Representations of Uncertainty in Artificial Intelligence: Probability and Possibility

Thierry Denœux, Didier Dubois, and Henri Prade

Abstract Due to its major focus on knowledge representation and reasoning, artificial intelligence was bound to deal with various frameworks for the handling of uncertainty: probability theory, but more recent approaches as well: possibility theory, evidence theory, and imprecise probabilities. The aim of this chapter is to provide an introductive survey that lays bare specific features of two basic frameworks for representing uncertainty: probability theory and possibility theory, while highlighting the main issues that the task of representing uncertainty is faced with. This purpose also provides the opportunity to position related topics, such as rough sets and fuzzy sets, respectively motivated by the need to account for the granularity of representations as induced by the choice of a language, and the gradual nature of natural language predicates. Moreover, this overview includes concise presentations of yet other theoretical representation frameworks such as formal concept analysis, conditional events and ranking functions, and also possibilistic logic, in connection with the uncertainty frameworks addressed here. The next chapter in this volume is devoted to more complex frameworks: belief functions and imprecise probabilities.

Thierry Denoeux
Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Heudiasyc (UMR 7253)
e-mail: `tdenoeux@utc.fr`

Didier Dubois
IRIT-CNRS, Université Paul Sabatier, Toulouse, France,
e-mail: `dubois@irit.fr`

Henri Prade
IRIT-CNRS, Université Paul Sabatier, Toulouse, France,
e-mail: `prade@irit.fr`

1 Introduction : An Historical Perspective

The question of including, hence modeling, uncertainty in scientific matters is not specific to the field of artificial intelligence. Historically, this concern already appears in the XVIIth century, with pioneering works of Huyghens, Pascal, chevalier de Méré, and Jacques Bernoulli. There existed at that time a major distinction between the objective notion of *chance* in connection with the study of games (of chance), and the subjective notion of *probability* in connection with the issue of unreliable testimonies at courts of law. With J. Bernoulli, chances are related to frequencies of events and are naturally additive, while subjective probabilities are not supposed to be so. This view is still present in the middle of the XVIIIth century with the works of Lambert. He proposed a combination rule which turns out to be a special case of Dempster's rule of combination; see [Shafer, 1978; Martin, 2006], and Chapter 4 of this volume. However, with the rapid development of physics and actuarial sciences later on, the interest for the non-additive side of probability eventually waned and the issue was forgotten for almost two centuries, while the additive view became prominent, with the works of Laplace, whether the focus was on frequentist probability or not. Noticeably, in the middle of the XXth century, in economics, not only the main approach to decision under (frequentist) risk after [von Neumann and Morgenstern, 1944], but also the mainstream theory of decision under (subjective) uncertainty relied on additive probability.

It is the emergence of computer sciences that brought issues related to knowledge representation and reasoning in the presence of imprecision, uncertainty, and conflicting information to the front. This went on till the 1980's almost independently of probability theory and the issue of decision-making. Instead, artificial intelligence first put the emphasis on logical and qualitative formalisms, as well as the modeling of linguistic information (in trends of research such as fuzzy set theory).

Indeed, the available information to be stored in a computer is often unreliable, as is human knowledge, so that reasoning is based on rules that may lead to uncertain conclusions even starting from sure premises. The need to handle uncertainty arose in fact with the emergence of the first expert systems at the beginning of the 1970's. One of the first and best known expert rule-based system, namely MYCIN [Shortliffe, 1976; Buchanan and Shortliffe, eds.], already proposed an ad hoc, entirely original, technique for uncertainty propagation based on degrees of belief and disbelief. This method will not be described here for lack of space, and because it is now totally outdated, especially due to its improper handling of exceptions in if-then rules. But the uncertainty propagation technique of MYCIN pioneered the new, more rigorous frameworks for uncertainty modeling that would appear soon after. On this point, see [Dubois and Prade, 1989], and [Lucas and van der Gaag, 1991] as well.

This chapter is structured in four sections. In Section 2, basic notions useful for describing the imperfection of information are defined and discussed. Section 3 deals with probability theory, focusing on the possible meanings of probability and the difficulty to handle plain incomplete information with probability distributions, as well as the connections between conditioning and logic. Section 4 deals with set functions extending the modalities of possibility and necessity, distinguishing between qualitative and quantitative approaches, and describing connections with reasoning tolerant to exceptions, formal concept analysis, probability and statistics. Section 5 explains the links between uncertain reasoning and Aristotelian logic, generalizing the square of opposition.

2 Imprecision, Contradiction, Uncertainty, Gradualness, and Granularity

Before presenting various representation frameworks (see [Halpern, 2003; Dubois and Prade, 2009; Liu, 2001; Parsons, 2001] for interesting focused overviews), it is useful to somewhat clarify the terminology. We call *information item* any collection of symbols or signs produced by observing natural or artificial phenomena, or by human cognitive activity, whose purpose is communication. Several distinctions are in order. First, one must separate so-called *objective* information items, coming from sensor measurements or direct observations of the world, from *subjective* ones, expressed by individuals and possibly generated without using direct observations of the outside world. Information items may be couched in numerical formats, especially objective ones (sensor measurements, counting processes), or in qualitative or symbolic formats (especially subjective ones, in natural language for instance). However the dichotomy objective numerical vs. subjective qualitative is not so clearcut. A subjective information item can be numerical, and objective observations can be qualitative (like a color perceived by a symbolic sensor, for instance). Numerical information can take various forms: integers, real numbers, intervals, real-valued functions, etc. Symbolic information is often structured and encoded in logical or graphical representations. There are also hybrid representation formats, like Bayesian networks [Pearl, 1988]. Finally, another important distinction should be made between *singular* and *generic* information. Singular information refers to particular facts and results from an observation or a testimony. Generic information pertains to a *class* of situations and expresses knowledge about it: it can be a law of physics, a statistical model stemming from a representative sample of observations, or yet commonsense statements such as “birds fly” (in this latter case the underlying class of situations is not precise: is it here a zoological definition, or the birds of any epoch, or of any place, etc. ?).

2.1 Imprecise Information

To represent the epistemic state of an agent, one must beforehand possess a language for representing the states of the world under interest, according to the agent, that is, model relevant aspects by means of suitable attributes. Let v be a vector of attribute variables[1] relevant for the agent, and let S be its domain (possibly not described in extension). S is then the set of (precise descriptions) of the set of possible states of affairs. A subset A of S is viewed as an event, or as a proposition that asserts $v \in A$.

An information item $v \in A$ possessed by an agent is said to be *imprecise* if it is not sufficient to enable the agent to answer a question of interest about v. Imprecision corresponds to the idea of *incomplete* or even missing information. The question to which the agent tries to answer is of the form what is the value of v, or more generally does v satisfy a certain property B, given that $v \in A$ is known? The notion of imprecision is not absolute. When concerned with the age of a person, the term *minor* is precise if the referential set is $S = \{minor, major\}$ and the question is : has this person the right of vote? In contrast if the question is to determine the age of this person and $S = \{0, 1, \ldots, 150\}$ (in years), the term *minor* is very imprecise.

The standard format of an imprecise information item is $v \in A$ where A is a *subset* of *S containing more that one element*. An important remark is that elements of A, seen as possible values of v are mutually exclusive (since the entity v possesses only one value). So, an imprecise information item takes the form of a disjunction of mutually exclusive values. For instance, to say that John is between 20 and 22 years old, that is, $v = age(John) \in \{20, 21, 22\}$ means to assume that $v = 20$ or $v = 21$ or $v = 22$. An extreme form of imprecise information is *total ignorance*: the value of v is completely unknown. In classical logic, imprecision explicitly takes the form of a disjunction (stating that $A \vee B$ is true is less precise than stating that A is true). The set A representing an information item is called an *epistemic set*.

Two imprecise information items can be compared in terms of informational content: an information item $v \in A_1$ is said to be *more specific* than another information item $v \in A_2$ if and only if A_1 is a proper subset of A_2.

The disjunctive view of sets used to represent imprecision contrasts with the more usual conjunctive view of a set as a collection of items forming a certain complex entity. It then represents a precise information item. For instance, consider the set of languages that John can speak, say $v = Lang(John)$. This variable is set-valued and stating that $Lang(John) = \{\text{English, French}\}$ is a precise information item, as it means that John can speak English and French only. In contrast, the variable $v' = NL(John)$ representing the native language of John is single-valued and the statement $NL(John) \in \{\text{English, French}\}$ is imprecise. The domain of v' is the set of all spoken languages while the domain of v is its power set. In the latter case, an imprecise information item pertaining to a set-valued variable is represented by a (disjunctive) set of (conjunctive) subsets.

[1] In fact, in this chapter, v denotes an ill-known entity that may be for instance a random variable in a probabilistic setting, or rather an imprecisely known entity but which does not vary strictly speaking.

2.2 Contradictory Information

An information item is said to be contradictory if it is of the form $v \in A$, where $A = \emptyset$. Under this form there is not much we can do with such an information item. In mathematics, the presence of a contradiction ruins any form of reasoning, and it is only used to prove claims by refutation (a claim is true because assuming its falsity leads to a contradiction). In artificial intelligence, contradiction often stems from the conflict between several information items, e.g., $v \in A$ and $v \in B$ where $A \cap B = \emptyset$. It is thus a natural situation that is to be expected each time there are several sources, and more generally if collected information items are numerous. Another cause of conflicting information is the presence of exceptions in generic information items such as rules, which may lead to simultaneously infer opposite conclusions. There are several approaches in the literature that aim at coping with contradictory information, and that are studied in this book:

- information fusion techniques that aim at restoring consistency, by deleting unreliable information items, taking into account the sources that deliver them, and analyzing the structure of the conflict between them. See Chapter 14 in this volume.
- argumentation methods that discuss the pros and the cons of deriving a proposition $v \in A$ using a graph-theoretic representation of an attack relation between conflicting arguments. See Chapter 13 in this volume.
- paraconsistent logics that try to prevent the infection of the contradiction affecting some variables or some subgroups of information items to other ones, by for instance changing the inference relation, thus avoiding the explosive nature of standard inference from inconsistent bases in classical logic. See Chapter 13 in this volume.
- nonmonotonic reasoning formalisms that try to cope with exceptions in rules by giving priority to conclusions of the most specific ones. See Chapter 2 in this volume.

2.3 Uncertain Information

An information item is said to be uncertain for an agent if the latter does not know whether it is true or false. If an elementary information item of the form of a proposition $v \in A$, where A contains a set of non-impossible values for v, is tainted with uncertainty, a token of uncertainty is attached to it. This token is a qualifier situated at the meta-level with respect to the information item. It can be numerical or symbolic: compare statements expressing uncertainty such as *The task will take at least one hour, with probability 0.7*, and *It is not fully sure that John comes to the meeting*. Uncertainty has two main origins: the lack of information, or the presence of conflicting information. A special case of the latter is aleatory uncertainty, where due to the variability of an observed phenomenon, it is difficult to predict the next event, hence the information item $v \in A$ that may describe it.

