Other uncertainty theories based on capacities

Sébastien Destercke, Didier Dubois

Other uncertainty theories based on capacities

S. Destercke, D. Dubois

Tuesday 3$^{\text{rd}}$ June, 2014

Contents

Chapter 1

Other uncertainty theories based on capacities

by Sébastien Destercke, Didier Dubois

Some of the simple mathematical models reviewed in Chapter **??** first emerged as basic building blocks of other uncertainty theories, and not as special instances of coherent lower previsions. In the following, the term *uncertainty theories* refers to several mathematical models and approaches devoted to the numerical or ordinal representation of beliefs induced by uncertainty due to the variability of phenomena, or the presence of incomplete information, or both. As these theories emerged from different areas and motivations, they led to different requirements, settings or axioms and to different tools for information processing. More precisely, we can enumerate several lines of thought motivating the development of uncertainty theories:

- The idea that in Bayesian approaches to statistics, prior probabilities on model parameters are hard to precisely specify, partly because there is no truly satisfying "non-informative" prior, and partly because probabilities elicited from experts are most of the time imprecise. This is the robust Bayesian approach [7].

- A questioning of the classical fair betting behavior framework of subjective probability [25] leading to upper and lower previsions for gambles (See Chapter **??**).

- Systematic deviations of human behavior from Savage's [71] axioms for rational decision-making (especially the sure-thing principle) in the presence of ambiguity about probabilities (like in Ellsberg urn experiment [44]). This is the non-additive expected utility trend, where subjective probability is replaced by more general set functions [11] (see Chapter **??**).

- The claim that incomplete information is better captured by sets of more or less plausible states of nature. It is a subjective, often ordinal, view of plausibility advocated by Shackle [73], where impossibility is equated to the absence of surprise.

This is also the starting point of the modal view of uncertainty first put forward by Hamblin [51] and Lewis [59], that is closely related to possibility theory [37].

- The handling of linguistic information understood as defining flexible constraints on values of parameters of interest (this is Zadeh's view of possibility theory [87]).

- The representation and fusion of uncertain evidence such as unreliable testimonies, a problem that goes back to the origins of probability theory. This is a topic of renewed interest in Artificial Intelligence and Robotics. This is the theory of belief functions by Shafer [74].

- The issue of plausible reasoning in inference tasks under incomplete statistical or commonsense information, where the choice of the proper reference class is the key problem. This concern goes from Kyburg's evidential probabilities to non-monotonic reasoning in Artificial Intelligence [58].

Apart from the approaches relying on convex sets of probability measures and lower expectations, presented in Chapters **??** and **??**, the two main uncertainty representations in the literature that tolerate imprecision are possibility distributions and random disjunctive sets. This chapter will devote special attention to the theories that have emerged from them. These formalisms follow different calculi and use different basic building blocks, hence different languages. Even if these building blocks can be ordered in terms of generality and mathematically unified under the umbrella of lower previsions, this difference in language create discrepancies when processing information that are as many impediments preventing a complete unification. Some notions, easily expressible in special cases of lower prevision using the language of the specialized theory become very difficult to express in the more general setting that does not use this language. For instance the setting of possibility theory enables many information fusion operations to be defined, that have no clear counterparts in more general theories. Likewise Dempster's rule of combination in evidence theory is expressed in terms of set functions that are not used in the theory of lower previsions, even these set-functions (Moebius transforms, commonality) can be mathematically defined for lower previsions. In some cases, bridges can be built and equivalent notions identified between these frameworks, but sometimes only to a limited extent, as each theory uses different primitive notions. There is therefore a need for both unifying concepts when possible and clarifying the underlying assumptions made by each theory, so as to lay bare situations to which they apply, and better situate them with respect to the lower prevision approach put forward in this book.

The route taken in this chapter is different from the one adopted so far. Indeed, the framework of desirability assumes an additive linear bipolar "currency" scale (the real line), and derives upper and lower probabilities and previsions from rationality requirements. By linear bipolar value scale, we mean a scale where the zero plays a key-role for separating the good from the bad [43], and that the positive and negative parts of the currency scale are ratio scales.

However, events can exist without the need to express them as numerical functions, and therefore one might want to develop uncertainty theories without such numerical

underpinnings—this is particularly true if one starts from symbolic logic. Moreover, there are situations where coherence may not be essential: for instance when one receives new information that contradicts old information, it may be desirable to let go of some of the initial information, and perform a revision step. The result of revision protects the input information but is no longer consistent with the prior one. At least it achieves minimal change of the prior information so as to accept the input one. A special case of this situation is theory revision of Gärdenfors and colleagues in the setting of classical logic ([1]) a setting subsumed by lower previsions. Another example will be Dempster's rule of conditioning that operates a revision but does not preserve coherence.

In the first part of this chapter, we start from epistemic logic nd derive the need for capturing imprecision in information representations. We then bridge the gap between uncertainty theories and epistemic logic showing that imprecise probabilities subsume modalities of possibility and necessity as much as probability. In fact this is the joint probabilistic and set-theoretic (and modal) nature of imprecise probabilities that makes this approach to uncertainty so rich. In the second part of this chapter, we present possibility and evidence theories, their origins, assumptions and semantics, discuss the connections between them and the general framework of imprecise probability. Finally, we point out the remaining discrepancies between the different theories regarding various basic notions, such as conditioning, independence or information fusion and the existing bridges between them.

1.1 Imprecise probability = modal logic + probability

Consider a variable X assuming values in a non-empty state space $\mathcal{X}$. The usual numerical representation of uncertainty consists of attaching to each statement of the form $x \in A \subset \mathcal{X}$ a degree of confidence $g(A)$ to each subset of $\mathcal{X}$. The confidence function thus defined is chosen to be a capacity (see Definition **??**).

The prototypical example of a confidence function is a probability measure. While the frequentist view of probability looks natural for repeatable events, the notion of subjective probability is more problematic for singular events in a situation of incomplete information [42]. A piece of information concerning X is said to be *incomplete* in a given context if it is not sufficient to answer relevant question in this context.

1.1.1 Boolean possibility theory and modal logic

The typical form of a piece of incomplete information is '$x \in E$' where E is a subset of $\mathcal{X}$ containing more than one element. An important remark is that elements in E are mutually exclusive, since X takes on a single value. Hence, a piece of incomplete information takes the form of a disjunction of mutually exclusive values. To emphasize this use of sets, we speak of *disjunctive* sets. Just like a subjective probability, a disjunctive set E represents the epistemic state of an agent about the ill-known value of X, and is attached to this agent.

This type of representation of incomplete information can be found in two areas:

classical logic and interval analysis. The two settings differ by the type of variable used to describe the state space $\mathcal{X}$: Boolean for logic, numerical for intervals. For instance, to say that *Peter is between 20 and 25 years old*, i.e., $x = age(Peter) \in E = \{20, 21, 22, 23, 24, 25\}$, is to suppose $x = 20$ or $x = 21$ or $x = 22$ or $x = 23$ or $x = 24$ or $x = 25$. On the real line, incomplete information is often of the form $a \leq X \leq b$, where $E = [a, b]$. In logic, information comes in the form of a belief base K containing several propositions ϕ_i expressed in some Boolean language. The corresponding epistemic state is the disjunctive set of models of K.

There exists confidence measures that are more natural than probability measures to handle this situation. From a consistent epistemic state $x \in E \neq \emptyset$, we can define two binary confidence functions, namely a *possibility measure* Π such that $\Pi(A) = 1$ if $E \cap A \neq \emptyset$ and 0 otherwise, and a *necessity measure* N such that $N(A) = 1$ if $E \subseteq A$ and 0 otherwise. It is clear that $\Pi(A) = 1$ if and only if proposition '$x \in A$' is not inconsistent with information item '$x \in E$', and that $N(A) = 1$ if and only if proposition '$x \in A$' is entailed by information item $x \in E$.

Basically, $\Pi(A) = 0$ means that A is impossible if '$x \in E$' is true. $N(A) = 1$ expresses that A is certain if '$x \in E$' is true. Moreover to say that A is impossible ($A \cap E = \emptyset$) is to say that the opposite event A^c is certain. So, functions N and Π are related by the conjugateness property:

$$N(A) = 1 - \Pi(A^c). \tag{1.1}$$

This is the Boolean version of possibility theory [37]. Possibilistic uncertainty is clearly at work in classical logic. If a set of formulas K in propositional logic has $E \subseteq \mathcal{X}$ as set of models (in this case, $\mathcal{X}$ is the set of interpretations of the language), and if proposition p is the syntactic form of proposition $x \in A$, then $N(A) = 1$ if and only if K implies p, and $\Pi(A) = 0$ if and only if $K \cup \{p\}$ is logically inconsistent.

However, in propositional logic, it cannot be expressed that $N(A) = 0$ nor $\Pi(A) = 1$. To do so, a modal logic is needed (Chellas [12]), that prefixes propositions with modalities such as possible ($\Diamond$) and necessary ($\Box$): in a modal belief base K^{mod}, $\Diamond p \in K^{mod}$ encodes $\Pi(A) = 1$, and $\Box p \in K^{mod}$ encodes $N(A) = 1$ (the latter is encoded by $p \in K$ in classical logic). The well-known conjugateness relation in modal logic reads: $\Diamond p = \neg \Box \neg p$.

