Partial Probability: Theory and Applications

Frans Voorbraak
Department of Mathematics and Computer Science, University of Amsterdam
Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands
fransv@wins.uva.nl

Abstract

In this paper, we describe an approach to handling partially specified probabilistic information. We propose a formalism, called Partial Probability Theory (PPT), which allows very general representations of belief states, and we give brief treatments of problems like belief change, evidence combination, and decision making in the context of PPT. We argue that the generality of PPT provide new insights in all the mentioned problem areas. More detailed treatments of these issues can be found in several papers referred to in the text.

Keywords. partial probability theory, partial ignorance, probabilistic belief change, conditioning, constraining, evidence combination, decision under partial ignorance, minimax regret, satisficing.

1 Introduction

In many situations, the available evidence does not determine a unique probability function. Choosing a so-called "least informative" probability function satisfying the evidence may be the best general solution, in case one insists on using probability functions to represent belief states. However, in our opinion it is better to represent the available information, and no more than the available information, by means of probabilistic constraints. Absence of information, or ignorance, should lead to indeterminate, imprecise, or partially specified probabilities.

We propose a formalism called Partial Probability Theory (PPT), which, as the name suggests, allows probability assignments to be partially determined. The partiality, or imprecision, of the probability assignments is not the only difference from classical probability theory. As will be explained below, PPT makes explicit some distinctions which cannot be made within the classical probabilistic formalism. As a direct result, several important aspects of handling uncertainty (other than pure probabilistic reasoning) become more visible in the partial case.

For example, even if the probability functions are fully determined, PPT does not simply reduce to probability theory, since in PPT it is possible to distinguish probability assignments determined by hard evidence from probability assignments based on some assumptions.

If probabilistic information is only partially specified, then it may be necessary to make some assumptions in addition to the available evidence in order to draw useful conclusions and make sensible decisions. The explicit incorporation of these assumptions in PPT is perhaps the most distinctive feature of PPT compared to other general approaches to indeterminate probabilities, such as [25].

Dempster-Shafer theory (DS theory) also allows the representation of ignorance next to uncertainty. (See [12].) PPT is not only more general than DS theory, but also makes it possible (and necessary) to explicitly formulate the assumptions underlying the controversial Dempster's combination rule used in DS theory. Dempster's rule has probably contributed much to the popularity of DS theory, since such a combination rule can be shown not to exist in the case of probability theory, even though some ad hoc rules have been proposed in this case. See, for example, [1]. However, the justification of Dempster's rule is rather problematic.

In PPT, the assumptions underlying Dempster's rule can be stated clearly, and it can be argued that these assumptions are very strong and unrealistic. The following alternative, probabilistically justified, approach to combining evidence is proposed: explicitly add some (preferably weak) assumptions to the available partial probabilistic information, as long as these assumptions are necessary to reach useful conclusions by means of purely probabilistic reasoning. In section 4 we illustrate this approach.

Another area where the generality of PPT can provide new insights is the area of belief change. Presently, the most popular theory, or paradigm, for studying belief change operations is commonly known as AGM theory. In this paradigm, the focus is typically on changes of belief states represented by belief sets, i.e., logically closed sets of propositional sentences, but sometimes more refined representations of belief states, such as probability functions, are considered.

In AGM theory, three distinct kinds of belief change operations are distinguished: expansions, revisions and contraction. An expansion is the incorporation of new information which is consistent with the old beliefs, a revision incorporates new information which is (possibly) inconsistent with the old beliefs., and a contraction is reduction of the old beliefs in order to be consistent with the new information.

In the context of belief sets, an expansion of a belief set K with new information A is uniquely determined: the expanded belief set is obtained by adding A to the set K and subsequently taking the logical closure. In a probabilistic setting, expansion is not uniquely determined, but it is widely accepted that conditioning is the probabilistic version of expansion.

However, in section 3, it is argued that conditioning is best considered to be a kind a belief change which is different than expansion and other kinds of belief change that can be recognised in the abstract context of belief sets. The argument is made in the context of PPT which is sufficiently general to represent both belief sets and probability functions.

An important aspect of handling uncertainty is decision making based on uncertain information. Traditionally, decision problems are divided into the following three kinds: (1) decisions under certainty, where the decision maker knows the state of nature, (2) decisions under risk, where the decision maker is uncertain about the true state of nature, but is able to quantify this uncertainty through a probability measure over the sample space consisting of the possible states of nature, and (3) decisions under strict uncertainty, or complete ignorance, where the decision maker is uncertain about the true state of nature, and is not able to quantify this uncertainty. In this case, he can only list the possible states.