The most usual representation of uncertainty consists in assigning to each proposition $v \in A$ or event $A \subseteq S$, a number $g(A)$ in the unit interval. This number expresses the agent's confidence in the truth of the proposition $v \in A$. Note that this proposition is ultimately only true or false, but the agent may currently ignore what its actual truth-value is. Natural conditions are required for the set function g:

$$g(\emptyset) = 0;\ \ g(S) = 1;\ \ \text{if } A \subseteq B \text{ then } g(A) \leq g(B). \tag{1}$$

Indeed the contradictory proposition $v \in \emptyset$ is impossible, and the tautological proposition $v \in S$ is certainly true. Moreover, if A is more specific than B (and thus implies B), a rational agent cannot trust $v \in A$ more than $v \in B$. When S is infinite, one must add suitable continuity properties with respect to monotonic sequences of subsets. Such a function g is often called a *capacity* [Choquet, 1953], or *fuzzy measure* [Sugeno, 1977], or yet plausibility function [Halpern, 2001] (not to be confused with the dual to belief functions, defined in the next chapter in this volume). An important consequence of (1) is in the form of two inequalities:

$$g(A \cap B) \leq \min(g(A), g(B)); \quad g(A \cup B) \geq \max(g(A), g(B)). \tag{2}$$

These inequalities suggest to consider extreme confidence measures g such that one of these inequalities is an equality, and more generally, when A and B are mutually exclusive, assume that $g(A \cup B)$ only depends on $g(A)$ and $g(B)$ [Dubois and Prade, 1982], i.e.,

$$\text{if } A \cap B = \emptyset \text{ then } g(A \cup B) = g(A) \oplus g(B). \tag{3}$$

for some binary operation $\oplus$ on $[0,1]$.

The *conjugate* set function, defined by $\overline{g}(A) = 1 - g(\overline{A})$, then satisfies the dual property $\overline{g}(A \cap B) = \overline{g}(A) \perp \overline{g}(B)$ if $A \cup B = S$ where $a \perp b = 1 - (1-a) \oplus (1-b)$ [Dubois and Prade, 1982]. The set functions g and $\overline{g}$ are said to be *decomposable*. Compatibility constraints with the Boolean algebra of events suggests considering operations $\oplus$ and $\perp$ that are associative, which leads to choose $\perp$ and $\oplus$ among triangular *norms* and *co-norms* [Klement et al, 2000] (they get their name from their role in the expression of the triangular inequality in stochastic geometry [Schweizer and Sklar, 1963]). The main possible choices for $a \perp b$ (resp. $a \oplus b$) are the operators minimum $\min(a,b)$, product $(a \times b)$, and truncated addition $\max(0, a+b-1)$ (resp. maximum $\max(a,b)$, probabilistic sum $a+b-a \times b$, and bounded sum $\min(1, a+b)$). Probability measures are recovered by defining $a \oplus b = \min(1, a+b)$ (equivalently $a \perp b = \max(0, a+b-1)$), and *possibility* measures and *necessity* respectively for $a \oplus b = \max(a,b)$ and for $a \perp b = \min(a,b)$. The use of more complex operators (like ordinal sums of the above ones) may make sense [Dubois et al, 2000b].

2.4 Graduality and Fuzzy Sets

Representing a proposition in the form of a statement that can only be true or false (or an event that occurs or not) is but a convenient convention. It is not always an ideal one. Some information items are not easily amenable to respecting this convention. This is especially the case for statements involving *gradual* properties, like in the proposition *John is young*, that may sometimes be neither completely true nor completely false: it is clearly more true if John is 20 than if he is 30, even if in the latter case, John is still young to some extent. Predicates like *young* can be modified by linguistic hedges. It makes sense to say *very young, not so young*, etc. Such linguistic hedges cannot be applied to Boolean predicates, like *single*. In other words, the proposition *John is young* is not Boolean, which denotes the presence of an ordering between age values to which it refers. This type of information can be taken into account by means of *fuzzy sets* [Zadeh, 1965]. A fuzzy set F is a mapping from S to a totally ordered set L often chosen to be the unit interval $[0,1]$. The value $F(s)$ is the membership degree of the element s in F. It evaluates the compatibility between the situation s and the predicate F.

Fuzzy sets are useful to deal with information items in natural language referring to a clear numerical attribute. Zadeh [1975] introduced the notion of *linguistic variable* with values in a linearly ordered linguistic term set. Each of these terms represents a subset of the numerical domain of the attribute, and these subsets correspond to a partition of this domain. For instance, the set of terms $T = \{$*young, adult, old*$\}$ forms the domain of the linguistic variable *age(John)* and partitions the domain of this attribute. Nevertheless it is not surprising to admit that the transitions between the ranges covered by the linguistic terms are gradual rather than abrupt. And in this situation, it sounds counterintuitive to set precise thresholds separating these continuous ranges. Namely, it sounds absurd to define the set $F =$*young* $\in T$ by a precise threshold $s_\star$ such that $F(s) = 0$ if $s > s_\star$ and $F(s) = 1$ otherwise, beyond which an individual suddenly ceases to be young. The membership function of the fuzzy set valued in the scale [0, 1], representing here the gradual property *young*, is but a direct reflection of the continuous domain of the attribute (here the age). This also leads to the idea of a fuzzy partition made of non-empty fuzzy subsets $F_1, \cdots, F_n$, often defined by the constraint $\forall s, \Sigma_{i=1,n} F_i(s) = 1$ [Ruspini, 1970].

If we admit that some sets are fuzzy and membership to them is a matter of degree, one issue is to extend the set-theoretical operations of union, intersection and complementation to fuzzy sets. This can be done in a natural way, letting

$$(F \cup G)(s) = F(s) \oplus G(s); \quad (F \cap G)(s) = F(s) \perp G(s); \quad \overline{F}(s) = 1 - F(s),$$

where $\oplus$ and $\perp$ are triangular co-norms and norms already encountered in the previous subsection. The choice $\oplus = \max$ and $\perp = \min$ is the most common. With such connectives, the De Morgan property between $\cup$ and $\cap$ are preserved, as well as their idempotence and their mutual distributivity. However, the excluded middle ($A \cup \overline{A} = S$) and contradiction laws ($A \cap \overline{A} = \emptyset$) fail. Choosing $\oplus = \min(1, \cdot + \cdot)$ and $\perp = \max(0, \cdot + \cdot - 1)$ re-install these two laws, at the cost of losing idempotence and mutual distributivity of $\cup$ and $\cap$. As to fuzzy set inclusion, it is oftentimes defined by the condition $F \subseteq G \Leftrightarrow \forall s, F(s) \leq G(s)$. A more drastic notion of inclusion requires the inclusion of the support of F (elements s such that $F(s) > 0$) in the core of G (elements s such that $G(s) = 1$). In agreement with the spirit of fuzzy sets, inclusion can also be a matter of degree. There are various forms of inclusion indices, of the form $d(F \subseteq G) = \min_s F(s) \to G(s)$, where $\to$ is a many-valued implication connective.

Fuzzy sets led to a theory of approximate reasoning and the reader is referred to a section dedicated to interpolation in Chapter 10 of this volume. Besides, since the mid-1990's, there has been a considerable development of formal fuzzy logics, understood as syntactic logical systems whose semantics is in terms of fuzzy sets. These works, triggered by the book by Hájek [1998], considerably improved the state of the art in many-valued logics developed in the first half of the XXth century (see [Dubois et al, 2007] for a detailed survey of both approximate reasoning and formal fuzzy logic.)

2.5 Degree of Truth vs. Degree of Certainty: A Dangerous Confusion

It is very crucial to see the difference between the degree of adequacy between a state of affairs and an information item (often called *degree of truth*) and a degree of certainty (confidence). Already, in natural language, sentences like *John is very young* and *John is probably young* do not mean the same. The first sentence expresses the fact that the degree of membership of *age(John)* (e.g., *age(John)* = 22) to the fuzzy set of young ages is for sure high. The degree of membership $F(s)$ evaluates the degree of adequacy between a state of affairs s_0, e.g., $s_0 = 22$, and the fuzzy category $F = young$. According to the second sentence, it is not ruled out that John is not young at all.

Degrees of truth and degrees of certainty correspond to distinct notions that occur in distinct situations with unrelated semantic contents. Moreover they are driven by mathematical frameworks that should not be confused despite their superficial resemblances as to the involved connectives. Indeed, as seen earlier in this text, truth degrees are usually assumed to be compositional with respect to all connectives like conjunction, disjunction, and negation (respectively corresponding to intersection, union, and complementation of fuzzy sets). However, full-fledged compositionality is impossible for degrees of certainty. This is because the Boolean algebra of standard events is not compatible with the structure of the unit interval, nor any finite totally ordered set with more than 2 elements [Dubois and Prade, 2001]: they are not Boolean algebras. For instance, probability is compositional only for negation ($Prob(\overline{A}) = 1 - Prob(A)$), and as we shall see later on, possibility (resp. necessity) measures are compositional only for disjonction (resp. conjonction). For instance one can be sure that $v \in A \cup B$ is true (especially if $B = \overline{A}$!), without being sure at all that any of $v \in A$, and $v \in B$ is true.

A typical situation where certainty and truth tend to be confused is when using a three-valued logic to capture partial information, changing Boolean interpretations of a language into three-valued ones. The usual truth set $\{0,1\}$ is turned into, say, $\{0,1/2,1\}$, with the idea that $1/2$ stands for *unknown* as in Kleene logic [Kleene, 1952]. Now the problem is that under the proposed calculus by Kleene with conjunction, disjunction and negation expressed by operations min, max and $1-(\cdot)$, respectively, the excluded middle law is lost. This is a paradox here as, since a proposition $v \in A$ can only be true or false, the composite proposition $v \in A$ *or* $v \notin A$ is always valid while, in the three-valued setting, it will have truth value 1/2 if $v \in A$ is set to *unknown*. The way out of the paradox consists in noticing that the negation of *unknown* is *known*, actually known to be true or known to be false. So the three alleged truth-values $\{0,1/2,1\}$ are degrees of certainty, and actually stand for the three non-empty subsets of $\{0,1\}$, 1/2 standing for the hesitation between true and false, namely $\{0,1\}$. And then it becomes clear the statement *either* $v \in A$ *is known to be true* or $v \in A$ *is known to be false* is not a tautology.

The Kleene approach to ignorance has been extended by Belnap [1977a; 1977b] to include contradictory information stemming from conflicting sources, adding a fourth truth value expressing contradiction. The 4-valued truth set forms a bilattice structure and is isomorphic to the four subsets of $\{0,1\}$ (now including $\emptyset$), equipped with two partial orderings: the truth-ordering (where the two new truth-values are incomparable and lie between true and false) and the information ordering (that coincides with inclusion of subsets in $\{0,1\}$). These "epistemic truth-values" are attached to atomic propositions, and truth-tables in agreement with the bilattice structure enable the epistemic status of complex propositions to be computed. The same kind of analysis as above applies regarding the use of compositional truth values in this logic (e.g., *true* in the sense of Belnap means approved by some source and disapproved by none, an epistemic stance). See [Dubois, 2012] for a discussion. Besides, Ginsberg [1990] used Belnap bilattices to propose a unified semantic view for various forms of non-monotonic inferences (see Chapter 2 in this volume and the subsection 4.2.1 in this chapter).