Within this setting, possibility and necessity measures naturally satisfy the following equalities, respectively:

$$\Pi(A \cup B) = \max\{\Pi(A), \Pi(B)\} \tag{1.2}$$

$$N(A \cap B) = \min\{N(A), N(B)\} \tag{1.3}$$

besides $N(\mathcal{X}) = \Pi(\mathcal{X}) = 1$ and $\Pi(\emptyset) = N(\emptyset) = 0$. Possibility measures are said to be *maxitive*, while necessity measures are said to be *minitive*. These properties are taken as postulates even when possibility and necessity values lie in the unit interval $[0, 1]$ and are no longer Boolean. They also do not rely on a numeric setting, as taking the minimum and maximum of N and Π also makes sense when they range on a lattice [21]. In the modal logic setting, these axioms respectively read $\Diamond(p \vee q) = \Diamond p \vee \Diamond q$ and $\Box(p \wedge q) = \Box p \wedge \Box q$ and are well-known to hold in regular modal logics. A simple modal

logic for Boolean possibility theory can be devised from a propositional language. It corresponds to a fragment of the logic KD. Atoms of this logic are of the form $\Box p$, where p is any propositional formula. Well-formed formulas in this logic are obtained by applying standard conjunction and negation to these atoms. Models of such modal formulas are epistemic states: for instance E is a model of $\Box p$ means that $E \subseteq A$, where A is the set of models of p [2].

Recall (see Section **??**) that a possibility measure Π (and its dual necessity measure N) is entirely determined by the knowledge of a so-called possibility distribution $\pi : \mathcal{X} \to [0, 1]$ such that $\pi(x) := \Pi(\{x\})$. In the following, we extend this modal framework for possibility and certainty to more general numerical settings.

1.1.2 A unifying framework for capacity based uncertainty theories

All numerical representations of incompleteness-tolerant uncertainty have the common feature that the uncertainty of each event $A \subseteq \mathcal{X}$ is characterized by two, respectively upper and lower, evaluations. We shall respectively call them (adopting a subjectively biased language) degrees of *epistemic possibility* and *certainty* (in order to differentiate them from coherent lower and upper probabilities), that will be denoted Ep and Cer. Epistemic possibility refers to a lack of surprise and generalizes logical consistency (hence the possibility modality and set function Π), while certainty generalizes logical entailment (hence the necessity modality and set function N). These two degrees define monotonic confidence functions (capacities) from the frame $\mathcal{X}$ to $[0, 1]$ such that

$$\forall A \subseteq \mathcal{X}, \mathrm{Cer}(A) \leq \mathrm{Ep}(A). \tag{1.4}$$

They are supposed to be conjugate to each other, namely:

$$\forall A \subseteq \mathcal{X}, \mathrm{Cer}(A) = 1 - \mathrm{Ep}(A^c). \tag{1.5}$$

The first condition, Eq. (1.4), postulates that an event must be epistemically possible prior to being certain, and the second condition, Eq. (1.5), says that an event is all the more certain as its opposite is less epistemically possible. This framework has the merit of unambiguously encoding three extreme epistemic attitudes pertaining to event A:

- The case when A is certainly true: $\mathrm{Cer}(A) = 1$ (hence $\mathrm{Ep}(A) = 1, \mathrm{Ep}(A^c) = 0, \mathrm{Cer}(A^c) = 0$).

- The case when A is certainly false: $\mathrm{Ep}(A) = 0$ (hence $\mathrm{Cer}(A) = 0$).

- The case when the agent does not know if A is true or false: $\mathrm{Cer}(A) = 0$ and $\mathrm{Ep}(A) = 1$ (then $\mathrm{Ep}(A^c) = 1$; $\mathrm{Cer}(A^c) = 0$).

It is clear that probability measures, as well as Boolean possibility and necessity measures are special cases of this setting. In the case of probability measures, $\mathrm{Cer}(A) = \mathrm{Ep}(A) = P(A)$. Similarly, lower and upper probabilities on event A are important general examples of epistemic certainty ($\mathrm{Cer}(A) = \underline{P}(A)$) and of epistemic possibility

($\mathrm{Ep}(A) = \overline{P}(A)$), respectively. It is clear that super-additivity implies $g(A) + g(A^c) \leq 1$, so that a super-additive capacity is a certainty function Cer, while a sub-additive capacity is an epistemic possibility function.

1.2 From imprecise probabilities to belief functions and possibility theory

All set functions proposed for the representation of uncertainty obey the above conventions for epistemic certainty and possibility functions, hence extending the framework detailed in Section 1.1.1. As seen in Chapter **??**, such set functions can be interpreted (without considering conditioning) as special cases of coherent upper and lower probabilities. However, they also stand as proper extensions of Boolean possibility theory, as discussed in this section.

A first idea to extend the framework of Section 1.1.1 is to consider a probability distribution over the set of epistemic states. A second idea is to turn epistemic sets into fuzzy sets, allowing for degrees of possibility.

1.2.1 Random disjunctive sets

The approach adopted in the theory of evidence (Shafer [74]) is somewhat opposite to the one of the lower prevision school. Instead of augmenting the probabilistic approach with higher order uncertainty due to incompleteness, described by sets of probabilities or equivalent models, the idea is to add higher order probabilistic information to the disjunctive set approach to incompleteness. So, instead of a single representation of the form $x \in E$, a (generally) discrete probability distribution is defined over the various possible assertions of the form $x \in E$ (assuming a finite frame $\mathcal{X}$).

Mass functions

Let m be a probability distribution over the power set $\wp(\mathcal{X})$. The function m is called *mass assignment*, $m(E)$ is the belief mass allocated to the set E, and any subset E of $\mathcal{X}$ such that $m(E) > 0$ is called a *focal set*. Let $\mathcal{F} \subseteq \wp(\mathcal{X})$ be the collection of focal sets of a given m. Usually, no positive mass is assigned to the empty set ($m(\emptyset) = 0$ is assumed). However, the Transferable Belief Model, after Smets [77], does not make this assumption. In this case $m(\emptyset)$ represents the degree of internal conflict of the mass assignment. This is an essential difference with the theory of lower previsions, since the notion of coherence is central in the latter, while allowing for $m(\emptyset) > 0$ precisely means that some degree of incoherence can be accepted in the Transferable Belief Model. Indeed, the condition $m(\emptyset) = 0$ is a form of normalisation. As m is a probability distribution, the condition $\sum_{E \subseteq \mathcal{X}} m(E) = 1$ must hold anyway. Also note that the mass assignment m is here considered as the basic building block of the theory, while in Section **??** it is just the Möbius transform of a lower probability, whose positivity is a useful by-product of the ∞-monotone property.

In this hybrid representation of uncertainty, it is important to understand the meaning of the mass function, and it is essential not to confuse $m(E)$ with the probability of occurrence of event E.

- Shafer [74] says $m(A)$ is the belief mass assigned to E only and to none of its subsets.

- One may also see $m(E)$ as the amount of probability to be shared between elements of E without being assigned yet, by lack of knowledge. This is in agreement with a frequentist view of the mass function and the idea of incomplete observation[1].

- An explanation in the subjective line consists in saying that $m(E)$ is the probability that the subject only knows that $x \in E$. So, there is an epistemic modality implicitly present in $m(E)$, but absent from $P(A)$ (see [2] for the connection between modal logic and mass assignments).

These views explain why the function m is not required to be inclusion-monotonic. It is allowed to have $m(E) > m(F) > 0$ even if $E \subset F$, when the subject is sure enough that what is known is of the form $x \in E$. In particular, $m(\mathcal{X})$ is the probability that the agent is completely ignorant about the value of a variable X. We have already mentioned in Section **??** a number of situations from which mass functions can arise, the most well-known being a multivalued mapping between a probability space and a space $\mathcal{X}$, as proposed by Dempster [26].

Note that m can be formally associated to a random set such that $m(E)$ is the probability of observing the (random) disjunctive set E. However, in Dempster, Shafer and Smets views, this is a random disjunctive set, in contrast with other views of random sets [53, 61], where sets are conjunctions of their elements, i.e., they represent real objects (like a region in a land). This difference is crucial when considering derived set-functions and operations such as conditioning, which may differ in both situations.

Belief and plausibility functions

A mass assignment m induces two set-functions, respectively a belief function Bel and a plausibility function Pl, defined by:

$$\text{Bel}(A) = \sum_{E \subseteq A, E \neq \emptyset} m(E); \tag{1.6}$$

$$\text{Pl}(A) = \sum_{E \cap A \neq \emptyset} m(E). \tag{1.7}$$

When $m(\emptyset) = 0$, it is clear that $\text{Bel}(\mathcal{X}) = \text{Pl}(\mathcal{X}) = 1$, $\text{Pl}(\emptyset) = \text{Bel}(\emptyset) = 0$, and $\text{Bel}(A) = 1 - \text{Pl}(A^c)$ so that these functions are another example of certainty (Cer = Bel) and epistemic possibility (Ep = Pl).

---

[1]See also the construction in the context of data imprecision, most notably pages **??** and **??**

At the interpretive level, Bel and Pl can be viewed as direct extensions of Boolean necessity and possibility functions respectively. They are recovered when $m(E) = 1$ for some focal subset E. In fact, $\mathrm{Bel}(A)$ is an expected necessity degree (in the sense of m), or a probability of provability [67]   (the probability that m logically implies A). Likewise, $\mathrm{Pl}(A)$ is an expected possibility degree (in the sense of m), or a probability of consistency (the probability that m is logically consistent with A). In terms of modal operations $\square, \lozenge$ introduced earlier, it is clear that $\mathrm{Bel}(A) = P(\square A)$ and $\mathrm{Pl}(A) = P(\lozenge A)$.

Belief functions and lower probabilities

If $m(\emptyset) = 0$, belief functions coincide with ∞-monotone capacities. However, if $m(\emptyset) > 0$ then $\mathcal{M}(\mathrm{Bel}) = \emptyset$ and the connection with imprecise probability theory is lost.