The case of partially quantified uncertainty, or partial ignorance, has not got the attention it deserves. This case is studied in section 5. The main conclusions are that decision making under partial ignorance is rather complex, since it is more general than decision making under complete ignorance, for which no generally accepted decision rule exists, but it seems possible to formalise several aspects of intuitively acceptable reasoning and decision making under partial ignorance.

The widely accepted decision rule for decisions under risk, i.e., maximising expected utility, can easily be extended to the case of partial ignorance, but this extended rule is in general rather weak. The reason is that the natural extension of the maximum expected utility (MEU) criterion to the partial case corresponds to a preference relation which is only a partial ordering. If the probabilities are highly indeterminate, then the extended maximum expected utility criterion leaves many choices open.

In section 5, we consider possible refinements of extended MEU criterion. We define extensions to the partial ignorance case of some decision rules originally proposed for decision under strict uncertainty. However, it is argued that these rules are not completely satisfactory as rules for optimising, and it is proposed to consider them to characterise acceptable actions. Especially the extension of what is known as the minimax regret criterion seems promising in this respect.

2 Partial Probability Theory

When using ordinary probability theory, the available information, or the belief state of an agent under consideration, is assumed to be represented by a probability space consisting of a sample space and a probability function over the sample space. In partial probability theory (PPT), a more refined representation is used.

A PPT belief state contains four elements: (1) a sample space, (2) a set of constraints on probability functions over the sample space representing hard, generic evidence, (3) another set of such constraints representing assumptions, and (4) a subset of the sample space representing the specific evidence. Formally, we propose the following definition.

Definition 1 (PPT belief state) *A* PPT belief state *is a quadruple $\langle \Omega, \mathcal{B}, \mathcal{A}, C \rangle$, where Ω is a sample space, $\mathcal{B}$ and $\mathcal{A}$ are sets of constraints on probability measures over Ω, and $C \subseteq \Omega$. The members of $\mathcal{B}$ are called belief or probability constraints, and $\mathcal{A}$ is called the set of assumptions.*

Throughout the paper, the sample space Ω is assumed to be finite. The constraints in $\mathcal{B}$ represent generic or general information about the relevant probabilities, whereas $\mathcal{A}$ represent general assumptions about these probabilities. The subset C of the sample space Ω represents specific information concerning the case at hand.

Example 1 *Consider a robot which has to recharge his battery. This can be done in two rooms, let us* *call them room 1 and 2. The rooms are equally far away from the robot. An example of generic information might be: "the door of room 1 is at least 40% of the time open". Suppose there is no other information available, and let p_i denote the probability of door i being open. Then $\mathcal{B} = \{p_1 \geq 0.4\}$. Since doors are typically sometimes open and sometimes closed, it might be reasonable to include $0 < p_i < 1$ in $\mathcal{A}$. However, such additional assumptions should be invoked only when they are necessary, for example, in case no reasonable decision can be made without assumptions. A good example of specific evidence is information about the state of the doors obtained by the sensors of the robot.*

Depending on the application, one can put restrictions on the form of the constraints. This allows the implementation of efficient algorithms for computing required probabilities, or probability bounds, in sufficiently simple special cases. One can also make use of the fact that there exist many probabilistic logics, with varying complexity and power of expression. See, for example, [3]. For many applications, a suitable formal language can be found in which it is possible to do formal reasoning with probabilistic constraints. Without restrictions on the form of the constraints, PPT has a high computational complexity, but it still is useful as a formal tool.

In PPT, partial ignorance can be represented by allowing the constraints not to determine a unique probability function. If desired, one can always add assumptions, such as the maximum entropy principle, to further constrain or even completely determine a probability function. But in contrast to probability theory, PPT distinguishes between hard evidence constraining the probabilities, and assumptions.

We write $\mathcal{B}^{\mathcal{A}}(\Omega)$ for the set of all probability measures over Ω which satisfy all the constraints in $\mathcal{A} \cup \mathcal{B}$. By definition, $\mathcal{B}^{\mathcal{A}}(\Omega) \subseteq \mathcal{B}(\Omega)$. Therefore, the conclusions warranted by $\mathcal{B}^{\mathcal{A}}(\Omega)$ are (weakly) stronger than those warranted by $\mathcal{B}(\Omega)$. However, it is intended to be understood that the conclusions warranted by $\mathcal{B}^{\mathcal{A}}(\Omega)$ depend on the assumptions represented in $\mathcal{A}$.