2.6 Granularity and Rough Sets

In the preceding sections, we did not question the assumptions that underlie the definition of the set S of states of the world. It should not be taken for granted, as it presupposes the definition of a language. The logical approach to Artificial Intelligence often starts from a set of statements expressed in a propositional language, to which it may assign degrees of confidence. Then the set S is the set of states or interpretations generated by these propositions (mathematically, the subsets of S are the smallest Boolean algebra supporting these propositions). This view has important consequences for the representation and the updating of bodies of information items. For instance, a new information item may lead to a refinement of S: this is called a change of granularity of the representation.

The simplest case of change of granularity is when the basic propositions are taken as atomic ones, or more generally when describing objects by attributes. Let Ω be a set of objects described by attributes $V_1, V_2, \ldots, V_k$ with respective domains $D_1, D_2, \ldots, D_k$. Then S is the Cartesian product $D_1 \times D_2 \times \cdots \times D_k$. Each element of S can be refined into several ones if a $(k+1)$th attribute is added. Clearly, nothing prevents distinct objects from having the same description in terms of such attributes. Then they are indiscernible by means of this language of description.

Consider a subset Θ of objects in Ω. It follows from the above remark, that in general Θ cannot be described precisely using such an attribute-based language. Indeed let R be an equivalence relation on Ω clustering objects having the same description: $\omega_1 R \omega_2$ if and only if $V_i(\omega_1) = V_i(\omega_2), \forall i = 1, \ldots, k$. Let $[\omega]_R$ be the equivalence class of object ω. We only have such equivalence classes to describe the set Θ, so only approximate descriptions of it can be used. The only thing we can do is to build upper and lower approximations Θ^* and Θ_* defined as follows:

$$\Theta^* = \{\omega \in \Omega : [\omega]_R \cap \Theta \neq \emptyset\}; \quad \Theta_* = \{\omega \in \Omega : [\omega]_R \subseteq \Theta\} \tag{4}$$

The pair (Θ^*, Θ_*) is called a *rough set* [Pawlak, 1991; Pawlak and Skowron, 2007]. Only subsets of objects such as Θ^* and Θ_* can be accurately described by means of combinations of attribute values $V_1, V_2, \ldots, V_k$.

There are various examples of situations where rough sets implicitly appear, for instance histograms or digital images correspond to the same notions of indiscernability and granularity, where equivalence classes correspond, respectively, to the bins of the histograms and to pixels.

The concept of rough set is thus related to the ones of indiscernibility and granularity, while the concept of fuzzy set is related to gradualness. It it is possible to build concepts where these two dimensions are at work, when the set to be approximated or the equivalence relation become fuzzy [Dubois and Prade, 1992]. Rough sets are also useful in machine learning to extract rules from incomplete data [Grzymala-Busse, 1988; Hong et al, 2002], as well as fuzzy decision rules [Greco et al, 2006] (see Chapter 12 in Volume 2).

3 Uncertainty: The Probabilistic Framework

Probability theory is the oldest uncertainty theory and, as such, the best developed mathematically. Probability theory can be envisaged as a chapter of mathematics. In that case, we consider a probability space, made of a set Ω (called a sample space) and an application v from Ω to S (called random variable), where oftentimes S is taken as the real line. In the simplest case S is a finite set which determines via v a finite partition of Ω. If $\mathscr{B}$ is the Boolean algebra generated by this partition, a probability space is actually the triple $(\Omega, \mathscr{B}, \mathscr{P})$, and P is a probability measure, i.e., an application from $\mathscr{B}$ to $[0, 1]$ such that:

$$P(\emptyset) = 0; \;\; P(\Omega) = 1; \tag{5}$$

$$\text{if } A \cap B = \emptyset \text{ then } P(A \cup B) = P(A) + P(B). \tag{6}$$

Elements of $\mathscr{B}$ are called measurable subsets of Ω. The probability distribution induced by v on S is then characterized by a set of weights $p_1, p_2, \ldots, p_{card(S)}$, defined by $p_i = P(v^{-1}(s_i))$, and such that

$$\sum_{i=1}^{card(S)} p_i = 1.$$

Probabilities of events can be extended to fuzzy events by considering the expectation of their membership functions [Zadeh, 1968], which indeed generalizes the usual expression $P(A) = \sum_{s_i \in A} p_i$ of a classical event.

Beyond the apparent[2] unity of the mathematical model of probability, there are strikingly different views of what probability means [Fine, 1983]. The aim of this section is to discuss some of these views, emphasizing some limitations of the representation of uncertainty by means of a unique probability distribution. This section is completed by a glance at De Finetti's conditional events and their three-valued logic, and at a very specific kind of probability distribution (so-called *big-stepped*) that play a noticeable role in the representation of default rules.

3.1 Frequentists vs. Subjectivists

If probability theory is considered as a tool for knowledge representation, one must explain what probability means, what is it supposed to represent. There are at least three understandings of probability, that have been proposed since its inception.

The simplest one is combinatorial. The set Ω is finite and p_i is proportional to the number of elements in $v^{-1}(s_i)$. Then a probability degree is just a matter of counting, for each event, the proportion of favorable cases over the number of possible ones. The well-foundedness of this approach relies on considerations about symmetry (a principle of indifference or insufficient reason, after Laplace), or the assumption that the phenomenon we deal with is genuinely random (like coin flipping, fair die tossing, etc.), and follows a uniform distribution.

The most common interpretation is frequentist. It is assumed that we accumulate observations (a finite n-element subset $\Omega(n)$ of the sample space Ω). Then frequencies of observing $v = s_i$ defined by:

$$f_i = \frac{card(v^{-1}(s_i) \cap \Omega(n))}{n}$$

[2] Apparent, because the mathematical settings proposed by Kolmogorov and De Finetti [1974] are different, especially for the notion of conditioning, even if the Kolmogorov setting seems to be overwhelmingly adopted by mathematicians.

can be obtained. When S is infinite, we can build a histogram associated to the random variable v by considering frequencies of elements of a finite partition of S (possibly adjusting a continuous distribution to it).

As the number of observations increases, $\Omega(n)$ becomes a representative sampling of Ω, and it is assumed that such frequencies f_i converge to probability values defined as limits, by $p_i = \lim_{n\to\infty} f_i$. To use this definition of probability, one must clearly have a sufficient number of observations available (ideally an infinite number) for the phenomenon under study. Under this view, the probability of a non-repeatable event makes no sense. Moreover, the frequentist probability distribution is a mathematical model of a physical phenomenon, hence objective, even if it can be part of the knowledge of an agent.

Under the third, subjectivist, view, the degree of probability $P(A)$ is interpreted as a degree of belief of an agent in the truth of the information item $v \in A$. Hence it should apply to any event, be it repeatable or not. What plays the role of frequencies for making subjective probability operational for non-repeatable events is the amount of money one should pay for a gamble on the occurrence or the non occurrence of event A. More precisely the degree of probability $P(A)$ for an agent is equated to the fair price this agent is willing to pay to a bookmaker for a lottery ticket with a 1 euro reward in case the event occurs. The price is fair in the sense that the agent would also agree to sell it at this price to the bookmaker, should the latter decide to buy it. Clearly the more the agent believes in A the greater (i.e., the closer to 1 euro) the price (s)he is likely to offer. This approach then relies on a rationality principle, called *coherence*, saying that the agent is not willing to lose money for sure. It ensures that degrees of belief (betting prices) behave in an additive way like probabilities. To see it, suppose the agent buys two lottery tickets, the first one to bet on A, the second one to bet on its complement $\overline{A}$. The agent is sure to have one winning ticket, which means a profit of $1 - P(A) - P(\overline{A})$ euros in relative value. Prices such that $P(A) + P(\overline{A}) - 1 > 0$ are not rational as it means a sure loss for the agent. However, prices such that $P(A) + P(\overline{A}) - 1 < 0$ are unfair and will lead the bookmaker to buy the tickets at those prices instead of selling them, to avoid sure loss on the bookmaker side. So the only choices left for the agent is to propose prices such that $P(A) + P(\overline{A}) = 1$. The same reasoning can be carried our for three mutually exclusive events, $A, B, \overline{A \cup B}$, leading to the constraint $P(A) + P(B) + P(\overline{A \cup B}) = 1$, which, since $P(\overline{A \cup B}) = 1 - P(A \cup B)$, leads to $P(A \cup B) = P(A) + P(B)$. Note that the probability degrees so-defined are personal, and may change across agents, contrary to frequentist probabilities.

Apparently, the subjectivist approach looks like a mere copy of the calculus of frequentist probabilities. In fact as shown by De Finetti [1974] and his followers [Coletti and Scozzafava, 2002], things are not so simple. First, in the subjectivist approach there is no such thing as a sample space. The reason is that a subjective probability is either assigned to a unique event (after betting one checks whether this event did occur or not), or to a single realization of a repeatable one (e.g., flipping this coin now). Next, on infinite spaces, only finite additivity (in contrast with σ-additivity for the frequentist approach) can be justified by the above betting paradigm. Finally, the initial data does not consist of statistics, but a collection of bets (prices c_i) on on the truth of propositions A_i in an arbitrary set thereof $\{A_j : j = 1,\ldots,m\}$, along with a number of logical constraints between those propositions. The state space S is then constructed based on these propositions and these constraints. It is assumed, by virtue of the coherence principle, that the agent assigns prices c_j to propositions A_j in agreement with the probability calculus, so that there is a probability distribution that satisfies $P(A_j) = c_j, j = 1,\ldots,m$. While the frequentist approach leads to assuming a unique probability distribution representing the random phenomenon (obtained via an estimation process from statistical data), this is not the case in the subjectivist setting, if the bets bear on arbitrary events. Indeed there may be several probability measures such that $c_j = P(A_j), \forall j = 1,\ldots,m$. Any of those probability functions is coherent but the available information may not allow us to select a single one. It may also occur that no such probability exists (then the bets are not coherent). To compute the probability degree $P(A)$ of some arbitrary event A based on a collection of pairs $\{(A_j, c_j) : j = 1,\ldots,m\}$, one must solve linear programming problems whose decision variables are probabilities p_i attached to singletons of S of the form: maximise (or minimise) $\sum_{s_i \in A} p_i$ under constraints $c_j = \sum_{s_k \in A_j} p_k, \forall j = 1,\ldots,m$.

It is then clear that the subjectivist approach to probability is an extension of the logical approach to artificial intelligence based on propositional logic and classical inference. The latter is recovered by assigning probability $c_j = 1$ to $A_j, j = 1,\ldots,m$, which enforces $P(A) = 1$ to all logical consequences A of $\{A_j : j = 1,\ldots,m\}$.