It is also important to note that, when a mass assignment m is generated by a random variable Y imprecisely observed through a multivalued mapping $\Gamma : \mathcal{Y} \to \wp(\mathcal{X})$, the credal set $\mathcal{M}(\mathrm{Bel})$ is usually less informative than Γ [63]. In that case, the belief function is viewed as a disjunctive set of random variables on $\mathcal{X}$, each being defined by a selection function $\phi : \mathcal{Y} \to \mathcal{X}$ of Γ. So, for each element $y \in \mathcal{X}$, the probability mass $P(\{y\})$ should be entirely transferred to one element $x = \phi(y) \in \Gamma(y)$, and the resulting set of probabilities $\mathcal{M}(\Gamma) = \{P_\phi : \phi \in \Gamma\}$ is not generally convex. In contrast, when building the credal set $\mathcal{M}(\mathrm{Bel})$, masses on focal subsets can be redistributed arbitrarily among the elements of the focal sets, and $\mathcal{M}(\Gamma) \subset \mathcal{M}(\mathrm{Bel})$.

Example 1.1. Consider a set of ages discretised in periods of 5 years ($\mathcal{Y} = \{y_1 = [16, 20], y_2 = [21, 25], \ldots\}$). We are interested in the knowledge on the precise age X of Pierre ($\mathcal{X} = \{15, \ldots, 100\}$). All we know is that *Pierre is between 21 and 25 years old*. Y is reduced to the Dirac measure $p(\{y_2\}) = 1$, with $\Gamma(y_2) = A = \{21, \ldots, 25\}$. The set $\mathcal{M}(\Gamma)$ corresponds to the 5 Dirac probabilities focusing on singletons in A. The corresponding credal set contains all probability measures with support within A. ♦

However, as the convex hull of $\mathcal{M}(\Gamma)$ equals $\mathcal{M}(\mathrm{Bel})$, they have the same behavioral consequences in the theory of lower previsions, i.e., expectations of any function $f : \mathbb{R} \to \mathcal{X}$ have the same lower and upper bounds for $\mathcal{M}(\Gamma)$ and $\mathcal{M}(\mathrm{Bel})$. However, this equality may not hold for other (non-linear) functionals like variance. In the above example, the set of possible variances of probability distributions in $\mathcal{M}(\Gamma)$ is reduced to the singleton $\{0\}$, while it is not so for distributions in $\mathcal{M}(\mathrm{Bel})$ (for instance, the uniform distribution, which has a non-null variance, is in $\mathcal{M}(\mathrm{Bel})$). See [16] for a discussion about variance in Dempster-Shafer theory.

1.2.2   Numerical Possibility Theory

The roots of possibility theory can be traced back to the degree of potential surprise introduced by the economist Shackle [73], and its revival is mainly due to Zadeh [87] (starting from the representation of linguistic information). Other scholars dealing with various aspects of possibility theory are Lewis [59], Dubois and Prade [36, 41, 34], and De Cooman [21]. In this section, we focus on the numerical representation of possibility and necessity.

Possibility distributions

The basic idea of possibility theory is to refine epistemic states $X \in E$ by introducing degrees of epistemic possibility of elements. The disjunctive set E is replaced by a fuzzy set, namely by a mapping $\pi \colon \mathcal{X} \to L$, where L is a possibility scale. L must be at least a complete lattice bounded from above and from below by 1 and 0 respectively; more often it is a totally ordered set, for instance the unit interval. It is generally assumed that $\pi(x) = 1$ for some x. This is not always required, meaning that the theory permits the presence of some inconsistency degree, as for the case of belief functions. More generally, following Shackle's ideas, $1 - \pi(x)$ is the amount of surprise that $X = x$, equating impossibility to full surprize. A possibility distribution is a purely epistemic construction representing in a graded way the knowledge of an agent.

This kind of representation has been found natural in the qualitative case where the possibility scale is not numerical (only an ordering of more or less possible values of X is useful in this case); this trend was initiated by Lewis [59] and pursued in the literature of non-monotonic reasoning and belief revision [30].

Zadeh [87] considers that natural language propositions modelled by fuzzy sets generate possibility distributions on domains of interest. For instance, the statement *John is tall* involves the gradual predicate *tall* delimiting an elastic constraint on the set of human heights $\mathcal{X}$, modelled by the membership function $\mu_{tall} \colon \mathcal{X} \to [0, 1]$. Then the piece of information *John is tall* defines the possibility distribution $\pi = \mu_{tall}$. Beyond these original motivations, possibility distributions seem to be ubiquitous in uncertainty theories, in connection with probability theory as seen later on.

Possibility and necessity measures

From the knowledge of π ranging on the unit interval, we can define two confidence functions, namely a *possibility measure* Π, already encountered in the previous chapter **??** (see section **??**) such that

$$\Pi(A) = \sup_{x \in A} \pi(x). \tag{1.8}$$

and a *necessity measure N* such that

$$N(A) = \inf_{x \notin A} 1 - \pi(x). \tag{1.9}$$

The possibility distribution π is here the primitive concept from which are induced the two measures Π and N. Again, $\Pi(A) = 1$ measures the degree of consistency of proposition '$x \in A$' with information item π, and $N(A)$ is a degree of logical entailment of proposition '$x \in A$' by information item π, thus extending the Boolean version of possibility theory [37]. The same conventions for extreme values of Π and N are valid: in fact, they are yet another example of certainty (Cer $= N$) and epistemic possibility (Ep $= \Pi$).

Outside their interpretation as specific cases of ∞-monotone capacities, another interpretation of numerical possibility distributions is the likelihood function in non-Bayesian statistics (Smets [75], Dubois et al. [33]). In the framework of an estimation

problem, one is interested in determining the value of some parameter $\theta \in \Theta$ that defines a probability distribution $P(\cdot \mid \theta)$ over $\mathcal{X}$. Suppose that we observed the event A. Then $L(\theta) := P(A \mid \theta)$, as a function of $\theta \in \Theta$, is not a probability distribution, but a likelihood function[2]: a value θ is considered as being all the more plausible as $P(A \mid \theta)$ is higher, and the hypothesis θ will be rejected if $P(A \mid \theta) = 0$ (or is below some threshold).

Often, this function is renormalized so that its maximum is 1. We are allowed to let $\pi(\theta) := P(A \mid \theta)$ (thanks to this renormalisation) and to interpret this likelihood function in terms of possibility degrees. In particular, it can be checked that for all $B \subseteq \Theta$, bounds for the value of $P(A \mid B)$ can be computed as:

$$\inf_{\theta \in B} P(A \mid \theta) \leq P(A \mid B) \leq \sup_{\theta \in B} P(A \mid \theta)$$

The reason for these inequalities is that $P(A|B)$ is a weighted arithmetic mean of the $P(A|\theta)$ where the weights are proportional to (unknown) prior probabilities attached to the parameter values $\theta \in B$. It shows that adopting the maxitivity axiom (Eq. (1.2)) corresponds to an optimistic computation of $P(A \mid B) = \Pi(B)$. It is easy to check that letting $P(A \mid B) := \sup_{\theta \in B} P(A \mid \theta)$ is the only way for building a confidence function about Θ from $P(A \mid \theta), \theta \in \Theta$. [3] Indeed, the monotonicity w.r.t. inclusion of the likelihood function $L(\cdot) = P(A \mid \cdot)$ forces $P(A \mid B) \geq \sup_{\theta \in B} P(A \mid \theta)$ to hold [13]. Also note that the maximum likelihood principle originally due to Fisher, which consists in choosing the value $\theta = \theta^*$ that maximizes $P(A \mid \theta)$, is in total agreement with possibility theory.

1.2.3   Overall picture

We can now draw a picture (Figure 1.1) of the different imprecision tolerant basic uncertainty representations and their formal relationships. It does not reflect the variety of existing viewpoints but it helps to situate the different theories in a common framework.

Once we have Figure 1.1 in mind, notions such as set intersection and unions then needs to be extended to possibility theory, random sets, lower previsions, etc. Similarly, probabilistic notions such as conditioning or independence modelling needs to be properly and meaningfully extended to other frameworks. Since the new frameworks are usually richer than basic ones, such extensions are usually not unique, whence the apparitions of discrepancies that needs to be understood and reconciled across the different theories.

1.3   Discrepancies between uncertainty theories

The generality of desirable gambles and lower previsions may suggest that the apparent diversity of uncertainty theories is delusive, and that a single formalism is at work, the choice of a representation framework depending on practical considerations. However this is not the end of the story.