The specific information C is incorporated by essentially conditioning the probability measures satisfying the constraints on C. More formally:

Definition 2 (extended Bayesian conditioning) *Let Π be a set of probability measures over Ω, and let $C \subseteq \Omega$, such that for some $P \in \Pi$, $P(C) > 0$. Then Π_C is defined to be the set $\{P_C : P \in \Pi, P(C) > 0\}$.*

The conditioning information C may seem to be redundant, since at any time the belief state $\langle \Omega, \mathcal{B}, \mathcal{A}, C \rangle$ can be replaced with $\langle C, \mathcal{B}', \mathcal{A}', C \rangle$, where $\mathcal{B}'$ and A' are sets of constraints on probability measures over C. However, the generality of definition 1 will turn out to be convenient, in particular when belief changes are studied. We return to this issue in section 3.

Conditioning a belief state $\langle \Omega, \mathcal{B}, \mathcal{A}, C \rangle$ on D results in the belief state $\langle \Omega, \mathcal{B}, \mathcal{A}, C \cap D \rangle$. Given definition 2 , the PPT belief state $\langle \Omega, \mathcal{B}, \mathcal{A}, C \rangle$ gives rise to the set $\mathcal{B}(\Omega)_C$ of probability measures satisfying the belief constraints, and to the set $\mathcal{B}^{\mathcal{A}}(\Omega)_C$ of probability measures additionally satisfying the assumptions. Hence conditioning a PPT belief state essentially consists of applying extended Bayesian conditioning to two sets of probability measures.

It should be noted that the sets of probability measures induced be a PPT belief state are not necessarily closed convex sets. One could of course use the closed convex hull $conv(\Pi)$ of Π, but taking these convex closures instead of the sets themselves, or the underlying constraints, results in loss of information.

Example 2 *Consider again example 1, but now suppose that according to one source one of the doors is twice as likely to be open as the other one, whereas according to a second source it is more probable that door 1 is open than that door 2 is open. That is, $\mathcal{B}_1 = \{p_1 = 2p_2 \vee p_2 = 2p_1\}$ and $\mathcal{B}_2 = \{p_1 > p_2\}$. Let Π_i be the set of probability measures compatible with $\mathcal{B}_i$. Then $(\Pi_1 \cap \Pi_2)$ supports the (intuitively correct) conclusion that door 1 is twice as likely to be open as door 2 ($p_1 = 2p_2$). But this conclusion is not supported by $conv(\Pi_1) \cap conv(\Pi_2)$.*

As is well known, sets of probability measures, induce in a natural way (possibly non-additive) measures, namely their lower and upper envelopes, and belief functions of DS Theory and possibility functions of Possibility Theory can be seen as special cases of such measures.

3 Belief Change

Presently, the most popular theory, or paradigm, for studying belief change operations is commonly known as AGM theory [5]. In this paradigm, the focus is typically on changes of belief states represented by belief sets, i.e., logically closed sets of propositional sentences, but sometimes more refined representations of belief states, such as probability functions, are considered. Since for any probability function, the set of propositions which are assigned probability 1 forms a belief set, there is indeed a natural precise sense in which probability functions are refinements of belief sets.

In AGM theory, three distinct kinds of belief change operations are distinguished: expansions, revisions and contraction. An expansion is the incorporation of new information which is consistent with the old beliefs, a revision incorporates new information which is (possibly) inconsistent with the old beliefs., and a contraction is reduction of the old beliefs in order to be consistent with the new information. In the context of belief sets, an expansion of a belief set K with new information A is uniquely determined: the extended belief set is obtained by adding A to the set K and subsequently taking the logical closure.

It turns out that conditioning exactly corresponds to expansion in the following sense: conditioning a probability function P with associated belief set K on information A results in a new probability function P' such that the belief set K' associated with P' is identical to the expansion of K with A. This is generally assumed to be a good reason for considering conditioning to be the probabilistic version of expansion.

However, in [19] it is argued that conditioning is best considered to be a kind a belief change which is different than expansion and other kinds of belief change that can be recognised in the abstract context of belief sets. The argument is made in the context of PPT which is sufficiently general to represent both belief sets and probability functions.

In the context of PPT, there exists a belief change operation, called constraining, which consists of adding new constraints to the set of probabilistic constraints and which can be shown to share several more properties with belief set expansion than the properties common to expansion and conditioning. To make this argument, it is essential to refer to the partial probabilistic context, since completely determined probability functions cannot be constrained any further. Therefore, the constraining operation does not apply to completely determined probability functions in a non-trivial sense.

Lets us formally define the effect of constraining on a set of probability measures.

Definition 3 (constraining) *Let Π be a set of probability measures over Ω, and let c be a probabilistic constraint. Then $\Pi_{\& c}$ is defined to be the set $\{P : P \in \Pi, P satisfies\ c\}$.*

This definition allows adding a constraint which is inconsistent with the old evidence. The resulting set of probability measures is empty. One can imagine several ways to define an alternative operation which in such a case returns an non-trivial belief state. Such operations are called revisions, and the main subject of AGM theory is to find reasonable conditions on revisions. However, we concentrate on the case where the new constraint is consistent with the old evidence.