There are other formal differences between frequentist and subjectivist probabilities when it comes to conditioning.

3.2 Conditional Probabilities

By considering S as the state space, it is implicitly assumed that S represents an exhaustive set of possible worlds. To emphasize this point of view we may as well write the probability $P(A)$ as $P(A \mid S)$. If further on the agent receives new information that comes down to restraining the state space, probabilities will be defined based on a different context, i.e., a non-empty subset $C \neq \emptyset \subset S$ and the probability $P(A)$ becomes $P(A \mid C)$ in this new context. Changing $P(A)$ into $P(A \mid C)$ essentially consists in a renormalization step for probabilities of states inside C, setting other probabilities to 0:

$$P(A \mid C) = \frac{P(A \cap C)}{P(C)} \tag{7}$$

We can indeed check that $P(A) = P(A \mid S)$. This definition is easy to justify in the frequentist setting, since indeed $P(A \mid C)$ is but the limit of a relative frequency.

Justifying this definition in the subjectivist case is somewhat less straightforward. The probability $P(A \mid C)$ is then assigned to the occurrence of a conditional event denoted by $A \mid C$[3]. The quantity $P(A \mid C)$ is again equated to the fair price of a lottery

[3] We come back to the logic of conditional events at the end of this section.

ticket for the conditional bet on $A \mid C$. The difference with a standard bet is that if the opposite of C occurs, the bet is called off and the amount of money paid for the ticket is given back to the agent [De Finetti, 1974]. The conditional event $A \mid C$ represents the occurrence of event A in the hypothetical context where C would be true. In this operational set-up it can be shown that the identity $P(A \cap C) = P(A \mid C) \cdot P(C)$ holds.

This definition of conditional probability contrasts with the one of Kolmogorov based on a quotient, which presupposes $P(C) \neq 0$, and proves to be too restrictive in the subjectivist setting. Indeed, in the latter setting, conditional probabilities are directly collected, so that conditional probability is the primitive concept in the subjectivist setting of De Finetti, and no longer derived from the unconditional probability function. The conditional probability satisfying $P(A \cap C) = P(A \mid C) \cdot P(C)$ still makes sense if $P(C) = 0$ (see [Coletti and Scozzafava, 2002]).

Under the subjectivist view, a body of knowledge consists of a set of conditional probability assignments $\{P(A_i \mid C_j) = c_{ij}, i = 1, \ldots, m; j = 1, \ldots, n\}$. Such conditional events correspond to various hypothetical contexts whose probability is allowed to be 0. The questions of interest are then (i) to derive a probability distribution in agreement with those constraints (actually a sequence of probability measures on disjoint parts of S (see [Coletti and Scozzafava, 2002]); (ii) to find induced optimal bounds on some conditional probability $P(A|C)$. For instance, one may consider the probabilistic syllogism already studied by Boole and De Morgan. Namely suppose the quantities $P(B|A)$, $P(C|B)$ are precisely known, what can be inferred about $P(C|A)$? It turns out that if $P(C|B) < 1$, we can only conclude that $P(C|A) \in [0,1]$. However when the values of $P(A|B)$ and $P(B|C)$ are known as well, we can compute non-trivial bounds on $P(C|A)$. These bounds can be found in [Dubois et al, 1993]. For example, it can be shown that

$$P(C|A) \geq P(B|A) \cdot \max\left(0, 1 - \frac{1 - P(C|B)}{P(A|B)}\right)$$

and that this lower bound is tight.

Yet another mathematical attempt to justify probability theory as the only reasonable belief measure is the one of R. T. Cox [1946]. To do so he relied on the Boolean structure of the set of events and a number of postulates, considered compelling. Let $g(A|B) \in [0,1]$ be a conditional belief degree, A, B being events in a Boolean algebra, with $B \neq \emptyset$:

i) $g(A \cap C|B) = F(g(A|C \cap B), g(C|B))$ (if $C \cap B \neq \emptyset$);
ii) $g(\overline{A}|B) = n(g(A|B))$, $B \neq \emptyset$, where $\overline{A}$ is the complement of A;
iii) function F is supposed to be twice differentiable, with a continuous second derivative, while function n is twice differentiable.

On such a basis, Cox claimed $g(A|B)$ is necessarily isomorphic to a probability measure.

This result is important to recall here because it has been repeated *ad nauseam* in the literature of artificial intelligence to justify probability and Bayes rule as the only reasonable approach to represent and process numerical belief degrees [Horvitz et al, 1986; Cheeseman, 1988; Jaynes, 2003]. However some reservations must be made. First, the original proof by Cox turned out to be faulty – see [Paris, 1994] for another version of this proof based on a weaker condition iii) : it is enough that F be strictly monotonically increasing in each place. Moreover, Halpern [1999a,b] has shown that the result does not hold in finite spaces, and needs an additional technical condition to get it in the infinite setting. Independently of these technical issues, it should be noticed that postulate (i) sounds natural only if one takes Bayes conditioning for granted; the second postulate requires self-duality, i.e., it rules out all other approaches to uncertainty considered in the rest of this chapter and in the next one; it forbids the representation of uncertainty due to partial ignorance as seen later on. Noticing that $P(A|B)$ can be expressed in terms of $P(A \cap B)$ and $P(\overline{A} \cap B)$, an alternative option would be to start with assuming $g(A|B)$ to be a function of $g(A \cap B)$ and $g(\overline{A} \cap B)$), adding the postulate $g((A|B)|C) = g(A|B \cap C)$, if $B \cap C \neq \emptyset$, but dropping (iii). This could lead to a general study of conditional belief as outlined in [Dubois et al, 2010]. The above comments seriously weaken the alleged universality of Cox results.

3.3 Bayes Rule: Revision vs. Prediction

Assuming that a single probability measure is available, the additivity property of probability theory implies two noticeable results for conditional probabilities, that are instrumental in practice:

- The theorem of total probability: If $\{C_1, \ldots, C_k\}$ forms a partition of S, then

$$P(A) = \sum_{i=1}^{k} P(A \mid C_i) P(C_i).$$

- Bayes theorem

$$P(C_j \mid A) = \frac{P(A \mid C_j) P(C_j)}{\sum_{i=1}^{k} P(A \mid C_i) P(C_i)}.$$

The first result makes it possible to derive the probability of an event in a general context S given the probabilities of this event in various subcontexts $C_1, \ldots, C_k$, provided they form a partition of the set of possible states, and if probabilities of these subcontexts are available. Bayes theorem is useful to solve classification problems: suppose k classes of objects forming a partition of S. If the probability that objects in each class C_j satisfy property A is known, as well as prior probabilities of classes C_j, then if a new object is presented that is known to possess property A, it is easy to compute the probability $P(C_j \mid A)$ that this object belongs to class C_j. Diagnosis problems are of the same kind, replacing “class” by “disease” and “observed property” by “symptom”. The use of conditional probabilities in Bayesian networks first proposed in [Pearl, 1988] is extensively discussed in Chapter 8 of Volume 2 of this treatise.

Most of the time, the information encoded in a probability distribution refers to some population. It represents *generic* information, with a frequentist meaning. One can use this information to infer beliefs about a particular situation, in which one has made partial, but unambiguous observations. This task is referred to as *prediction*. If $P(A \mid C)$ is the (frequentist) probability of event A in context C, one measures the agent's confidence $g(A \mid C)$ in proposition A, when only information C is known, by the quantity $P(A \mid C)$, assuming the current situation is typical of context C. The agent's belief about proposition A is updated from $g(A) = P(A)$ to $g(A \mid C) = P(A \mid C)$ after observing that C is true in the current situation, and nothing else. Conditioning is thus used to update the agent's contingent beliefs about the current situation by exploiting generic information. For instance, probability measure P represents medical knowledge (often compiled as a Bayesian network). Contingent information C represents test results for a given patient. Conditional probability $P(A \mid C)$ is then the probability that disease A is present for patients with test results C; this value also measures the (contingent) probability that the particular patient under consideration has disease A. We can remark that, under inference of this kind, the probability measure P does not change. One only applies generic knowledge to the reference class C, a process called *focalization*.

In the context of subjective probability à la De Finetti, to say that a probability distribution P is known means to know $P(A \mid C)$ for all events in all contexts. The agent only chooses the conditional probability of the event of interest in the context that is in agreement with the information on the current situation.

These views of conditioning differ from a revision process leading to a change of probability measure. Indeed some authors justify conditional probability in terms of belief revision [Gärdenfors, 1988]. The quantity $P(A \mid C)$ is then viewed as the *new* probability of A when the agent learns that C occurred. A basic principle of belief revision is minimal change: the agent revises its beliefs minimally while absorbing the new information item, interpreted as the constraint $P(C) = 1$. Under this view, the nature of the prior probability, and of the input information is the same, as is the posterior probability. In this revision scenario (see Chapter 14 of this volume), the probability function can be generic (e.g., frequentist, population-based) or singular (a subjective probability) and the input information is of the same kind as P (we learn that C has actual probability 1). The revision problem is then defined as follows: find a new probability P' as close as possible to P such that $P'(C) = 1$, which obeys minimal change [Domotor, 1985]. Using a suitable measure of relative information (e.g., Kullback-Leibler relative entropy) it can be shown that $P'(A) = P(A \mid C), \forall A$.

This revision scenario contrasts with the one of making predictions based on generic knowledge (in the form of a probability measure P describing the behavior of a population) and singular information items describing a situation of interest, even if the same tool, conditional probability, is used. As will be seen later on in this chapter and in the next one, the two tasks (revision vs. prediction) will no longer be solved by the same form of conditioning in more general uncertainty theories.

3.4 Probability Distributions and Partial Ignorance

The so-called *Bayesian* approach to subjective probability theory postulates the unicity of the probability distribution as a preamble to any kind of uncertainty modeling (see, for instance, [Lindley, 1982]), which could read as follows: any state of knowledge is representable by a single probability distribution. Note that indeed, if, following the fair bet procedure of De Finetti, the agent decides to directly assign subjective probabilities via buying prices to all *singletons* in S, the coherence principle forces this agent to define a unique probability distribution in this case. However, it is not clear that the limited perception of the human mind makes the agent capable of providing real numbers with infinite precision in the unit interval as prices. The measurement of subjective probability should address this issue in some way. If one objects that perhaps the available knowledge of the agent hampers the assignment of precise prices, the Bayesian approach sometimes resorts to selection principles such that the Laplace Principle of Insufficient Reason that exploits symmetries of the problem, or the maximal entropy principle [Jaynes, 1979; Paris, 1994]. Resorting to the latter in the subjectivist setting is questionable because it would select the uniformly distributed probability whenever it is compatible with the imprecise probabilistic information, even if imprecise probabilities suggest another trend.