---

[2]see also **??**
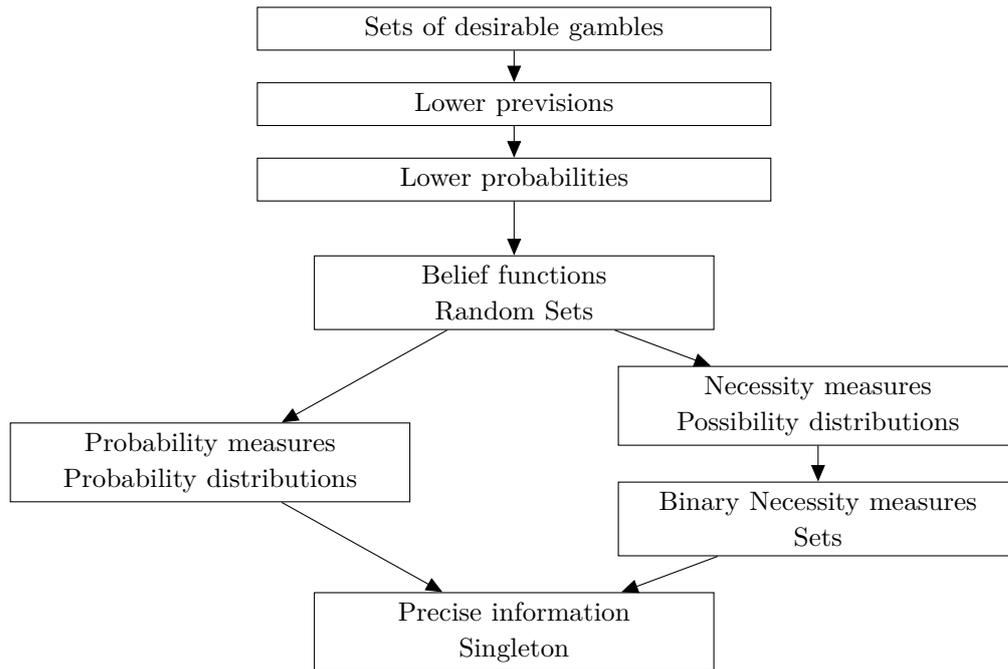
[3]See also the profile-likelihood approach in section **??**

```
┌─────────────────────────────────┐
│      Sets of desirable gambles  │
└─────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│         Lower previsions        │
└─────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│        Lower probabilities      │
└─────────────────────────────────┘
              │
              ▼
┌─────────────────────────────────┐
│          Belief functions       │
│           Random Sets           │
└─────────────────────────────────┘
```

Figure 1.1: Summary of uncertainty theories and basic representations

Boxes: Sets of desirable gambles → Lower previsions → Lower probabilities → Belief functions / Random Sets → (Probability measures / Probability distributions) and (Necessity measures / Possibility distributions → Binary Necessity measures / Sets) → Precise information / Singleton

On the one hand, each uncertainty theory has its own history and was motivated by different problems, so a proper unification also needs to figure out whether these various stories fit together. Already in standard probability theory, there is more than one story leading to a probability measure: noticeably, the frequentist and the subjectivist stance [47]. Each one leads to a different definition of probability (for instance, sigma-additivity, central in Kolmogorov approach, is not part of the subjectivist approach after De Finetti). This distinction between frequentist and subjectivist views is still present across all uncertainty theories.

On the other hand, each uncertainty theory can be seen as a language for describing and solving problems. These languages rely on primitive notions that are used in all subsequent definitions, and these are not equally expressive. To summarize, one may say that the basic building block in possibility theory is the possibility distribution; in evidence theory it is the Möbius mass function; in Walley theory the basic building block is the coherent lower prevision, to which can be associated a corresponding credal set. In each theory, everything is defined by means of the primitive building block. So there will be some discrepancies on important concepts, such as conditioning or fusion operations. This section provides some examples of such discrepancies.

1.3.1   Objectivist vs. subjectivist standpoints

A very important distinction must be drawn between singular and generic information. It underlies the opposition between frequentist and subjectivist sides of uncertainty theories, but is in some sense more fundamental. *Generic* information refers to a collection of cases or a population of situations. It could be a physical law, a statistical model built from a sample of observations, or a piece of commonsense knowledge like "birds fly". A probability distribution can be used to represent the behaviour of a population, and has then a frequentist flavor. One can then speak of the true (or the best) probability distribution underlying this population. A credal set may represent incomplete knowledge of this best model and it then underlies a currently unreacheable probability measure induced by this population. However, the population need not even be well-defined for the concept of generic information to make sense.

In contrast, *singular* information refers to a particular situation, the occurrence of a single event: for instance, an observation (a patient has fever at a given time point), or a testimony (the crazy driver's car was blue). A probability degree attached to a piece of singular information is a degree of belief, and is not necessarily related to frequentist information. The idea of probability degrees justified by betting odds precisely aims at attaching probabilities to unique events, bypassing the need to resort to frequentist notions. This is the spirit of Walley's lower previsions. However an agent often has some experience about the world, and possesses some generic knowledge, be it imprecise. It has been argued that one can use generic knowledge on the frequency of an event to derive a degree of belief in the singular occurrence of this event in a specific situation (Hacking [50]). This reasoning looks valid for imprecise probabilities as well.

Evidence theory after Shafer clearly stands on the subjectivist side, degrees of belief being derived from the perceived reliability of testimonies about singular events. Like in Walley's approach, a belief function is a full-fledged representation of belief, and Smets has insisted that referring to an ill-known probability degree $P(A)$ lower bounded by $Bel(A)$ made no practical sense in the approach. On the contrary Dempster's view is fully compatible with a frequentist interpretation of belief functions. Indeed, if the underlying probability space $\mathcal{Y}$ reflects statistical information, the corresponding belief function is a lower probability having a frequentist flavour. So the random set underlying a belief function may have frequentist or subjectivist interpretations.

Similar comments make sense for possibility distributions as well. It has been shown that while they can be built from subjective linguistic information as suggested by Zadeh, they are present in frequentist statistics, in the form of nested prediction, dispersion or confidence intervals or likelihood functions. But a possibility measure can serve as a special type of upper prevision in the subjectivist view as well [22].

1.3.2   Discrepancies in conditioning

Most of the time, a probability distribution refers to a population. This is a form of generic information, typically frequentist. This information can be used for inferring beliefs about a particular situation for which we have incomplete but clear-cut observations. This is

called prediction. In this framework, $P(A \mid C)$ is the (frequentist) probability of event A in context C and it can quantify the belief of the agent in A when it has been observed that C is true in the current situation and nothing else (assuming that the situation is typical of environment C). The degree of belief of the agent in context C has changed from $P(A)$ to $P(A \mid C)$. Conditioning here is used for updating the beliefs of the agent about the current situation by exploiting (but not modifying) generic information. One only applies the available generic knowledge to a reference class C, what is called *focusing* by Dubois, Moral, and Prade [32]. For instance, the probability measure P represents medical knowledge (often compiled under the form of a Bayesian network). The singular information item C represents the results of tests for a patient. $P(A \mid C)$ is the probability of having disease A for patients for whom C has been observed; this value also estimates the singular probability (belief) that this patient has this disease.

When probability P is subjective, it has a singular nature (when betting on the occurrence of a non-repeatable event). In this case conditioning can be interpreted as the revision of the singular probability function *by a piece of information of the same nature*. In this case, information C is interpreted as $P(C) = 1$, which represents a constraint that has to be taken into account when revising P. For instance [42], in a criminal affair where the investigator suspects Peter, Paul and Mary with probabilistic confidence degrees $\frac{1}{4}$, $\frac{1}{4}$ and $\frac{1}{2}$ respectively, and then learns that Peter has an alibi, i.e. $P(\{\text{Mary, Paul}\}) = 1$. We then have to revise these singular probabilities. The use of conditional probability for handling this revision of the probabilities is often proposed (and justified by the minimal change principle, already mentioned above), which yields probabilities $\frac{1}{3}$ and $\frac{2}{3}$ for Paul and Mary respectively. However, the problem of revising P is different from the one of updating singular beliefs on the basis of generic information.

In the case where the generic knowledge of the agent is represented by imprecise probabilities, Bayesian prediction is generalized by performing a sensitivity analysis on the conditional probability. Let $\mathcal{M}$ be a generic credal set on $\mathcal{X}$. For each proposition A a lower bound $\underline{P}(A)$ and an upper bound $\overline{P}(A)$ of the probability degree of A are known. In the presence of singular observations summarized under the form of a context C, the belief of an agent in a proposition A could be represented by the interval $[\underline{P}(A \mid C), \overline{P}(A \mid C)]$ defined by

$$\underline{P}(A \mid C) = \inf\{x : P(A \cap C) = x \cdot P(C), P \in \mathcal{M}\}$$
$$\overline{P}(A \mid C) = \sup\{x : P(A \cap C) = x \cdot P(C), P \in \mathcal{M}\}.$$

which is the conditioning rule obtained by means of the natural extension of Section **??**. It may happen that the interval $[\underline{P}(A \mid C), \overline{P}(A \mid C)]$ is larger than $[\underline{P}(A), \overline{P}(A)]$, which corresponds to a loss of information (a dilation in the sense of Seidenfeld and Wasserman [72]) in specific contexts. This property reflects the idea that the more singular information is available about a situation, the less informative is the application of generic knowledge to it (since the number of relevant statistical data, i.e. where C is true, may become very small). We see that this form of conditioning does not correspond at all to the idea of enriching generic information, it is only a matter of querying it. Indeed the input information pertains to a singular situation, and focuses generic knowledge (the credal set) on a specific subset of possible worlds, on which the credal set may be totally

uninformative (typically if the lower probability of C is zero, that is when nothing supports C in the credal set, we get $[\underline{P}(A \mid C), \overline{P}(A \mid C)] = [0, 1]$).

If belief and plausibility functions are used as generic knowledge, $m(E)$ represents, for instance, the proportion of imprecise results, of the form $x \in E$, in a statistical test on a random variable X. In this framework, prediction in context C consists in evaluating the weight function $m(\cdot \mid C)$ induced by the mass function m on the subset C of states, taken as the new frame. Three cases should be considered:

- $E \subseteq C$: In this case, $m(E)$ remains assigned to E.

- $E \cap C = \emptyset$: In this case, $m(E)$ no longer matters and is eliminated.

- $E \cap C \neq \emptyset$ and $\overline{E} \cap C \neq \emptyset$: In this case, some fraction $\alpha_E \cdot m(E)$ of $m(E)$ remains assigned to $E \cap C$ and the rest, i.e. $(1 - \alpha_E) \cdot m(E)$, is allocated to $\overline{E} \cap C$. But this sharing process is unknown.