In the context of belief sets, it is possible to obtain any belief state from the ignorant belief state by a series of expansions. In PPT, constraining, but not conditioning, has the analogous property. This is one of the main reasons we prefer to constraining and not conditioning to be the probabilistic version of expansion.

Roughly speaking, it can be argued that the principal result of constraining is a decrease in ignorance, whereas conditioning is aimed at reducing uncertainty. Of course, the measurement of ignorance and uncertainty in a partial probabilistic setting is a controversial issue, but in [19] we propose the following provisional measures of these notions which at least allow the above statement to be made more precise.

Definition 4 (uncertainty) *Let Π be a set of probability measures. The* uncertainty *$U(\Pi)$ of Π is defined as follows.*

$$U(\Pi) = \max\{H(P) : P \in conv(\Pi)\},$$

where $H(P)$ is the entropy of P, and $conv(\Pi)$ is the closed convex hull of Π.

Definition 5 (ignorance) *Let Π be a set of probability measures ove Ω. The* ignorance *$I(\Pi)$ of Π is defined as follows.*

$$I(\Pi) = \frac{\sum_{A \subseteq \Omega} (\Pi_{up}(A) - \Pi_{low}(A))}{2^{|\Omega|} - 2}.$$

In other words, the uncertainty of Π is defined to be the maximum uncertainty (measured in entropy) of the probability measures in the set Π, and its ignorance is an average of the ignorance with respect to each event (measured in the difference between the upper and lower probability bounds).

Both constraining and conditioning can be used to change the maximally ignorant and maximally uncertain belief state, which has no constraints and allows all probability measures, into a belief state which is minimally ignorant and minimally uncertain. Hence we cannot claim that constraining is exclusively used for reducing ignorance and conditioning is only used to reduce uncertainty. However, there does exist some bias.

For example, constraining is guaranteed to (weakly) reduce ignorance. Also, constraining can be used to strictly reduce the ignorance of all belief states which are not minimally ignorant already. Neither property holds for conditioning, although the second property is not true for a natural generalisation of conditioning known as minimum cross entropy update. On the other hand, if a set contains just one probability measure, then constraining can not reduce its uncertainty, whereas conditioning can.

Details of the above can be found in [19]. In [7] an axiomatic characterisation of conditioning and constraining is given. We do not claim to be the first to distinguish the notions of conditioning and constraining. For example, in [2] a similar distinction is made.

4 Combining Evidence

The combination of evidence is an extremely important problem when reasoning with uncertainty. Instances of this problem include the pooling of opinions of different experts, the weighting of different arguments in favour of or against some conclusion, and the fusion of several sensor readings.

In [17, 18, 21], the problem of combining evidence is discussed in the context of PPT. In this context, the assumptions underlying Dempster's rule can be stated clearly, and it can be argued that these assumptions are very strong and unrealistic. The following alternative, probabilistically justified, approach is proposed: explicitly add some (preferably weak) assumptions to the available partial probabilistic information, as long as these assumptions are necessary to reach useful conclusions.

A rule for minimal combination, i.e., without adding assumptions and even dropping the previously made assumptions, can be formalised as follows.

Definition 6 (minimal PPT combination)
The minimal combination of $\langle \Omega, \mathcal{B}_1, \mathcal{A}_1, C_1 \rangle$ *and* $\langle \Omega, \mathcal{B}_2, \mathcal{A}_2, C_2 \rangle$ *is defined to be the PPT belief state* $\langle \Omega, \mathcal{B}_1 \cup \mathcal{B}_2, \emptyset, C_1 \cap C_2 \rangle$.

In example 1, this combination rule was already applied. The combination is called minimal, since the combined belief state does not contain any of the assumptions present in the original belief states. The reason is that some of these assumptions may be inconsistent with $\mathcal{B}_1 \cup \mathcal{B}_2$. A less cautious combination would include the assumptions $\mathcal{A}_1 \cup \mathcal{A}_2$, or a subset of this set, if the full set is inconsistent with $\mathcal{B}_1 \cup \mathcal{B}_2$. One can also consider some additional assumptions, in particular concerning the interaction of the evidence.

In [17, 18, 21], the following concrete example involving sensor fusion is discussed in some detail.