Applying the Bayesian credo as recalled above forces the agent to use a single probability measure as the universal tool for representing uncertainty whatever its source. This stance leads to serious difficulties already pointed fifty years ago [Shafer, 1976]. For one, it means we give up making any difference between uncertainty due to incomplete information or ignorance, and uncertainty due to a purely random process, the next outcome of which cannot be predicted. Take the example of die tossing. The uniform probability assignment corresponds to the assumption that the die is fair. But if the agent assigns equal prices to bets assigned to all facets, how can we interpret it? Is it because the agent is sure that the die is fair and its outcomes are driven by pure randomness (because, say, they could test it hundreds of times prior to placing the bets)? Or is it because the agent who is given this die, has just no idea whether the die is fair or not, so has no reason to put more money on one facet than on another one? Clearly the epistemic state of the agent is not the same in the first situation and in the second one. But the uniformly distributed probability function is mute about this issue.

Besides, the choice of a set of mutually exclusive outcomes depends on the chosen language, e.g., the one used by the information source, and several languages or points of view can co-exist in the same problem. As there are several possible representations of the state space, the probability assignment by an agent will be language-dependent, especially in the case of ignorance: a uniform probability on one state space may not correspond to a uniform one on another encoding of the same state space for the same problem, while in case of ignorance this is the only representation left to the betting agent. Shafer [1976] gives the following example. Consider the question of the existence of extra-terrestrial life, about which the agent has no idea. If the variable v refers to the claim that life exists outside our planet ($v = l$), or not ($v = \neg l$), then the agent proposes $P_1(l) = P_1(\neg l) = \frac{1}{2}$ on $S_1 = \{l, \neg l\}$.

However it makes sense to distinguish between animal life (*al*), and vegetal life only (*vl*), which leads to the state space $S_2 = \{al, vl, \neg l\}$. The ignorant agent is then bound to propose $P_2(al) = P_2(vl) = P_2(\neg l) = \frac{1}{3}$. As l is the disjunction of *al* and *vl*, the distributions P_1 and P_2 are not compatible with each other, while they are supposed to represent ignorance. A more casual example comes from noticing that expressing ignorance by means of a uniform distribution for $v \in [a,b]$, a positive interval, is not compatible with a uniform distribution on $v' = \log v \in [\log(a), \log(b)]$, while the agent has the same ignorance on v and v'.

Finally, it is not easy to characterize a single probability distribution by assigning lottery prices to propositions that do not pertain to singletons of the state space. Probability theory and classical logic, understood as knowledge representation frameworks, do not get along very conveniently. A maximal set of propositions to each of which the same lower bound of probability strictly less than 1 is assigned is generally not deductively closed. Worse, the conditioning symbol in probability theory is not a standard Boolean connective. The values $Prob(A|B)$ and $Prob(B \to A) = Prob(\overline{B} \cup A)$ can be quite different from each other, and will coincide only if they are equal to 1 [Kyburg, Jr. and Teng, 2012]. A natural concise description of a probability distribution on the set of interpretations of a language is easily achieved by a Bayesian network, not by a weighted set of propositional formulas.

Besides, in first-order logic, we should not confuse an uncertain universal conjecture [Gaifman and Snir, 1982] (for instance, $Prob(\forall x, P(x) \to Q(x)) = \alpha$) with a universally valid probabilistic statement (for instance, $\forall x, Prob(P(x) \to Q(x)) = \alpha$, or $\forall x, Prob(Q(x)|P(x)) = \alpha$). Extensions of Bayesian networks to first-order logical languages can be found in [Milch and Russell, 2007]. Finally we give a number of references to works that tried to reconcile probabilistic and logical representations (propositional, first-order, modal) in various ways: [Halpern, 1990; Bacchus, 1991; Nilsson, 1993; Abadi and Halpern, 1994; Marchioni and Godo, 2004; Jaeger, 2001; Halpern and Pucella, 2002, 2006; Jaeger, 2006]. See Chapter 9 in Volume 2 for a detailed account of probabilistic relational languages.

The above limitations of expressive power of single probability distributions have motivated the emergence of other approaches to uncertainty representations. Some of them give up the numerical setting of degrees of belief and use ordinal or qualitative structures considered as underlying the former subjectivist approaches. For instance [Renooij and van der Gaag, 1999; Parsons, 2001; Bolt et al, 2005; Renooij and van der Gaag, 2008] for works that try to provide a qualitative counterpart of Bayesian nets. Another option is to tolerate incomplete information in the probabilistic approaches, which leads to different mathematical models of various level of generality. They are reviewed in the rest of this chapter and in the next chapter in this volume. Possibility theory is the simplest approach of all, and is found in both qualitative and quantitative settings [Dubois and Prade, 1998].

3.5 Conditional Events and Big-Stepped Probabilities

Instead of considering a conditional probability function $P(\cdot \mid C)$ as a standard probability distribution on C, De Finetti [1936] was the first scholar to consider the set of conditional probabilities $\{P(A \mid C) : A \subseteq S, C \neq \emptyset\}$ as a probability assignment to *three-events, or conditional events* $A \mid C$. A conditional event can be informally understood as a conditional statement or an if-then rule: if all the currently available information is described by C, then conclude that A holds.

A three-event is so called because it partitions the state space S into three disjoint sets of states s:

- Either $s \in A \cap C$; s is called an example of the rule "if C then A". The three-event is considered as true at state s, which is denoted by $t(A \mid C) = 1$;
- or $s \in \overline{A} \cap C$; s is called a counter-example of the rule "if C then A". The three-event is considered as false at state s, which is denoted by $t(A \mid C) = 0$;
- or $s \in \overline{C}$; then the rule "if C then A" is said not to apply to s. In this case the three-event takes a third truth-value at s, which is denoted by $t(A \mid C) = I$ where I stands for inapplicable.

A three-event $A \mid C$ can thus be interpreted as a pair $(A \cap C, \overline{A} \cap C)$ of disjoint sets of examples and counter-examples. A qualitative counterpart of Bayes rule holds, noticing that as the set-valued solutions of the equation $A \cap C = X \cap C$ are all sets $\{X : A \cap C \subseteq X \subseteq A \cup \overline{C}\}$, which is another possible representation of $A \mid C$ (as an interval in the Boolean algebra of subsets of S). This definition of conditional events as pairs of subsets suggests a natural consequence relation between conditional events defined as follows [Dubois and Prade, 1994]:

$$B \mid A \vDash D \mid C \Leftrightarrow A \cap B \vDash C \cap D \text{ and } C \cap \overline{D} \vDash A \cap \overline{B}$$

which reads: all examples of $B|A$ are examples of $D|C$ and all counter-examples of $D \mid C$ counter-examples of $B \mid A$. Note that only the second condition coincides with the deductive inference between material conditional counterparts of the three-events. Material conditionals highlight counter-examples of rules, not examples. When ordering the truth-values as $0 < I < 1$, this inference also reads $B|A \vDash D|C \Leftrightarrow t(B \mid A) \leq t(D \mid C)$.

Representing if-the rules by conditional events avoids some paradoxes of material implications, such as the confirmation paradox: in the material implication representation, the rule *if C then A* is the same as its contrapositive version *if* $\overline{A}$ *then* $\overline{C}$. If we use material implication, we are bound to say that an example confirms a rule if it makes this material implication true. So, both $s_1 \in A \cap C$ and $s_2 \in \overline{A} \cap \overline{C}$ confirm the rule. But this is questionable: suppose the rule means all ravens are black. Then meeting a white swan would confirm that all ravens are black [Hempel, 1945]. This anomaly does not occur with conditional events as $A \mid C$ is not equivalent to $\overline{C} \mid \overline{A}$: they have the same counterexamples (e.g., white ravens) since they have the same material conditional representations, but they do not have the same examples:

s_2 is an example of $\overline{C} \mid \overline{A}$ (e.g., a white swan), but this three-event does not apply to $s_1 \in A \cap C$ [Benferhat et al, 2008].

It is worth noticing that a conditional probability $P(A \mid C)$ is indeed the probability of a conditional event $A \mid C$ since $P(A \mid C)$ is entirely determined by the pair $(P(A \cap C), P(\overline{A} \cap C))$. Moreover, if all probabilities of singletons are positive, and $B \mid A \vDash D \mid C$, it is clear that $P(B \mid A) \leq P(D \mid C)$.

A three-valued logic for conjunctions of conditional events was developed by Dubois and Prade [1994]. A three-valued extension of standard conjunction is used where the third truth-value I is a semi-group identity. The three-valued logic truth-table for conjunction and the above inference rule offer an alternative simple semantics for the non-monotonic inference system P [Kraus et al, 1990] that captures exception-tolerant reasoning, where conditional events $B \mid A$ model generic rules of the form: *generally if A then B* (see also the section on non-monotonic inference in Chapter 2 in this volume and Section 4.2.1 of the present chapter). Non-monotonicity manifests itself by the fact that the inference $B \mid A \vDash B \mid A \cap C$ does not hold (the latter has less examples than the former), so that, like in probability theory, conditional events $B \mid A$ and $\overline{B} \mid (A \cap C)$ can coexist in the same rule base without ruling out any possible world (contrary to material conditionals in propositional logic that would enforce $A \cap C = \emptyset$). Under this logic, to infer a plausible conclusion F from a state of knowledge described by the epistemic set E, and a conditional base (a set of conditional events) $\mathscr{C}$ that encodes generic information, means to infer the conditional event $F \mid E$ from a conditional event obtained as a suitable conjunction of a subset of conditional events in $\mathscr{C}$ [Benferhat et al, 1997].

Note also that under this inference scheme, the conditional event $A \cap B \mid C$ follows from $\mathscr{C} = \{A \mid C, B \mid C\}$, so that the set of plausible conclusions obtained from C will be deductively closed. But as pointed our earlier, $P(A \cap B \mid C) > 1 - \theta$ does not follow from $P(A \mid C) \geq 1 - \theta$ and $P(B \mid C) > 1 - \theta$, however small θ may be. In particular, if we minimally define A as an accepted belief whenever $P(A \mid C) > P(\overline{A} \mid C)$ (in other words $P(A \mid C) > 1/2$), we see that contrary to what happens with conditional events, a set of probabilistically accepted beliefs will not be closed in the sense of classical deduction. To ensure compatibility between symbolic inference between conditional events and accepted beliefs in the above sense, we can restrict the set of probability distributions to a subset for which deductive closure will be respected. This kind of probability measure is called *big-stepped probability* and is defined as follows by the condition:

$$\forall i < n-1, p_i > \sum_{j=i+1,..,n} p_j \text{ where } p_i = P(s_i) \text{ and } p_1 > .. > p_{n-1} \geq p_n > 0.$$

For an example of big-stepped probability distribution when $n = 5$, consider $p_1 = 0.6, p_2 = 0.3, p_3 = 0.06, p_4 = 0.03, p_5 = 0.01$. This type of exponential-like (or super-decreasing) probability distributions are at odds with uniform distributions. They offer a full-fledged probabilistic semantics to the logic of conditional events and Kraus, Lehmann and Magidor [1990]'s system P for coping with exceptions in rule-based systems [Benferhat et al, 1999b; Snow, 1999].