The third case corresponds to incomplete observations E that neither confirm nor disconfirm C. The scalar value α_E is the proportion of the population for which C is true among those observations for which all that is known is that E holds. We have not enough information in order to know if, for each of these observations, C is true or not, since only E is known and it overlaps both C and its complement. Suppose that the values $\{\alpha_E, E \subseteq S\}$ were known. Note that $\alpha_E = 1$ and $\alpha_E = 0$ in the first and second cases, respectively. Then, we can build a mass function $m_\alpha^C(\cdot)$. Note that a renormalisation of this mass function is necessary, in general, as soon as $\mathrm{Pl}(C) < 1$ (letting $m_\alpha(\cdot \mid C) = \frac{m_\alpha^C(\cdot)}{Pl(C)}$). If one denotes by $\mathrm{Bel}_\alpha(A \mid C)$ and $\mathrm{Pl}_\alpha(A \mid C)$ the belief and plausibility functions obtained by focusing on C, based on the allocation vector α, the conditional belief and plausibility degrees on C are defined by [20]:

$$\mathrm{Bel}(A \mid C) = \inf_\alpha Bel_\alpha(A \mid C), \qquad (1.10)$$

and

$$\mathrm{Pl}(A \mid C) = \sup_\alpha Pl_\alpha(A \mid C). \qquad (1.11)$$

One still obtains belief and plausibility functions (Jaffray [52]), and necessity and possibility measures if we start with such measures (Dubois and Prade [40]). The following results show that what is obtained is a generalization of Bayesian inference [45]: if $\mathrm{Bel}(C) > 0$, then

$$\mathrm{Bel}(A \mid C) = \inf\{x : P(A \cap C) = x \cdot P(C), P \geq \mathrm{Bel}\} = \frac{Bel(A \cap C)}{\mathrm{Bel}(A \cap C) + \mathrm{Pl}(\overline{A} \cap C)},$$

$$\mathrm{Pl}(A \mid C) = \sup\{x : P(A \cap C) = x \cdot P(C), P \geq \mathrm{Bel}\} = \frac{Pl(A \cap C)}{\mathrm{Pl}(A \cap C) + \mathrm{Bel}(\overline{A} \cap C)}.$$

It corresponds again to Walley's updating rule for upper and lower probabilities in Section **??** based on the natural extension. It is easy to see that $Pl(A \mid C) = 1 - Bel(\overline{A} \mid C)$, and that these formulas generalize probabilistic conditioning. This closed form already

appears in the first report by Walley [79], and in [20, 45]. It is present with a different notation in the seminal paper of Dempster [26]. Jaffray [52] expresses this conditioning in terms of mass functions, laying bare the Möbius transform of $\text{Bel}(\cdot \mid C)$ in terms of m.

Note that if $\text{Bel}(C) = 0$ and $\text{Pl}(s) > 0, \forall s \in C$ (no value in C is impossible but C has no certainty) then all the focal sets of m either overlap (possibly contain) C or are disjoint from C, none being contained in it. In this case, from equations (1.10) and (1.11), $Bel(A \mid C) = 0$ and $\text{Pl}(A \mid C) = 1$ for all non-trivial (i.e. $A \neq C$ and $A \neq \emptyset$) events $A \subset C$: one cannot infer anything in context C. However, if $\text{Pl}(A) = 0$ for some $A \subset C$, then $\text{Pl}(A \mid C) = 0$ as well. So this kind of conditioning usually expressed using credal sets, can be couched in terms of mass functions as well.

However the usual type of conditioning in evidence theory, called 'Dempster rule of conditioning', proposed by Shafer [74] and Smets [77], systematically assumes $\alpha_E = 1$ as soon as $E \cap C \neq \emptyset$ in the above construction. It supposes a transfer of the full mass of each focal set E to $E \cap C \neq \emptyset$ (followed by a renormalisation). This means that we interpret the new information C as allowing us to restrict to probability functions such that $P(C)$ is maximal: situations where C is false are considered as impossible. If one denotes by $\text{Pl}(A \parallel C)$ the plausibility function based on this assumption, we have:

$$\text{Pl}(A \parallel C) = \frac{\text{Pl}(A \cap C)}{\text{Pl}(C)}.$$

This constitutes another generalisation of probabilistic conditioning. The conditional belief is then obtained by duality $\text{Bel}(A \parallel C) = 1 - \text{Pl}(\overline{A} \parallel C)$. With this conditioning, the size of focal sets diminishes, thus information becomes more precise, and when $\text{Pl}(C) = 1$, the intervals $[\text{Bel}, \text{Pl}]$ become tighter, in the sense that $[\text{Bel}(A \parallel C), \text{Pl}(A \parallel C)] \subseteq [\text{Bel}(A), \text{Pl}(A)]$ for all $A \subseteq C$ (they are also always tighter than those obtained by focusing [56]).

Dempster rule of conditioning thus corresponds to a process where information is enriched, or more generally significantly altered, which contrasts with focusing. In particular, when $\text{Pl}(C) < 1$ this kind of conditioning violates coherence in the sense of Walley, contrary to the previous one [81].

However, if the credal set represents singular uncertain evidence, this conditioning performs a revision of the underlying credal set. The incurred lack of coherence then is not surprising. When revising our beliefs, if the new information is at odds with previous beliefs, it can be expected that the new beliefs be conflicting with the previous one, not that it be coherent with it. On the contrary, the focusing conditioning operation for prediction of singular events out of generic knowledge preserves coherence because the prediction is a consequence of the credal set, and the latter is never altered, just exploited in a subset of $\mathcal{X}$.

In the more general framework of imprecise probabilities, a blind application of the revision process by a piece of information C consists in adding the supplementary constraint $P(C) = 1$ to the family $\mathcal{M}$, i.e.

$$\underline{P}(A \parallel C) = \inf\{P(A \mid C), P(C) = 1, P \in \mathcal{M}\};$$

$$\overline{P}(A \mid\mid C) = \sup\{P(A \mid C), P(C) = 1, P \in \mathcal{M}\}.$$

If $\overline{P}(C) = (1)$ it coincides with Dempster rule of conditioning (and then, coherence is preserved). But clearly, it may happen that the set $\{P \in \mathcal{M}, P(C) = 1\}$ is empty (it is always the case in the classical Bayesian framework since $\mathcal{M}$ is a singleton). Dempster rule of conditioning then corresponds to applying the maximal likelihood principle (Gilboa and Schmeidler [49]) and we replace the condition $P(C) = 1$ by $P(C) = \overline{P}(C)$ in the above equation. In this way, we generalize Dempster rule (which is recovered if $\overline{P}$ is a plausibility function) to 2-monotone lower probabilities (unless $\overline{P}(C) = 0$).

This type of conditioning has nothing to do with focusing. Indeed, Shafer and Smets view the Möbius mass function m only as singular information, and no generic knowledge is taken into account in the picture[77]. In the crime example (originally due to Smets), suppose that the organizer of the crime tossed a coin for deciding whether a man or a woman is recruited to be the killer. This piece of uncertain singular information is represented by the mass function $m(\{\text{Peter, Paul}\}) = \frac{1}{2}$ (there is no information available about Peter alone and Paul alone), and $m(\{\text{Mary}\}) = \frac{1}{2}$. Now, if we learn that Peter has an alibi, the focal set $\{\text{Peter, Paul}\}$ reduces to $\{\text{Paul}\}$ and we deduce after revision, that $P(\{\text{Mary}\}) = P(\{\text{Paul}\}) = \frac{1}{2}$. We have seen that the Bayesian approach shares the mass $m(\{\text{Peter, Paul}\})$ equally between Peter and Paul, so that Bayesian conditioning then yields $P(\{\text{Mary}\}) = 2 \cdot P(\{\text{Paul}\}) = \frac{2}{3}$. This result may sound debatable when dealing with uncertain singular pieces of information (let alone at a court of law).

As seen from the above discussion, discrepancies between belief functions and lower previsions concerning conditioning can be solved to a large extent. Shafer and Walley use different conditioning rules in their theories because they deal with different kinds of problems. Once it is acknowledged that Walley's conditioning is tailored for prediction, and Dempster rule of conditioning is tailored for revision of singular uncertain information, some kind of reconciliation can take place. In the case of possibility measures, both forms of conditioning make sense and preserve consonance [40, 82].

1.3.3   Discrepancies in notions of independence

Independence plays an important role in uncertainty theories, both for practical and theoretical reasons[4]. When dealing with precise probabilities, independence can be modelled in two conceptually different but formally equivalent ways. Consider two variables X, Y assuming their values in possibility spaces $\mathcal{X}, \mathcal{Y}$. Let P_X, P_Y be the marginal probability measures on $\mathcal{X}, \mathcal{Y}$, and P be the joint probability measure modeling our uncertainty about the value that (X, Y) assume jointly. According to the first point of view, Y is independent of X if learning the value of Y does not change the uncertainty about X, i.e., if

$$P_X(A \mid B) = P_X(A), \ \forall A \subseteq \mathcal{X}, B \subseteq \mathcal{Y} \quad [\text{marginalization}]. \tag{1.12}$$

---

[4]See also the discussion in section **??**.