Example 3 *Consider an autonomous vehicle which has to perform some subtle manoeuvres for which it needs to know its distance to its nearest obstacle with an accuracy of 0.01 metre. To measure this distance, it can use three sensors, let us call them* S_1, S_2, *and* S_3. *They each provide the vehicle's CPU with an integer between 0 and 999, representing the measurement in centimetres. They cannot detect objects which are removed 10 metre or more.*

Each sensor is reliable 50% *of the time. That is,* 50% *of the time, the sensor is working properly and the returned number is the correct distance. The remaining* 50% *of the time, the sensor is in some way disturbed and the returned number is not properly related to the true distance, although the number may of course happen to be correct by shear luck.*

Suppose that on a particular occasion, all three sensors return the number 454. Intuitively, this provides strong support that 454 is the actual distance to the nearest obstacle, even though the sensors are rather unreliable when taken in isolation. Of course, such sensor information never gives complete certainty about the actual distance, but let us say that a 5% *failure rate would still be acceptable. So if there is at least a* 95% *chance that a reading is correct, the vehicle should use this reading to guide its manoeuvres.*

It can be argued that the approach of combining evidence in PPT described above gives better results than Dempster's rule using weaker (explicit) assumptions than the assumptions implicitly underlying the application of Dempster's rule. For example, Dempster's rule does not sanction the conclusion that 454 is the actual distance with the required 95% certainty, whereas intuition and (under weak assumptions) PPT do.

In DS theory, the evidence of the above example can be represented by three identical mass functions m_1, m_2 and m_3 over the sample space $\Omega = \{0, 1, 2, \ldots, 999\}$. Each of them assigns 0.5 mass to $\{454\}$ and 0.5 mass to Ω. Combining these three mass functions by means of Dempster's combination rule leads to the mass function $m_1 \oplus m_2 \oplus m_3$, given by $m_1 \oplus m_2 \oplus m_3(\{454\}) = 0.875$, and $m_1 \oplus m_2 \oplus m_3(\Omega) = 0.125$. This result does not give the required 95% guarantee that 454 is correct.

Moreover, in order to apply Dempster's rule, one has to assume that the pieces of evidence are (DS) independent, in the sense that the sources behave independently. That is, the (un)reliability of one sensor is independent form the (un)reliability of the other sensors. More precisely, if we write R_i for the proposition that sensor S_i is reliable, and r_i for the propositional variable denoting either R_i or $\overline{R_i}$. Then one has to assume that $P(r_i, r_j) = P(r_i)P(r_j)$, whenever $1 \leq i < j \leq 3$.

This assumption of DS independence is quite strong, and practically never satisfied by sensors, since they usually have common causes for their unreliable behaviour (the movement of the vehicle, the size and shape of the obstacles, et cetera). To make matters worse, this assumption of DS independence is not sufficient to justify Dempster's rule, as is shown in [16] In addition, one has to assume that each instantiation of r_1, r_2, r_3 is equally confirmed by the evidence. This additional assumption is explained below.

Let x_1 (y_2, z_3) denote the fact that the reading x (y, z) is obtained from sensor S_1 (S_2, S_3). In the case of receiving evidence $454_1, 454_2, 454_3$, the assumption of equal confirmation is represented by the equation $P(r_1, r_2, r_3|454_1, 454_2, 454_3) = P(r_1, r_2, r_3)$.

The reasoning giving a confidence of 0.875 in 454 goes as follows. The chance of 454 being the actual distance is at least the probability that at least one of the sensors is reliable. By DS independence the prior probability of this event is 0.875, and by the assumption of equal confirmation, this probability is not changed after receiving the evidence. The assumption of equal confirmation is not plausible since agreeing sensors are more likely to be reliable (in which case they *have* to agree) than unreliable (in which case there is only a very small chance of agreeing).

In PPT, the evidence provided by the reading 454 of sensor S_1 can be captured by the PPT belief state $\langle \Theta_1, \mathcal{B}_1, \emptyset, C_1 \rangle$, where Θ_1 is the set of triples $\langle w_0, x_1, r_1 \rangle$ such that w denotes the real distance to the nearest obstacle, and x_1 and r_1 are as defined above. It follows that $0 \leq w \leq 999$, $0 \leq x \leq 999$, $r_1 \in \{R_1, \overline{R_1}\}$, and if $r_1 = R_1$, then $w = x$. Further, $\mathcal{B}_1 = \{P(R_1) = 0.5\}$, and $C_1 = \{\langle w_0, x_1, r_1 \rangle \in \Theta | x = 454\}$.

Minimally combining the three pieces of evidence gives the PPT belief state $\langle \Theta, \mathcal{B}, \emptyset, C \rangle$, where Θ is the set of tuples $\langle w_0, x_1, y_2, z_3, r_1, r_2, r_3 \rangle$, where $w, x, y, z \in \{0, \ldots, 999\}$, $r_i \in \{R_i, \overline{R_i}\}$, and if $r_i = R_i$, then $w =$ the reading of sensor S_i. Further, $\mathcal{B} = \{P(R_1) = P(R_2) = P(R_3) = 0.8\}$, and C is the subset of Θ characterised by $x = y = z = 454$.