4 Possibility Theory

Basic building blocks of possibility theory go back to a seminal paper by Zadeh [1978] and further works by Dubois and Prade [1988] quite independently of the works of an English economist, Shackle [1961] who had outlined a similar theory some thirty years before (in terms of so-called *degrees of surprize* to be equated to degrees of impossibility). Actually, Zadeh and Shackle did not have the same intuitions in mind. Zadeh viewed his possibility distributions as representing flexible constraints representing pieces of fuzzy information in natural language (viz. "what is the possibility that John is more than 30 years old assuming he is *young*"?). In contrast Shackle tried to offer a representation of how the human mind handles uncertainty that is supposedly more faithful than probability theory. After the publication of Zadeh's paper, it soon became patent that possibility distributions were not necessarily generated from the representation of gradual properties in natural language (like *young*), but that they allowed to formalize a gradual notion of epistemic states by extending the disjunctive view of sets to fuzzy sets, whereby degrees of possibility, understood as plausibility, can be assigned to interpretations induced by any propositional language.

Possibility measures are maximum-decomposable for disjunction. There have companion set-functions called necessity measures, obtained by duality, that are minimum-decomposable for conjunction. They can be completed by two other set-functions that use the same basic setting. This general framework is first recalled in the following subsections. Then the distinction between qualitative and quantitative possibility theories is recalled. Qualitative possibility theory is best couched in possibilistic logic, which is briefly outlined. This section is completed by an exposition of the relationships between qualitative possibility theory and non-monotonic reasoning, and the modeling of default rules. We end the section by a possibility-theory rendering of formal concept analysis, which was originally developed in a very different perspective.

4.1 General Setting

Consider a mapping π_v from S to a totally ordered scale L, with top denoted by 1 and bottom by 0. It can be the unit interval as suggested by Zadeh, or generally any finite chain such as $L = \{0, 0.1, 0.2, \ldots, 0.9, 1\}$, or a totally ordered set of symbolic grades. The possibility scale can be the unit interval as suggested by Zadeh, or generally any finite chain, or even the set of non-negative integers. For convenience, it is often assumed that the scale L is equipped with an order-reversing map denoted by $\lambda \in L \mapsto 1 - \lambda$. More generally L can be a complete lattice with a top and a bottom element, denoted by 1 or 0 respectively. The larger $\pi_v(s)$, the more possible, i.e., plausible the value s for the variable v, that supposedly pertains to some attribute (like the age of John in Section 2.4). The agent information about v is captured by π_v called a *possibility distribution*. Formally, the mapping π is the membership function of a fuzzy set [Zadeh, 1978], where membership grades are interpreted in terms of plausibility. If the possibility distribution stems from gradual linguistic properties, plausibility is measured in terms of distance to fully plausible situations, not in terms of, e.g., frequency. Function π represents the state of knowledge of an agent (about the actual state of affairs), also called an *epistemic state* distinguishing what is plausible from what is less plausible, what is the normal course of things from what is not, what is surprising from what is expected. It represents a flexible restriction on what is the actual state with the following conventions (similar to probability, but opposite to Shackle's potential surprise scale)[4]:

- $\pi(s) = 0$ means that state s is rejected as impossible;
- $\pi(s) = 1$ means that state s is totally possible (= plausible).

If the universe S is exhaustive, at least one of the elements of S should be the actual world, so that $\exists s, \pi(s) = 1$ (normalised possibility distribution). This condition expresses the consistency of the epistemic state described by π. Distinct values may simultaneously have a degree of possibility equal to 1. In the Boolean case, π is just the characteristic function of a subset $E \subseteq S$ of mutually exclusive states, ruling out all those states considered as impossible. Possibility theory is thus a (fuzzy) set-based representation of incomplete information. There are two extreme cases of imprecise information

- *Complete ignorance*: without information, only tautologies can be asserted. It is of the form $v \in S$, corresponding to the possibility distribution $\pi_v^?(s) = 1, \forall s \in S$.
- *Complete knowledge*: it is of the form $v = s_0$ for some value $s_0 \in S$, corresponding to the possibility distribution $\pi_v^{s_0}(s) = 1$ if $s = s_0$ and 0 otherwise. Note that it is the value 0 that brings information in π_v.

Possibility theory is driven by the *principle of minimal specificity*. It states that *any hypothesis not known to be impossible cannot be ruled out*. It is a minimal commitment, cautious information principle. Basically, we must always try to maximize possibility degrees, taking constraints into account. Measures of possibilistic specificity have been proposed in a way similar to probabilistic entropy [Higashi and Klir, 1982].

4.1.1 The Two Basic Set-Functions

Plausibility and certainty evaluations, induced by the information represented by a distribution π_v, pertaining to the truth of proposition $v \in A$ can then be defined. We speak of degrees of possibility and necessity of event A:

$$\Pi(A) = \max_{s \in A} \pi_v(s); \quad N(A) = 1 - \Pi(\overline{A}) = \min_{s \notin A} 1 - \pi_v(s) \tag{8}$$

[4] If $L = \mathbb{N}$, the conventions are opposite: 0 means possible and ∞ means impossible.

By convention $\Pi(\emptyset) = 0$ and then $N(S) = 1$. $\Pi(S) = 1$ (hence $N(\emptyset) = 0$) follows if π_v is normalized. The symbol $1 - (\cdot)$ should not suggest these degrees are numerical. It is just the order-reversing map on L.

When distribution π_v takes value on the binary scale $\{0,1\}$, i.e., there is a subset $E \subseteq S$ such that $\pi_v(s) = 1 \Leftrightarrow s \in E$, it is easy to see that $\Pi(A) = 1$ if and only if the proposition $v \in A$ is not inconsistent with the information item $v \in E$, i.e., if $A \cap E \neq \emptyset$. Likewise, $N(A) = 1$ if and only if proposition $v \in A$ is implied by the information item $v \in E$ (since $E \subseteq A$). $\Pi(A) = 0$ means that it is impossible that the assertion $v \in A$ is true if $v \in E$ is true. $N(A) = 1$ expresses that the assertion $v \in A$ is certainly true if $v \in E$ is true.

Functions N and Π are tightly linked by the duality property $N(A) = 1 - \Pi(\overline{A})$. This feature highlights a major difference between possibility and necessity measures and probability measures that are self dual in the sense that $P(A) = 1 - P(\overline{A})$.

The evaluation of uncertainty in the style of possibility theory is at work in classical and modal logics. If K is a set of propositional formulas in some language, suppose that E is the set of its models. Consider a proposition p which is the syntactic form of the proposition $v \in A$, 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. Of course, the presence of p inside K encodes $N(A) = 1$, while the presence of its negation $\neg p$ in K encodes $\Pi(A) = 0$. In contrast, in the propositional language of K, one cannot encode $N(A) = 0$ nor $\Pi(A) = 1$, e.g., that $v \in A$ is unknown. To do this inside the language, one must use the formalism of modal logic (see Chapter 2 in this volume), that prefixes propositions by modalities of possibility ($\Diamond$) and necessity ($\Box$): in a modal base K^{mod}, $\Diamond p \in K^{mod}$ directly encodes $\Pi(A) = 1$, and $\Box p \in K^{mod}$ encodes $N(A) = 1$ (the latter merely encoded by $p \in K$ in propositional logic). The duality relation between Π and N is very well known in modal logic, where it reads $\Diamond p = \neg\Box\neg p$. A simple modal logic (a very elementary fragment of the KD logic), called MEL (for minimal epistemic logic), has been defined by Banerjee et Dubois [2014] with a semantics in terms of non-empty subsets of interpretations ($\{0,1\}$-valued possibility distributions (a similar idea was first suggested by Mongin [1994]). The satisfaction of $\Box p$ by an epistemic set E means that $E \subseteq A$, if p encodes $v \in A$.

In the possibilistic setting one distinguishes three extreme epistemic attitudes pertaining to an information item $v \in A$:

- the certainty that $v \in A$ is true: $N(A) = 1$, hence $\Pi(A) = 1$;
- the certainty that $v \in A$ is false: $\Pi(A) = 0$, hence $N(A) = 0$:
- ignorance pertaining to $v \in A$: $\Pi(A) = 1$, and $N(A) = 0$.

These attitudes can be refined as soon as L contains at least one value differing from 0 or 1 leading to situations where $0 < N(A) < 1$ or $0 < \Pi(A) < 1$.

It is easy to verify that possibility and necessity measures saturate inequalities (2) verified by capacities:

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)). \tag{9}$$

$$N(A \cap B) = \min(N(A), N(B)). \tag{10}$$

Possibility measures are said to be *maxitive* and are fully characterized by the maxitivity property (9) in the finite case; necessity measures are said to be *minitive* and are fully characterized by the minitivity property (10) in the finite case, including when these functions take values in $[0,1]$.

In general possibility and necessity measures do not coincide. It is impossible for a set function to be at the same time maxitive and minitive for all events, except in case of complete knowledge ($E = \{s_0\}$). Then $N = \Pi$ also coincide with a Dirac $\{0,1\}$-valued probability measure.

Observe that we only have

$$N(A \cup B) \geq \max(N(A), N(B)) \text{ and } \Pi(A \cap B) \leq \min(\Pi(A), \Pi(B)),$$

and it may occur that the difference is maximal. Indeed in the $\{0,1\}$-valued case, if it is not known whether A is true or false (namely, $A \cap E \neq \emptyset$ and $\overline{A} \cap E \neq \emptyset$), then $\Pi(A) = \Pi(\overline{A}) = 1$ and $N(A) = N(\overline{A}) = 0$; however, by definition $\Pi(A \cap \overline{A}) = \Pi(\emptyset) = 0$ and $N(A \cup \overline{A}) = N(S) = 1$.

4.1.2 Two Decreasing Set Functions. Bipolarity

Yet another set function Δ and its dual companion ∇ (first introduced in 1991, see [Dubois and Prade, 1998]) can be naturally associated with the possibility distribution π_v in the possibilistic framework:

$$\Delta(A) = \min_{s \in A} \pi_v(s); \quad \nabla(A) = 1 - \Delta(\overline{A}) = \max_{s \notin A} 1 - \pi_v(s) \tag{11}$$

Observe first that in contrast with Π and N, Δ and ∇ are decreasing functions with respect to set inclusion (hence to the logical consequence relation). Function Δ is called *strong possibility* or *guaranteed possibility* since inside set A, the degree of possibility is never less than $\Delta(A)$ (while Π is only a weak possibility degree that just measures consistency); dually, function ∇ is a measure of weak necessity, while N is a measure of strong necessity. Besides, the following inequality hold:

$$\forall A, \max(\Delta(A), N(A)) \leq \min(\Pi(A), \nabla(A))$$

provided that both π_v and $1 - \pi_v$ are normalised.