We call this definition *causal* independence, since what is expressed here is the absence of (direct or indirect) causal influence of Y on X (this view is called *epistemic* in Chapter **??**). Note that Eq. (1.12) looks asymmetric. Only the axioms of probability theory make it symmetric when P is non-null. The second point of view, called here *decompositional* independence (and called *formalist* in Chapter **??**), requires the joint model to be computable from the marginals alone. This condition is expressed by

$$P(A, B) = P_X(A) \cdot P_Y(B), \ \forall A \subseteq \mathcal{X}, B \subseteq \mathcal{Y} \quad \text{[factorisation]}, \quad (1.13)$$

which is symmetric by definition. Eqs. (1.12) and (1.13) are equivalent when P is assumed to be non-null everywhere, and correspond to stochastic independence. The two approaches are no longer equivalent in the other uncertainty theories, and they give rise to different independence notions and joint models. Also, the meaning of Eq. (1.12) depends on the choice of a conditional capacity $g(A \mid B)$, hence the corresponding causal independence notion depends on the considered conditioning operation. Since probabilistic conditioning has multiple generalisations, Eq. (1.12) gives rise to different notions.

The extension of Eq. (1.12) to evidence theory has been investigated by Ben Yaghlane et al. [5]. Let $m(\cdot \mid\mid B)$ denote the mass distribution obtained by applying Dempster's conditioning to m, and m_X denote the marginal mass on $\mathcal{X}$ obtained for any $A \subseteq \mathcal{X}$ by $m_X(A) = \sum_{B \subseteq \mathcal{Y}} m(A \times B)$. In this setting, Y is said to be causally independent from X if the following condition

$$m_X(A \mid\mid B) = m_X(A), \ \forall A \subseteq \mathcal{X} \quad (1.14)$$

holds for every $B \subseteq \mathcal{Y}$, where $m_X(\cdot \mid\mid B)$ is the mass obtained by conditioning m and marginalizing it on $\mathcal{X}$. This condition turns out to be symmetric and to be equivalent to Shafer's definition of *cognitive independence* that requires the joint plausibility measure of the Cartesian product of events be the product of marginal plausibilities.

The extension of Eq. (1.12) to possibility theory has been studied by De Campos and Huete [19]. In this case, one starts from a joint possibility distribution π defined over $\mathcal{X} \times \mathcal{Y}$. Causal independence of Y from X comes down to stating that for every $x \in \mathcal{X}$, $y \in \mathcal{Y}$, we have

$$\pi(x \mid\mid y) = \pi_X(x) \quad (1.15)$$

where $\pi_X(x) = \max_{y \in \mathcal{Y}} \pi(x, y)$ is the marginal possibility distribution on X and conditional possibility is defined by

$$\pi(x \mid\mid y) = \frac{\pi(x, y)}{\pi_Y(y)}. \quad (1.16)$$

Again, this independence notion turns out to be symmetric, and corresponds to requiring that the joint possibility distribution be the product of the marginal distributions. Other extensions using other T-norm operators instead of the product have been studied [8, 18], however, such extensions, e.g. using minimum instead of product, can suffer from continuity problems in the numerical setting [21]. The problem of characterising the most

conservative joint model that has some given marginals and satisfies Conditions (1.14) and (1.16) is still open. Likewise, the causal independence notion associated to the prediction conditioning [45] based on *focusing* [32] has received little attention in the frameworks of possibility theory and evidence theory.

This is in contrast with the theory of lower previsions, where it corresponds to the well-studied notions of epistemic irrelevance and independence presented in Chapter **??**, but where extensions of Conditions (1.14) and (1.16) to more general frameworks than belief functions (i.e. coherent lower probabilities and previsions) have not been investigated. This may be explained by the fact that neither n-monotonicity nor supremum-preserving properties of marginals are preserved in the most conservative joint model obtained with such an assumption. For example, Miranda and De Cooman [64] have attempted to embed epistemic independence in possibility theory, reaching the conclusions that the limited expressiveness of possibility distributions may not be sufficient to model epistemic independence. Extensions of Eq. (1.13) are usually more straightforward and mostly consists in applying product rules to the basic building block of each theory. In evidence theory, the extension applies the product rule to the marginal mass distributions m_X, m_Y. This comes down to building the joint model m over $\mathcal{X} \times \mathcal{Y}$ such that

$$m(A \times B) = m_X(A)m_Y(B). \tag{1.17}$$

Shafer called this kind of independence *evidential*, but it is often called random set independence. It expresses independence between sources informing about X and Y, respectively. In possibility theory, the obtained joint model [19] from π_X, π_Y using the product rule is simply

$$\pi(x, y) = \pi_X(x)\pi_Y(y). \tag{1.18}$$

and is equivalent to Eq. (1.15) when possibility degrees are positive. The decompositional independence in the framework of lower previsions, called *strong independence*, is detailed in Chapter **??**. There exist some formal links between the three notions, since the credal set induced by the possibility distribution (1.18) is included in the credal set induced by *strong independence*, while the latter credal set is included in the one induced by the mass distribution given by Eq. (1.17) (see Fetz [46], for example). Finally, it is worth mentioning that the so-called non-interactive [87] joint possibility distribution $\pi_\wedge$ that has the following form

$$\pi_\wedge(x, y) = \min(\pi_X(x), \pi_Y(y)) \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \tag{1.19}$$

does not have strong connections with the previously mentioned joint models [78]. This is due to the fact that it assumes a complete dependence between the sources supplying the possibility distributions (via the choice of the same α-cuts for both sources) instead of independence between them [5], and no knowledge on the dependence between X and Y [3].

Discrepancies between independence notions across the different theories appear more severe than for conditioning. However, works attempting to fill in the gaps between

---

[5]in which case random set independence is retrieved.

these notions are more recent than for conditioning [17]. As mentioned previously, what form would causal independence take using Dempster conditioning for lower previsions is still unknown. Studies linking the focusing rule with a causal independence notion in possibility and evidence theory should be carried out. Discrepancies among decompositional independence notions are less severe, and the main interest of such a notion is computational.

1.3.4 Discrepancies in fusion operations

Fusion is the operation consisting in merging uncertain information from several sources. This notion emerged first in the theory of belief functions, where the rule of combination originally due to Dempster plays a major role. In probability theory, this question is seldom addressed, due to the fact that it has been traditionally envisaged (since the XIXth cenrury) from a statistical and frequentist point of view, despite the fact that the problem of merging testimonies (i.e., non-statistical evidence) motivated early research in probability back in the XVIIth and XVIIIth centuries [68].

In subjective probability, one is often concerned with a single agent, however the problem of modeling expert opinions in reliability has prompted probabilists to address the issue of fusion of subjective probabilities supplied by experts [15]. Walley [80] himself devoted a long report to this problem.

Considering that imprecise probabilities generalise both probabilities and sets, the two building blocks of uncertain information fusion are the fusion of probabilities and the fusion of sets. There are basically two approaches to merging probabilities

- The non-Bayesian approach [15] supposes that n sources (experts) supply n subjective probability distributions P_i pertaining to the location of a variable x. Reliability weights $r_i \in [0, 1]$ such that $\sum_{i=1}^{n} r_i = 1$ are attached to each source. Then the merging rule prescribes to consider the convex mixture $\sum_{i=1}^{n} r_i P_i$ as the result of the fusion. This form of fusion is basically the only possible way of retrieving a probability measure as the result [62]. It presupposes each expert is a kind of random source and the fusion can be seen as the direct extension of the statistical view relying on identical, yet independent processes

- On the contrary the Bayesian view considers a holistic approach to the fusion, whereby there is a joint probability that each source produces the output x_i when the real value of the quantity under concern is x. The fusion tries to compute the probability distribution $P(x|x_1, x_2, \ldots x_n)$, when the observations $x_1, \ldots x_n$ are obtained and assuming a known likelihood and prior. This is a very demanding approach in terms of data, and the requirement of a prior is questionable as the reason for undertaking a fusion process is generally a lack of prior knowledge.

In contrast, if information comes in the form of epistemic states modelled by sets, the fusion can be envisaged from a logical perspective [38]. Suppose one source provides the information items $x \in A$ and $x \in B$. There are three fusion modes:

- A first option is to perform the conjunction of the two sets. It presupposes that both sources can be trusted, and the two pieces of information are consistent.

- A second option is to perform the disjunction of the two sets. It is enough to assume that one source is reliable without knowing which one of the two. It then makes sense even if the pieces of information are inconsistent.

- A third option is to count the number of times each value of x is considered possible by the sources. This is in line with the accumulation method of statistics, assuming sources are independent and identical.

In case of $n > 2$ sources, the two first options would often respectively yield inconsistent or uninformative results, and the most natural approach is then to find maximal consistent subsets of sources, perform conjunction among the information items from each source in each group and perform a disjunction of the results, an idea going back to Rescher and Manor [70] in the logical setting. Alternatively one may assume the knowledge of a number of reliable sources, and perform a *k-out-of-n* kind of merging [39].

The above set-based modes of fusion directly extend to possibility and evidence theories. For instance the conjunction mode exploits fuzzy set intersections to combine possibility theories and random set intersections for combining belief functions. Here, we can observe some discrepancies between the two approaches:

- There are many conjunctive fusion operations for merging possibility distributions, usually using T-norms operators [54]. In particular, if dependencies between sources are unknown the minimum idempotent operator applies, and if independence is assumed one can perform the pointwise product of possibility distributions. One may even perform a fusion mode assuming that a prescribed number of sources may lie (Mundici [65]). In general, merging possibility distributions under the conjunctive mode yields a possibility distribution, but the result may be subnormalized ($\pi(x) < 1, \forall x$). A renormalization step is often considered, assuming the remaining values of positive possibility, however implausible, contain the true one.