In PPT one cannot conclude much based on the evidence alone, since the available information about the sensors simply does not allow strong conclusions about the value of $P(r_1, r_2, r_3|454_1, 454_2, 454_3)$. However, relatively weak additional assumptions are sufficient to obtain at least as strong conclusions as DS theory. For example, one could assume that $P(\overline{R_1}, \overline{R_2}, \overline{R_3}|454_1, 454_2, 454_3) \leq P(\overline{R_1}, \overline{R_2}, \overline{R_3})$, which is much weaker and far more plausible than the corresponding equality.

In [17, 18, 21], we propose some relatively weak assumption that result in a confidence of approximately 0.999999 that 454 is the actual distance if all three sensors say it is, and we show how even if only two of the three sensors agree, a high confidence of approximately 0.998 can be justified.

5 Decision Making

An important aspect of handling uncertainty is decision making based on uncertain information. For decisions under risk, where the decision maker is uncertain about the true state of nature, but is able to quantify this uncertainty through a probability measure over the sample space consisting of the possible states of nature, maximising expected utility is widely accepted to be a rational decision rule. Several rules have been proposed for decisions under strict uncertainty, or complete ignorance, where the decision maker is uncertain about the true state of nature, but is not able to quantify this uncertainty and can only list the possible states.

If the probabilities are partially determined, the decision maker typically finds himself somewhere between these two extreme cases. Actually, the case of partial ignorance is the general case, which includes both risk and strict uncertainty. Therefore, it should be no surprise that decision making under partial ignorance is rather complex, since it is more general than decision making under complete ignorance, for which no generally accepted decision rule exists. However, it still seems possible to formalise several aspects of intuitively acceptable reasoning and decision making under partial ignorance.

In [20, 22] several naive decision strategies for partial ignorance are described and formalised in the context of PPT. These strategies can be informally described as follows: (1) (try to) learn more, (2) be satisfied with satisfactory rather than (necessarily) optimal choices, and (3) base one's actions on (reasonable) assumptions in addition to the available knowledge. It is argued that sometimes a combination of these strategies is necessary. For example, when deciding whether either to learn more or to act on the basis of the available information, it may be necessary to make some assumptions in order to sufficiently determine the (expected) value of the possible information acquiring actions.

The widely accepted decision rule for decisions under risk, i.e., maximising expected utility, can easily be extended to the case of partial ignorance, by defining the following preference relation on actions:

$$a \geq_{\Pi} b \text{ iff } \forall P \in \Pi(EU_{P,U}(a) \geq EU_{P,U}(b)).$$

Here Π denotes the set of probability measures satisfying the constraints, and $EU_{P,U}(a)$ denotes the expected utility, given a probability measure P and utility function U, of the action a. Now choose an action which is maximally preferred according to the above relation $\geq_\Pi$.

This extension to the case of partial ignorance of the maximum expected utility criterion is quite natural and has been proposed before. In particular, in [4] and [8] this criterion has been studied in case the probabilities of the possible states of nature were not necessarily exactly known, but could at least be *ranked*.

However, this extended rule is in general rather weak. For example, the rule does not allow the robot to make a choice in the situation of example 1, whereas many people would intuitively prefer going to room 1 for which substantial evidence of being open is available.

The extended rule is weak since it corresponds to a preference relation which is only a partial ordering. (As is well known, [11] establishes a correspondence between maximising expected utility and a total preference ordering when deciding under risk.) The correspondence between preference and maximising utility under partial ignorance has been made precise by [6]. If the probabilities are highly indeterminate, then the extended maximum expected utility criterion leaves many choices open.

Some authors propose that one should base one's decisions on a particular probability measure compatible with the available evidence. This probability measure is sometimes allowed to be arbitrarily chosen from the candidates, and sometimes it is assumed to be in some sense least informative, such as the measure with the maximum entropy or the pignistic probability measure of [15].

In [10] it is concluded that if one decides on the basis of a single probability measure one could just as well use the strict Bayesian approach from the start. However, even if decisions are based on a probability measure, a richer belief representation can still be used for other activities, such as combining evidence or belief revision. (Compare our argument using example 2 against the restriction to convex sets of probability measures.) Other authors allow decision makers to be indecisive between incomparable maximally preferred actions. See, for example, [25]. See [9] for a more extended overview of different approaches to deciding under partial ignorance.

In [20, 22] we propose to refine the extended MEU criterion by introducing some additional criteria obtained from decision rules originally proposed for decision under complete ignorance. These refinements can support intuitive preferences not supported by the extended MEU rule. For example, they support the preference for room 1 in example '1.