Characteristic properties of Δ and ∇ are:

$$\Delta(A \cup B) = \min(\Delta(A), \Delta(B)); \quad \Delta(\emptyset) = 1. \tag{12}$$

$$\nabla(A \cap B) = \max(\nabla(A), \nabla(B)); \quad \nabla(S) = 0. \tag{13}$$

From the standpoint of knowledge representation, it is interesting to consider the case when the possibility distribution π_v only takes a finite number of distinct values $\alpha_1 = 1 > \cdots > \alpha_n > \alpha_{n+1} = 0$. It can then be described by n nested subsets $E_1 \subseteq \cdots \subseteq E_i \subseteq \cdots \subseteq E_n$ where $\pi_v(s) \geq \alpha_i \Leftrightarrow s \in E_i$. One can then verify that $\Delta(E_i) \geq \alpha_i$, while $N(E_i) \geq 1 - \alpha_{i+1}$ for $i = 1, \ldots, n$, and that $\pi_v(s) = \max_{E_i \ni s} \Delta(E_i) = \min_{E_i \not\ni s}(1 - N(E_i))$ (with conventions $\max_\emptyset = 0$ et $\min_\emptyset = 1$). A distribution π_v can thus be seen as a weighted disjunction of sets E_i, from the point of view of Δ, and as a weighted conjunction of sets E_i from the point of view of N. The reading of π_v viewed from Δ offers a positive understanding of the possibility distribution, expressing to which extent each value is possible, while viewed from N, π_v expressed to what extent each value is not impossible (since each value s is all the more impossible as it belongs to fewer subsets E_i).

These positive and negative flavors respectively attached to Δ and N lay the foundation of a *bipolar* representation of information in possibility theory [Benferhat et al, 2008]. The idea of bipolarity refers to an explicit handling of positive or negative features of information items [Dubois and Prade, eds.]. There are several forms of bipolarity and we only focus on the case when it comes from the existence of distinct sources of information. In the possibilistic setting, *two* possibility distributions δ_v and π_v are instrumental to respectively represent values that are guaranteed possible for v and values that are just known to be not-impossible (because not ruled out). The concept of bipolarity applies to representing knowledge as well as preferences. These distributions are differently interpreted: when representing knowledge $\delta_v(s) = 1$ means that s is certainly possible because this value or state has been actually observed, and, when representing preferences, s is an ideal choice. Moreover, when representing knowledge, $\delta_v(s) = 0$ just means that nothing is known about this value that has not been observed, and, when representing preferences, that the choice s is not at all attractive. In contrast, when representing knowledge, $\pi_v(s) = 1$ means that s is not impossible (just feasible when representing preference), but $\pi_v(s) = 0$ means that s is completely ruled out (or not acceptable for preferences). Intuitively, any state that is guaranteed possible should be among the non-impossible situations. So there is a coherence condition to be required: $\delta_v \leq \pi_v$. It corresponds to a standard fuzzy set inclusion). In possibilistic logic presented further on, the distribution π_v stems from constraints of the form $N(A_i) \geq \eta_i$, and distribution δ_v from statements of the form $\Delta(B_j) \geq \delta_j$ where $A_i \subseteq S, B_j \subseteq S$, and $\eta_i \in L, \delta_j \in L$. The idea of bipolar representation is not limited to possibility theory, even if it was not often considered in other frameworks (see [Dubois et al, 2000a]).

4.1.3 Possibility and Necessity of Fuzzy Events

The set functions Π, N, Δ et ∇ can be extended to fuzzy sets. The (weak) possibility of a fuzzy event F is defined by $\Pi(F) = \sup_s \min(F(s), \pi_v(s))$ [Zadeh, 1978]; still using duality, the necessity of a fuzzy event then reads $N(F) = 1 - \Pi(\overline{F}) = \inf_s \max(F(s), 1 - \pi_v(s))$. Functions Π and N still satisfy, respectively, maxitivity (9) and minitivity (10) properties. The values $\Pi(F)$ and $N(F)$ turn out to be special cases of Sugeno integrals (see Chapter 16 in this volume). Possibility and necessity of fuzzy events are instrumental to evaluate the extent to which a flexible condition is satisfied by an ill-known piece of data [Cayrol et al, 1982]; in particular, if $\pi_v = F$, only $N(F) \geq 1/2$ obtains, which at first glance may be questionable. To get $N(F) = 1$, the condition $\forall s\ \pi_v(s) > 0 \Rightarrow F(s) = 1$ is needed, which means the inclusion of the support of π in the core of F so that any value that is possible even to a very low extent be fully in agreement with F. Such evaluations have been applied to fault diagnosis problems using a qualitative handling of uncertainty, where one may separate anomalies that more or less certainly appear when a failure occurs, from anomalies that more or less possibly appear [Cayrac et al, 1996; Dubois et al, 2001]. Functions Δ and ∇ extend similarly to fuzzy events as $\Delta(F) = \inf_s \max(1 - F(s), \pi_v(s))$, letting $\nabla(F) = 1 - \Delta(\overline{F})$ by duality, while preserving respective properties (12) and (13).

Set functions N and Δ on fuzzy events are also very useful to represent fuzzy if-then rules (see also Chapter 10 in this volume) of the form *the more v is F, the more it is* sure *that y is G*, and *the more v is F, the more it is* possible *that y is G* respectively, where F (but possibly G as well) are gradual properties represented by fuzzy sets [Dubois and Prade, 1996]. Indeed, the first type of rule expresses a constraint of the form $N(G) \geq F(s)$ while the second one is better modeled by the inequality $\Delta(G) \geq F(s)$. However, the first type of rule, where $1 - F(s)$ is viewed as the degree of possibility that the conclusion G is false, while in the second type of rule $F(x)$ is the minimal degree of possibility that the conclusion G holds, which corresponds to the following possibility distributions on the joint domain of (x, y):

$$\pi_{x,y}(s,t) \leq \max(1 - F(s), G(t)) \text{ and } \pi_{x,y}(s,t) \geq \min(F(s), G(t)).$$

Definitions of the strong necessity and possibility functions compatible with these inequalities are not the ones based on Zadeh's weak possibility of a fuzzy event. Based on the following equivalence: $c \leq \max(a, 1-b) \Leftrightarrow (1-a) \to (1-c) \geq b$, where $\to$ is Gödel implication

$$u \to v = \begin{cases} 1 \text{ si } u \leq v, \\ v \text{ otherwise,} \end{cases}$$

the following extensions of strong necessity and possibility of fuzzy events N et Δ must be used: $N(G) = \inf_s (1 - F(s)) \to (1 - \pi_v(s))$ and $\Delta(G) = \inf_s F(s) \to \pi_v(s)$. These evaluations do reduce to strong necessity and possibility of standard events, like the ones in the previous paragraph, but the necessity function satisfies $N(G) = 1$ when $\pi_v = G$ (since we expect some equivalence between statements such as *it is sure that John is young* and *John is young*). Likewise, $\Delta(G) = 1$ when $\pi_v = G$. See [Dubois et al, 2017a] for a systematic analysis of extensions of the four set functions of possibility theory to fuzzy events. The two types of fuzzy rules reflect a bipolar view of a standard rule R of the *if* $v \in A$ *then* $y \in B$, which, on a Cartesian product of domains $S \times T$ can be represented either by the constraint $R(s,t) \geq (A \times B)(s,t)$ pointing out examples, or by the constraint $\overline{R}(s,t) \geq (A \times \overline{B})(s,t) \Leftrightarrow R(s,t) \leq (\overline{A} + B)(s,t)$ excluding counter-examples, where the overbar means complementation and where $A + B = \overline{\overline{A} \times \overline{B}}$. The view of an if-then rule as a conditional event $B|A$ is thus retrieved.

4.1.4 Conditioning in Possibility Theory: Qualitative vs. Quantitative Settings

Since the basic properties in possibility are based on minimum, maximum and an order-reversing map on the uncertainty scale ($1-(\cdot)$ on the unit interval, and $1-\alpha_k = \alpha_{m-k}$) on a bounded chain $\{\alpha_0, \cdots, \alpha_m\}$), it is not imperative to use a numerical setting for the measurement of possibility and necessity. When the set functions take values in the unit interval, we speak of *quantitative possibility theory*. When they take values in a bounded chain, we speak of *qualitative possibility theory* [Dubois and Prade, 1998]. In both cases, possibility theory offers a simple, but non trivial, approach to non-probabilistic uncertainty. The two versions of possibility theory diverge when it comes to conditioning. In the qualitative case, there is no product operation, and the counterpart of Bayes rule is naturally expressed replacing it by the minimum operation on the bounded chain L:

$$\Pi(A \cap B) = \min(\Pi(A \mid B), \Pi(B)). \tag{14}$$

This equation has no unique solution. In the spirit of possibility theory, one is led to select the least informative solution, according to minimal commitment, namely when $B \neq \emptyset$, and $A \neq \emptyset$:

$$\Pi(A \mid B) = \begin{cases} 1 & \text{if } \Pi(A \cap B) = \Pi(B), \\ \Pi(A \cap B) & \text{otherwise.} \end{cases} \tag{15}$$

This is just like conditional probability, except that we no longer make a division by $\Pi(B)$. When $\Pi(B) = 0$, $\Pi(A \mid B) = 1$ as soon as $A \neq \emptyset$. It reflects the idea than you may destroy available information when conditioning on an impossible event. Conditional necessity is defined by duality as[5]:

$$N(A \mid B) = 1 - \Pi(\overline{A} \mid B) = \begin{cases} 0 & \text{if } \Pi(\overline{A} \cap B) = \Pi(B); \\ N(A \cup \overline{B}) & \text{otherwise.} \end{cases}$$

The least specific solution to equation (14) does capture an ordinal form of conditioning due to the following result:

$$N(A \mid B) > 0 \iff \Pi(A \cap B) > \Pi(\overline{A} \cap B)$$

when $\Pi(B) > 0$. Intuitively, it means that a proposition A is an accepted belief in context B if it is more plausible than its negation in this context. Like with probability, one may have that $\Pi(A \cap B) > \Pi(\overline{A} \cap B)$ while $\Pi(\overline{A} \cap B \cap C) > \Pi(A \cap B \cap C)$ in a more restricted context $B \cap C$. An alternative approach to conditional possibility is the one of Coletti and Vantaggi [2006], in which coherent possibility assessments

[5] The Bayesian-like rule in terms of necessity measures, $N(A \cap B) = \min(N(A \mid B), N(B))$, is trivial. Its least specific solution, minimizing necessity degrees, is $N(A \mid B) = N(A \cap B) = \min(N(A), N(B))$, which defines in turn $\Pi(A \mid B) = \Pi(\overline{B} \cup A)$. It comes down to interpret a conditional event as a material implication.

on conditional events are defined based on equation (14), in the style of De Finetti's conditional probability.