- There is essentially one conjunctive merging rule in evidence theory (Dempster's rule of combination) and it presupposes independence between sources. Usually the resulting belief function is subnormal in the sense that $m(\emptyset) > 0$, which can be renormalized (however Smets recommends not to do it so as to keep track of the conflict). This combination rule is the cornerstone of Shafer's theory of evidence, and the possibility of decomposing any belief functions into simple support functions is studied in Shafer's book.[6] This situation contrasts with the case of possibility theory where many operations are available. The possibilistic product rule is in some agreement with the unnormalized Dempster rule of combination because the obtained contour functions are proportional to each other. But merging consonant

---

[6]This turns out to be possible only for a subclass of so-called support belief functions. This difficulty is overcome by Smets [76] by enlarging the notion of simple support belief function.

belief functions by the conjunctive rule does not preserve consonance, and the normalization factors in both theories differ. Worse, there is no clear idempotent merging rule in belief function theory; in particular it is generally impossible to find a joint mass function whose marginals are the source information items and whose contour function would be the minimum of the two contour functions [28]. However Denoeux [27] has enlarged the set of combination rules for belief functions, including an idempotent one, by exploiting Smets [76] decomposition of belief functions via Dempster's combination rule.

Disjunctive fusion modes in possibility theory and evidence theory have similar discrepancies. Random set disjunction corresponds to computing the product of belief functions (or necessity measures) [35]. Since the latter applies to probability measures, the enlarged setting enables the product of probability measures to make sense as a disjunctive combination thereof, yielding random sets having at most two-element realizations, a benefit of considering a unified view of uncertainty theories.

Moreover, the convex combination of belief or plausibility functions and necessity or possibility measures makes sense, but does not preserve consonance. However it comes down to extending the counting mode of fusion to possibility distributions, when restricting to contour functions.

Turning to the general setting of credal sets, Walley [80] extensively discusses the fusion problem in the context of upper and lower probabilities. He proposes a number of properties a rational fusion method should satisfy, which can be adapted to possibility theory and belief functions. Given two familles $\mathcal{M}_1$ and $\mathcal{M}_2$ of probabilities provided by two reliable sources, it is natural to consider that the result of a fusion is the intersection $\mathcal{M} = \mathcal{M}_1 \cap \mathcal{M}_2$, when non-empty. The most natural way of merging n probability functions is to consider the credal set of which they are the vertices. The latter is an idempotent rule adopting a disjunctive mode of combination. This rule has the merit of satisfying all criteria requested by Walley. It is clear that the rule consisting of considering the convex hull of $\mathcal{M}_1 \cup \mathcal{M}_2$ (called the unanimity rule by Walley) extends this rule to sets of probabilities. Walley also considers the hybrid fusion rule based on maximal consistent subsets of sources, as well as the credal set extension of the weighted average probability combination rule.

The fusion of imprecise probabilities and the fusion of belief functions are not really in agreement. It sounds hard to justify Dempster rule of combination in terms of imprecise probabilities and, in the same way, it is not easy to express the mass function induced by $\mathcal{M}_1 \cap \mathcal{M}_2$ in terms of the mass functions induced by $\mathcal{M}_1$ and $\mathcal{M}_2$. Walley rejects Dempster rule of combination on the basis of its failing to satisfy criteria he considers important [81]. One may apply the idempotent fusion of imprecise probabilities to belief functions, by performing the intersection of sets $\mathcal{M}_i = \{P : P(A) \geq \mathrm{Bel}_i(A), \forall A \subseteq S\}$ for $i = 1, 2$. But the lower bounds of the induced probabilities are not belief functions in general. Chateauneuf [10] explores these issues in some detail.

The setting of uncertainty theories explicitly handling incomplete information seems to be more flexible than pure probability theory for the purpose of merging uncertain information. However many questions remain open on this topic, regarding the most

appropriate rules of combination of credal sets and their relationships with similar combination modes in more specialised settings.

1.4   Further reading

We have tried in this chapter to present the main uncertainty theories explicitly coming to grips with imprecision, together with their original motivations. In contrast with the theory of lower previsions, the other uncertainty theories rely on the use of special cases of capacities. We have provided a brief account of some discrepancies between the different theories on central concepts such as conditioning, independence and combination rules, pointing out the existing bridges between these discrepancies. We could have done the same with methods of information comparison.

In this chapter, we have not discussed other emerging approaches such as info-gap theory [4] or the game-theoretic approach of Shafer and Vovk, described in Chapter **??**, for their links with the theories introduced in this chapter largely remain to be studied in more detail.

In this chapter, we have also focused on quantitative representations of uncertainty, leaving aside qualitative representations. However, a purely ordinal framework looks natural when dealing with subjective uncertainty assessments. Indeed, it is easier for an agent to assert that one proposition is more credible than another, rather than providing a numerical belief degree. Ordinal or qualitative approaches represent uncertainty by means of a confidence relation between propositions (or events) obeying some natural properties like consistency with logical deduction. An uncertainty relation formalizes a pairwise comparison of events in terms of relative likelihood. The idea of representing uncertainty by relations over sets of propositions (or events) dates back to De Finetti [23], Koopman [55] and Ramsey [69]. These authors tried to find ordinal counterparts to subjective probabilities. Later on, philosophers such as Lewis [60] have considered other kinds of uncertainty relations in the framework of modal logic, including what eventually turned out to be comparative counterparts to possibility measures [29]. The survey of bridges between (imprecise) probability theory and modal logic [83] is also interesting to consult. Starting from this basic ordinal framework, one may start to redefine the notions of conditioning, independence, fusion, etc. in order to have a full-fledged qualitative uncertainty theory.

Qualitative approaches may also rely on a finite scale of confidence levels. Possibility theory is particularly well adapted to such a qualitative framework, as the minimum and maximum operators are well defined on any lattice (see de Cooman [21] for lattice-valued possibility measures, and [6] and [31] for discussions about qualitative conditioning and qualitative independence, respectively). A recent paper on conditional comparative uncertainty structures is [14].

Conversely, one can search for conditions under which a relation $\leq_\kappa$ on events is compatible with some numerical representations. When only one confidence function $g : \wp(\mathcal{X}) \to [0,1]$ is considered, then it is natural to require $\leq_\kappa$ to be complete and transitive. A confidence function g is then said to represent the relation $\leq_\kappa$ as soon as

$$A \leq_\kappa B \text{ if and only if } g(A) \leq_\kappa g(B).$$

Conditions under which a relation $\leq_\kappa$ is representable by a probability measure, a possibility measure, a belief function and a lower probability were respectively studied by Koopman [55], Dubois [29], Wong *et al.* [86] and Capotorti *et al.* [9]. The reader may benefit from consulting books like the one by Fine [47] and the survey paper by Fishburn [48], for an account of early results on ordinal versions of probability.

On the other hand, one may relax the completeness of $\leq_\kappa$ and search under which conditions a partial preorder is representable by pairs of dual confidence functions. There are however less results concerning such representations. Section 4 of Chapter **??** relates partial preferences with desirable gambles, however gambles are naturally numerical, and can hardly be characterised from the qualitative standpoint. Recently, Nerhing [66] has explored under which conditions a partial preorder $\leq_\kappa$ where $A \leq_\kappa B$ is interpreted as $P(A) \leq_\kappa P(B)$ is representable by a credal set.

Bibliography

[1] C. Alchourròn, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] M. Banerjee and D. Dubois. A simple modal logic for reasoning about revealed beliefs. In C. Sossai and G. Chemello, editors, *ECSQARU '09: Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 5590, pages 805–816, Berlin, 2009. Springer.

[3] C. Baudrit, I. Couso, and D. Dubois. Joint propagation of probability and possibility in risk analysis: Towards a formal framework. *International Journal of Approximate Reasoning*, 45:82–105, 2007.

[4] Y. Ben-Haim. *Information-Gap Decision Theory. Decisions Under Severe Uncertainty.* Academic Press, San Diego, CA, 2001.

[5] B. Ben Yaghlane, P. Smets, and K. Mellouli. Belief function independence: I. The marginal case. *International Journal of Approximate Reasoning*, 29:47–70, 2002.

[6] S. Benferhat, D. Dubois, and H. Prade. Non-monotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence*, 92:259–276, 1997.

[7] J. Berger, E. Moreno, L. Pericchi, M. Bayarri, J. Bernardo, J. Cano, J. De la Horra, J. Martín, D. Ríos-Insúa, B. Betrò, A. Dasgupta, P. Gustafson, L. Wasserman, J. Kadane, C. Srinivasan, M. Lavine, A. O'Hagan, W. Polasek, C. Robert, C. Goutis, F. Ruggeri, G. Salinetti, and S. Sivaganesan. An overview of robust Bayesian analysis. *TEST*, 3:5–124, 1994.

[8] B. Bouchon-Meunier, G. Coletti, and C. Marsala. Independence and possibilistic conditioning. *Annals of Mathematics and Artificial Intelligence*, 35:107–123, 2002.

[9] A. Capotorti, G. Coletti, and B. Vantaggi. Non additive ordinal relations representable by lower or upper probabilities. *Kybernetika*, 34:79–90, 1998.

[10] A. Chateauneuf. Combination of compatible belief functions and relation of specificity. In R. Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theory of Evidence*, pages 97–114. John Wiley & Sons, Inc, New York, NY, 1994.

[11] A. Chateauneuf and M. Cohen. Cardinal extensions of the EU model based on the Choquet integral. In D. Bouyssou, D. Dubois, M. Pirlot, and H. Prade, editors, *Decision-Making Process: Concepts and Methods*, pages 401–434. John Wiley & Sons, Inc, USA, 2009.

[12] B. Chellas. *Modal Logic*. Cambridge University Press, UK, 1978.

[13] G. Coletti and R. Scozzafava. Coherent conditional probability as a measure of uncertainty of the relevant conditioning events. In T. Nielsen and N. Zhang, editors, *ECSQARU '03: Proceedings of the 7th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 2711, pages 407–418, Berlin, 2003. Springer.