In particular, we propose the following extended minimax regret rule.

Definition 7 (extended minimax regret) *Define the P,U-regret $R_{P,U}(a)$ and the maximal regret $R_{\Pi,U}(a)$ of an action a as follows.*

$$R_{P,U}(a) = M_{P,U} - EU_{P,U}(a).$$

$$R_{\Pi,U}(a) = \sup\{R_{P,U}(a) : P \in \Pi\}.$$

Here, $M_{P,U} = max\{EU_{P,U}(b) : b \in A\}$. According to the extended minimax regret rule one should choose an action with minimal maximal regret.

This rule has been proposed before, and is known as the Γ-minimax regret rule in the field of robust Bayesian analysis. The following proposition summarises some nice properties of this decision rule.

Proposition 1 *The extended minimax regret rule generalises both Savage's minimax regret rule and the ordinary MEU rule, and it refines the extended MEU rule. More precisely,*

1. *In the case of complete ignorance, the extended minimax regret rule reduces to Savage's minimax regret rule.*
2. *In the case of decision under risk, the extended minimax regret rule reduces to the ordinary MEU rule.*
3. *The extended minimax regret rule refines extended MEU in the sense that $a \geq_\Pi b$ implies $R_{\Pi,U}(a) \leq R_{\Pi,U}(b)$.*

However, there are some problems with using the extended minimax regret rule. For example, in [20, 22] we also consider refining extended MEU with an extension of Wald's maximin return rule and since the extended maximin return rule and the extended minimax regret rule do not necessarily agree, one has to choose between the two rules (and other related rules).

Example 4 *Consider again example 1, but now suppose that we know that $0.4 \leq p_1 \leq 0.8$ and $0.5 \leq p_2 \leq 0.6$. Then a_1 and a_2 are incomparable according to extended MEU (and both maximally preferred). The maximal regret of a_2 is higher than that of a_1, since $R(a_1) = 0.2$ and $R(a_2) = 0.3$. The expected security level $ES(a)$ of an action a is defined as $\inf_{P\in\Pi} U_{P,U}(a)$, and the extended maximin return rule prefers actions with higher expected security level.*

Thus a_1 is preferred according to the extended minimax regret rule, whereas a_2 is preferred according to the extended maximin return rule, since $ES(a_1) = 0.4$ and $ES(a_2) = 0.5$.

A perhaps even more serious (and well-known) problem is that the ordering obtained by comparing maximal regret does not satisfy independence of irrelevant alternatives, i.e. a choice between two alternatives based on the (extended) minimax regret rule can be influenced by the availability of a third alternative.

Since the maximum regret of an action gives some upper bound on the suboptimality of the action, it can be interpreted as a kind of satisfaction level of the action. Although optimising based on comparing extended maximal regret can be shown not to be independent of irrelevant alternatives, this independence does hold when one considers satisficing based on the maximal regret. Therefore, we propose to use the minimax regret rule for satisficing rather than for optimising.

The connection between the extended minimax regret criterion and the notion of satisficing, as introduced by Simon [13, 14], is made more precise in [24].

6 Summary and Conclusions

We described an approach to handling partially specified probabilistic information. Our proposed Partial Probability Theory (PPT) allows very general representations of belief states, and includes many previously proposed formalisms as special cases. Of of the characteristic features of PPT is the explicit treatment of assumptions.

We briefly discussed problems like belief change, evidence combination, and decision making in the context of PPT. In the area of belief change, we argued that the probabilistic variant of the notion of expansion of AGM theory is not conditioning, but constraining. The point of view of partially specified probability is necessary to clearly distinguish conditioning and constraining.

Concerning evidence combination, we pointed out that in order to reach useful conclusions in a partial probabilistic setting, it may be unavoidable to make assumptions about the interaction of the pieces of evidence to be combined. We used a concrete robotic sensor fusion example to illustrate our approach and to compare it with Dempster's combination rule.

In the area of decision making under partial ignorance, the rule of maximising expected utility is quite weak and should be supplemented by other considerations. We do not propose a simple stronger decision rule, but we consider it in general unavoidable to base one's decisions on assumptions or to opt for satisficing alternatives rather than to insist on optimising.

We do not claim that PPT automatically includes solutions to all problems encountered when probabilities are allowed to be partially specified, but we propose PPT as a tool for expressing these problems and compare different solutions.

Acknowledgements

The investigations were carried out as part of the PIONIER-project Reasoning with Uncertainty, subsidized by the Netherlands Organization of Scientific Research (NWO), under grant pgs-22-262.

