eds.), Washington, DC, July 9-11, 1993, Morgan & Kaufmann, San Mateo, CA, 500-505.

Smets Ph. (1993b) No Dutch Book can be built against the TBM even though update is not obtained by Bayes rule of conditioning. In: Probabilistic Methods in Expert Systems (Proc. of the Inter. Workshop on Probabilistic Expert Systems, Roma, Oct. 14-15, 1993) (R. Scozzafava, ed.), S.I.S., Roma, Italy, 181-204.

Smets P. (1993c) Probability of deductibility and belief functions. Tech. Report IRIDIA-93-5-3, Forthcoming in Logique et Analyse.

Smets P., Kennes R. (1994) The transferable belief model. Artificial Intelligence, 66, 191-234.

Spohn W. (1988) Ordinal conditional functions : a dynamic theory of epistemic states. In : Causation in Decision, Belief Change, and Statistics (W.L. Harper, B. Skyrms, eds.), Kluwer Academic Publ., Dordrecht, 105-134.

Teller P. (1973) Conditionalization and observation. Synthese, 26, 218-258.

Walley P. (1991) Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London.

Williams M.A.(1994) Transmutations of knowledge systems. Proc. the 4th Intern. Conf. on Principles of Knowledge Representation and Reasoning (KR' 94), Bonn, Germany, J. Doyle et al., eds., Morgan Kaufmann, San Francisco, Ca., 619-632.

Williams P.M. (1980) Bayesian conditionalisation and the principle of minimum information. Brit. J. Philos. Sci., 31, 131-144.

Williams P.M. (1982) Discussion of Shafer G: Belief functions and parametric models. J. Roy. Statist. Soc., B44, p. 342.

Wilson N. (1993) Decision making with belief functions and pignistic probabilities. in Symbolic and Quantitative Approaches to Reasoning and Uncertainty. Clarke M., Kruse R. and Moral S. (eds.), Springer Verlag, Berlin, pg. 364-371.

Yager R.R. (1982) Generalized probabilities of fuzzy events from fuzzy belief structures. Information Sciences, 28, 45-62.

Yen J. (1990) Generalizing the Dempster-Shafer theory to fuzzy sets. IEEE Trans. on Systems, Man and Cybernetics, 20(3), 559-570.

Zadeh L.A. (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-28.

Zadeh L.A. (1979) Fuzzy sets and information granularity. In: Advances in Fuzzy Set Theory and Applications (M.M. Gupta, R.K. Ragade, R.R. Yager, eds.), North-Holland, Amsterdam, 3-18.

Zadeh L.A. (1984) Fuzzy probabilities. Information Processing and Management, 19, 148-153.