[14] G. Coletti and B. Vantaggi. A view on conditional measures through local representability of binary relations. *International Journal of Approximate Reasoning*, 47:268–283, 2008.

[15] R. Cooke. *Experts in Uncertainty*. Oxford University Press, Oxford, UK, 1991.

[16] I. Couso and D. Dubois. On the variability of the concept of variance for fuzzy random variables. *IEEE Transactions on Fuzzy Systems*, 17:1070–1080, 2009.

[17] I. Couso and S. Moral. Independence concepts in evidence theory. *International Journal of Approximate Reasoning*, 51:748–758, 2010.

[18] L. de Campos and J. Huete. Independence concepts in possibility theory: part i. *Fuzzy Sets and Systems*, 103:127–152, 1999.

[19] L. de Campos and J. Huete. Independence concepts in possibility theory: part ii. *Fuzzy Sets and Systems*, 103:487–505, 1999.

[20] L. de Campos, M. Lamata, and S. Moral. The concept of conditional fuzzy measure. *International Journal of Intelligent Systems*, 5:237–246, 1990.

[21] G. de Cooman. Possibility theory - Part i: Measure and integral theoretics groundwork; Part ii: Conditional possibility; Part iii: Possibilistic independence. *International Journal of General Systems*, 25:291–371, 1997.

[22] G. de Cooman and D. Aeyels. Supremum preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.

[23] B. de Finetti. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937. English translation in [57].

[24] B. de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.

[25] B. de Finetti. *Theory of Probability: A Critical Introductory Treatment*. John Wiley & Sons, Chichester, 1974–1975. English translation of [24].

[26] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[27] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172:234–264, 2008.

[28] S. Destercke and D. Dubois. Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory. *Information Sciences*, 19:4075 – 4100, 2011.

[29] D. Dubois. Belief structure, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence*, 5:403–416, 1986.

[30] D. Dubois, H. Fargier, and H. Prade. Ordinal and probabilistic representations of acceptance. *Journal of Artificial Intelligence Research*, 22:23–56, 2004.

[31] D. Dubois, L. Fariñas Del Cerro, A. Herzig, and H. Prade. A roadmap of qualitative independence. In D. Dubois, H. Prade, and E. Klement, editors, *Fuzzy Sets, Logics and Reasoning about Knowledge*, pages 325–350. Kluwer Academic Publishers, Dordrecht, 1999.

[32] D. Dubois, M. Moral, and H. Prade. Belief change rules in ordinal and numerical uncertainty theories. In D. Dubois and H. Prade, editors, *Handbook Defeasible Reasoning and Uncertainty Management Systems. Volume 3: Belief Change*, pages 311–392. Kluwer Academic Publishers, Dordrecht, 1998.

[33] D. Dubois, S. Moral, and H. Prade. A semantics for possibility theory based on likelihoods. *Journal of Mathematical Analysis and Applications*, 205:359–380, 1997.

[34] D. Dubois, H. Nguyen, and H. Prade. Possibility theory, probability and fuzzy sets: Misunderstandings, bridges and gaps. In *Fundamentals of Fuzzy Sets*, pages 343–438. Kluwer Academic Publishers, 2000.

[35] D. Dubois and H. Prade. A set-theoretic view of belief functions: Logical operations and approximation by fuzzy sets. *International Journal of General Systems*, 12:193–226, 1986.

[36] D. Dubois and H. Prade. Properties of measures of information in evidence and possibility theory. *Fuzzy Sets and Systems*, 24:161–182, 1987.

[37] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty.* Plenum Press, New York, 1988.

[38] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4:244–264, 1988.

[39] D. Dubois and H. Prade. Possibility theory and data fusion in poorly informed environments. *Control Engineering Practice*, 2:811–823, 1994.

[40] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1996.

[41] D. Dubois and H. Prade. Possibility theory: Qualitative and quantitative aspects. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 1, pages 169–226. Kluwer Academic Publishers, Dordrecht, 1998.

[42] D. Dubois, H. Prade, and P. Smets. Representing partial ignorance. *IEEE Transactions on Systems, Man, and Cybernetics*, 26:361–377, 1996.

[43] Didier Dubois and Henri Prade. An introduction to bipolar representations of information and preference. *International Journal of General Systems*, 23:866–877, 2008.

[44] D. Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.

[45] R. Fagin and J. Halpern. A new approach to updating beliefs. In P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, volume 6, pages 347–374. North-Holland, Amsterdam, 1991.

[46] T. Fetz. Sets of joint probability measures generated by weighted marginal focal sets. In G. de Cooman, T. Fine, and T. Seidenfeld, editors, *ISIPTA '01: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 171–178, Maastricht, 2001. Shaker.

[47] T. Fine. *Theories of Probability: An Examination of Foundations.* Academic Press, New York, 1973.

[48] P. Fishburn. The axioms of subjective probability. *Statistical Science*, 1:335–358, 1986.

[49] I. Gilboa and D. Schmeidler. Updating ambigous beliefs. *Journal of Economic Theory*, 59:33–49, 1993.

[50] I. Hacking. *The Emergence of Probability.* Cambridge University Press, 1975.

[51] C. Hamblin. The modal "probably". *Mind*, 60:234–240, 1959.

[52] J. Jaffray. Bayesian updating and belief functions. *IEEE Transactions on Systems, Man and Cybernetics*, 22:1144–1152, 1992.

[53] D. Kendall. Foundations of a theory of random sets. In E. Harding and D. Kendall, editors, *Stochastic Geometry.* Wiley, 1974.

[54] E. Klement, R. Mesiar, and E. Pap. *Triangular Norms.* Kluwer Academic Publisher, Dordrecht, 2000.

[55] B. Koopman. The bases of probability. *Bulletin of American Mathematical Society*, 46:763–774, 1940.

[56] H. Kyburg Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[57] H. Kyburg Jr. and H. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964. Second edition (with new material) 1980.

[58] H. Kyburg Jr. and C.-M. Teng. *Uncertain Inference*. Cambridge University Press, Cambridge, 2001.

[59] D. Lewis. Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2:418–446, 1973.

[60] D. Lewis. *Counterfactuals*. Blackwell, Worcester, UK, 1986.

[61] G. Matheron. *Random Sets and Integral Geometry*. John Wiley and Sons, 1975.

[62] K. McConway. Marginalization and linear opinion pools. *Journal of the American Statistical Association*, 76:410–414, 1981.

[63] E. Miranda, I. Couso, and P. Gil. Random sets as imprecise random variables. *Journal of Mathematical Analysis and Applications*, 307:32–47, 2005.

[64] E. Miranda and G. de Cooman. Epistemic independence in numerical possibility theory. *International Journal of Approximate Reasoning*, 103:23–42, 1999.

[65] D. Mundici. The logic of Ulam's game with lies. In C. Bicchieri, editor, *Knowledge, Belief and Strategic Interaction*, pages 275–284. Cambridge University Press, 1998.

[66] K. Nehring. Imprecise probabilistic beliefs as a context for decision-making under ambiguity. *Journal of Economic Theory*, 144:1054–1091, 2009.

[67] J. Pearl. Reasoning with belief functions: An analysis of compatibility. *International Journal of Approximate Reasoning*, 4:363–389, 1990.

[68] F. Pichon, D. Dubois, and T. Denoeux. Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53:159–175, 2012.

[69] F. Ramsey. Truth and probability. In J. Kyburg and H. Smokler, editors, *Studies in Subjective Probability*, pages 23–52. Krieger Publishing Co., NY, 2nd edition, 1998.

[70] N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, 1:179–219, 1970.

[71] L. Savage. *The Foundations of Statistics*. John Wiley & Sons Inc., New York, 1954.

[72] T. Seidenfeld and L. Wasserman. Dilation for sets of probabilities. *The Annals of Statistics*, 21:1139–54, 1993.

[73] G. Shackle. *Decision, Order and Time in Human Affairs.* Cambridge University Press, UK, 1961.

[74] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[75] P. Smets. Possibilistic inference from statistical data. In A. Ballester, D. Cardus, and E. Trillas, editors, *Second World Conference on Mathematics at the Service of Man*, pages 611–613. Universidad Politecnica de Las Palmas, Spain, 1982.

[76] P. Smets. The canonical decomposition of a weighted belief. In *IJCAI '95: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1896–1901. Morgan Kaufmann, 1995.

[77] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.

[78] F. Tonon and S. Chen. Inclusion properties for random relations under the hypotheses of stochastic independence and non-interactivity. *International Journal of General Systems*, 34:615–624, 2005.

[79] P. Walley. Coherent lower (and upper) probabilities. Technical report, University of Warwick, Coventry, 1981. Statistics Research Report.

[80] P. Walley. The elicitation and aggregation of beliefs. Technical report, University of Warwick, Coventry, 1982.

[81] P. Walley. Belief function representations of statistical evidence. *The Annals of Statistics*, 15:1439–1465, 1987.

[82] P. Walley and G. de Cooman. Coherence of rules for defining conditional possibility. *International Journal of Approximate Reasoning*, 21:63–107, 1999.

[83] P. Walley and T. Fine. Varieties of modal (classificatory) and comparative probability. *Synthese*, 41:321–374, 1979.

[84] P. Williams. Notes on conditional previsions. Technical report, University of Sussex, February 1975. Published as [85] ****************** unify? **********************.

[85] P. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383, 2007. Original: [84] ****************** unify? **********************.

[86] S. Wong, Y. Yao, P. Bollmann, and H. Burger. Axiomatization of qualitative belief structures. *IEEE Transactions on Systems, Man, and Cybernetics*, 21:259–276, 1991.

[87] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.