References

[1] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis*, (second edition) Springer, New York, 1980.

[2] D. Dubois, and H. Prade, Evidence, knowledge and belief functions, *International Journal of Approximate Reasoning*, 6 (1992) 295-319.

[3] R. Fagin, J.Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities, *Proceedings 3rd IEEE Symposium on Logic in Computer Science* (1988) 277-291. Expanded version: *IBM Research Report* RJ 6190 (1988). Also in *Information and Computation* 87 (1990) 78-128.

[4] P.C. Fishburn. *Decision and Value Theory*, Wiley, New York, 1964.

[5] P. Gärdenfors, *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, MIT Press, Cambridge MA, 1988.

[6] F.J. Giron and S. Rios. Quasi-Bayesian Behaviour: A more realistic approach to decision making?, *Bayesian Statistics*, J.M. Bernardo et al. (eds.), University Press, Valencia, 1980, 17-38.

[7] A.J. Grove and J.Y. Halpern. Updating sets of probabilities, *Proceedings 14th Conference on Uncertainty in AI* (1998) 173-182.

[8] Z.W. Kmietowicz and A.D. Pearman, *Decision Theory and Incomplete Knowledge*, Gower, 1981.

[9] R. Loui, Decisions with indeterminate probabilities, *Theory and Decision* 21 (1986) 283-309.

[10] P. Maher, What is wrong with strict bayesianism? *PSA 1986 Vol. 1*, A. Fine and P. Machamer (eds.), (1986) 450-457.

[11] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behaviour*, Princeton University Press, Princeton, 1944.

[12] G. Shafer, *A Mathematical Theory of Evidence* (Princeton U.P., Princeton, 1976).

[13] H. A. Simon, A behavioral model of rational choice, *Quarterly Journal of Economics*, 69 (1955) 99-118. Reprinted in H. A. Simon, *Models of Bounded Rationality. Vol. 2*, MIT Press, Cambridge Mass. (1982) 239-258.

[14] H. A. Simon, Theories of bounded rationality, *Decision and Organization*, C. B. Radner and R. Radner (eds.), Amsterdam: North-Holland (1972) 161-176. Reprinted in H. A. Simon, *Models of Bounded Rationality. Vol. 2*, MIT Press, Cambridge Mass. (1982) 408-423.

[15] P. Smets, Constructing the pignistic probability function in a context of uncertainty, *Uncertainty in Artificial Intelligence 5*, M. Henrion, R.D. Schachter, L.N. Kanal, and J.F. Lemmer (eds.), Elsevier Science Publ., Amsterdam,1990, 29-39.

[16] F. Voorbraak, On the justification of Dempster's rule of combination, *Artificial Intelligence* 48 (1991) 499-515.

[17] F. Voorbraak, Combining unreliable pieces of evidence, *ILLC Research Report CT-95-07*, ILLC, University of Amsterdam, 1995.

[18] F. Voorbraak, Evidence combination in AI and robotics, in: J.-J. Ch. Meyer and L.C. van der Gaag (eds.), *NAIC'96: Proceedings of the Eight Dutch Conference on Artificial Intelligence*, Utrecht University, 1996, 369-380.

[19] F. Voorbraak, Probabilistic belief expansion and conditioning, *ILLC Research Report LP-96-07*, ILLC, University of Amsterdam, 1996 (to appear in Journal of Logic, Language and Information).

[20] F. Voorbraak, Decision analysis using partial probability theory, in: J. Doyle et al. (eds.), *Qualitative Preferences in Deliberation and Practical Reasoning* (AAAI-97 Spring Symposium), Stanford University, 1997, 113-120. Electronically available at www.kr.org/qdt.

[21] F. Voorbraak, Combining evidence under partial ignorance, in D. Gabbay et al. (eds.), *Qualitative and Quantitave Practical Reasoning* (LNAI 1244), Springer, Berlin,1997, 574-588.

[22] F. Voorbraak, Deciding under partial ignorance, *EUROBOT'97: 2nd EUROMICRO Workshop on Advanced Mobile Robots*, IEEE/CS Press, 1997, 66-72.

[23] F. Voorbraak and N. Massios, Decision-theoretic planning for autonomous robotic surveillance, *Proceedings ECAI'98 Workshop Decision Theory meets Artificial Intelligence. Qualitative and Quantitative Approaches*, University of Brighton, 1998, 23-32. (Selected for publication in a special issue of Applied Intelligence.)

[24] F. Voorbraak, Optimising and satisficing under partial ignorance, in: *Proceedings LOFT3*, Torino, Italy, 1998. Also available as *LLC Research Report CT-1998-07*, Amsterdam: ILLC, University of Amsterdam.

[25] P. Walley. *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.