In the case of quantitative possibility theory, the lack of continuity of the set function $\Pi(A \mid B)$ in Equation (15) [de Cooman, 1997] has led to replace minimum by the product in this equation, mimicking conditional probability:

$$\Pi(A \mid B) = \frac{\Pi(A \cap B)}{\Pi(B)} \text{ provided that } \Pi(B) \neq 0.$$

As we shall see, it coincides with Dempster's rule of conditioning in evidence theory (see the next chapter in this volume). More generally, on the unit interval, the product can be extended to a triangular norm, and this general setting has been studied by Coletti and Vantaggi [2009] under the coherence approach in the style of De Finetti.

A major difference between possibility and probability theories concern independence. While stochastic independence between events with positive probability is a symmetric, negation-invariant, notion, since $Prob(B|A) = Prob(B)$ is equivalent to $Prob(A \cap B) = Prob(A) \cdot Prob(B)$ and to $Prob(B|\overline{A}) = Prob(B)$, this is no longer the case for possibilistic independence, several versions of which exist. For instance, in qualitative possibility theory, the equality $N(B|A) = N(B) > 0$ expresses that learning A does not question the accepted belief B and is not equivalent to $N(A|B) = N(A) > 0$ nor to $N(B|\overline{A}) = N(B) > 0$. Another form of independence is $N(B|A) = N(B) = N(\overline{B}|A) = N(\overline{B}) = 0$, which means that learning A leaves us ignorant about B; see [Dubois et al, 1999] for a complete study. There exist several definitions of conditional possibilistic independence between variables, in qualitative possibility theory, one being symmetric ($\Pi(x,y|z) = \min(\Pi(x|z), \Pi(y|z))$) and one being asymmetric ($\Pi(x|z) = \Pi(x|z,y)$); see [Ben Amor et al, 2002]. In the quantitative setting, independence between variables ($\forall x,y,z, \Pi(x|y,z) = \Pi(x|z)$) is symmetric since it is equivalent to $\forall x,y,z, \Pi(x,y|z) = \Pi(x|z) \cdot \Pi(y|z)$. The notion of possibilistic independence has also been studied in [Coletti and Vantaggi, 2006].

Conditional probability is the basis of representation of uncertain information in the form of Bayesian networks. There also exist graphical possibilistic representations in quantitative possibility theory, and in qualitative possibility theory as well (see Chapter 8 in Volume 2) and some variants of possibilistic independence are useful to develop local uncertainty propagation methods.

4.2 Qualitative Possibility Theory

The main application of qualitative possibility theory is the development of possibilistic logic, an extension of classical logic that handles qualitative uncertainty, and is useful for encoding non monotonic reasoning and dealing with inconsistency. Besides, the basic setting of formal concept analysis can be seen as a set-valued counterpart of possibility theory, which leads to an interesting parallel between the two theories. We first present possibilistic logic. Note that qualitative possibility theory can be used for decision under uncertainty. Decision-theoretic foundations of qualitative possibility theory are presented in Chapter 17 of this volume.

4.2.1 Possibilistic Logic

The building blocks of possibilistic logic [Dubois et al, 1994; Dubois and Prade, 2004] are pairs made of a (well-formed) formula of classical logic (propositional, or first order), and a weight (or level) which may be qualitative or numerical. The weights usually belong to a totally ordered scale, but may only belong to a lattice structure with a smallest and a greatest element).

Necessity-based possibilistic logic In its basic version, possibilistic logic only allows to consider conjunctions of pairs of the form (p, α) where p is a propositional logic formula associated with a weight α belonging to the interval $(0,1]$ (or to a finite totally ordered scale). The weight α is understood as a lower bound of a necessity mesure, i.e., the pair (p, α) encodes a constraint of the form $N(p) \geq \alpha$. It either corresponds to a piece of information (one is certain at level α that p is true), or a preference (p then represents a goal to be reached with priority α). The decomposability property of necessity mesures (10) ensures that we make no difference between $(p \wedge q, \alpha)$ and $(p, \alpha) \wedge (q, \alpha)$, and thus possibilistic bases, which are sets of such possibilistic pairs, can be expressed as conjunctions of weighted clauses.

Let $B^N = \{(p_j, \alpha_j) \mid j = 1, \ldots, m\}$ be a possibilistic base. At the semantic level, a possibility distribution π over the set of interpretations satisfies B^N (denoted by $\pi \models B^N$) if and only if $N(p) \geq \alpha_j, j = 1, \ldots, m$. The least specific possibility distribution that satisfies B^N exists and is of the form

$$\pi_B^N(s) = \min_{j=1,\ldots,m} \pi_{(p_j,\alpha_j)}(s) = \min_{j:\ s \models \neg p_j} 1 - \alpha_j,$$

where $\pi_{(p_j,\alpha_j)}(s) = 1$ if $s \models p_j$ and $1 - \alpha_j$ otherwise. Thus an interpretation s is all the more possible as it does not violate any formula p_j with a high priority level α_j, and $\pi \models B^N$ if and only if $\pi \leq \pi_B^N$.

The possibility distribution π_B^N provides a description "from above" (each pair (p_j, α_j) combined by min restricts the set of interpretations regarded as possible to some extent). It takes the form of a min-max combination, since $\pi_{(p_j,\alpha_j)}(s)$ is of the form $\max(M(p_j)(s), 1 - \alpha_j)$, where $M(p)$ denotes the characteristic function of the set of models of p. So, B^N can be expressed as a conjunction of weighted clauses, i.e., the extension of a conjunctive normal form, in agreement with the fact (p, α) and (q, α) is equivalent to $(p \wedge q, \alpha)$.

Basic possibilistic logic possesses the cut rule

$$(\neg p \vee q, \alpha); (p \vee r, \beta) \vdash (q \vee r, \min(\alpha, \beta)).$$

This rule is sound and complete for refutation, with respect to possibilistic semantics. It should be noticed that the probabilistic counterpart to this rule, namely

$$Prob(\neg p \vee q) \geq \alpha; Prob(p \vee r) \geq \beta) \vdash Prob(q \vee r) \geq \max(0, \alpha + \beta - 1)$$

is sound, but not complete with respect to probabilistic semantics. This is related to the fact that the deductive closure of possibilistic base $\{(p_j, \beta_j) \text{ with } \beta_j \geq \alpha\}_{j=1,n}$ only contains formulas with weights at least α, while this is wrong in general for the set of probabilistic formulas $\{p_j | Prob(p_j) \geq \alpha\}_{j=1,n}$ after closure with the corresponding resolution rule (except if $\alpha = 1$).

Dual possibilistic logic with guaranteed possibility weights A dual representation for possibilistic logic bases relies on guaranteed possibility functions. A formula is then a pair $[q, \beta]$, understood as the constraint $\Delta(q) \geq \beta$, where Δ is a guaranteed possibility (anti-)measure. It thus expresses that *all* the models of q are at least possible, at least satisfactory at level β. A Δ-base $B^\Delta = \{[q_i, \beta_i] \mid i = 1, \dots, n\}$ is then associated with the distribution

$$\pi_B^\Delta(s) = \max_{i=1,\dots,n} \pi_{[q_i,\beta_i]}(s) = \max_{i:\ s\ \models q_i} \beta_i,$$

with $\pi_{[q_i,\beta_i]}(s) = \min(M(q_i)(s), \beta_i)$. We define $\pi \models B^\Delta$ if and only if $\Delta(q_i) \geq \beta_i, \forall i = 1, \dots, n$, which is equivalent to $\pi \geq \pi_B^\Delta$. So, π_B^Δ provides a description "from below" of the distribution representing an epistemic state. Taking advantage of decomposability property (12) of guaranteed possibility measures, it is easy to see that the set $\{[p, \alpha], [q, \alpha]\}$ is equivalent to the formula $[p \vee q, \alpha]$. Then putting classical logical formulas in disjunctive normal form, we can always rewrite a dual possibilistic base B^Δ into an equivalent base where all formulas q_i are conjunctions of literals.

A base B^Δ in dual possibilistic logic can always be rewritten equivalently in terms of a standard possibilistic logic N-base B^N [Benferhat and Kaci, 2003; Benferhat et al, 2008], and conversely, in such a way that $\pi_B^N = \pi_B^\Delta$. However, note that Δ-based possibilistic logic obeys an inference rule different from the above resolution rule for N-bases: $[\neg p \wedge q, \alpha]; [p \wedge r, \beta] \vdash [q \wedge r, \min(\alpha, \beta)]$. It propagates guaranteed possibility levels in agreement with the decreasingness of set function Δ (indeed, if $r = \top$, and $q \vdash p$, then $\alpha = 1$ since $\Delta(\bot) = 1$, and the rule concludes $[q, \beta]$ from $[p, \beta]$).

A set of pieces of possibilistic Boolean information (with a finite number of possibility levels) can thus be represented by a possibility distribution on interpretations, but also in a more compact manner under the form of a finite set of formulas associated either with a certainty (resp. priority) level, or with a level of guaranteed possibility (resp. satisfaction) when modeling knowledge (resp. preferences). Moreover, graphical representations of possibilistic bases in terms of possibilistic networks (either based on qualitative or on quantitative conditioning) have been proposed, with exact translations from one type of representation to the other [Benferhat et al, 2002]. For an introduction to possibilistic networks and their algorithms, the reader is referred to Chapter 8 in Volume 2. Possibilistic networks are also useful for preference modeling [Ben Amor et al, 2017] (see also Chapter 7 in this volume).

There exist different variants of possibilistic logic where a logical formula is, in particular, associated with lower bounds of (weak) possibility measures. They can express different forms of ignorance by asserting that two opposite events are both at least somewhat possible). Other kinds of weights can be attached to logical formulas such as time slots where one is more or less certain that the formula is true, or subsets of sources or agents that are certain to various extents that the formula is true; see [Dubois and Prade, 2004, 2014] for references. For further developments on multiple agent possibilistic logic, see [Belhadi et al, 2013].

Generalized possibilistic logic Another type of extension allows for negations or disjunctions of basic possibilistic formulas (and not only conjunctions as in standard possibilistic logic). It then results into a two-tiered logic, named "*generalized possibilistic logic*" (GPL) [Dubois et al, 2017c], where connectives can be placed inside or outside basic possibilistic formulas. Its semantics is in terms of *subsets* of possibility distributions. Indeed, elementary formulas in the logic GPL encode lower or upper bounds on the necessity or the possibility of logical formulas. GPL is both a generalization of the minimal epistemic logic MEL [Banerjee and Dubois, 2014] (where weights are only 1 or 0), and of standard possibilistic logic, in full agreement with possibility theory. GPL has been axiomatized and inference in GPL has been shown sound and complete w.r.t. semantics in terms of subsets of possibility distributions.

GPL appears as a powerful unifying framework for various knowledge representation formalisms. Among others, logics of comparative certainty, and reasoning about explicit ignorance can be modeled in GPL. There also exists a close connection between GPL and various existing knowledge representation formalisms. It includes possibilistic logic with partially ordered formulas [Touazi et al, 2015], the logic of conditional assertions of Kraus et al [1990], three-valued logics [Ciucci and Dubois, 2013], and the 5-valued "equilibrium logic" of Pearce [2006] as well as answer set programming [Dubois et al, 2012] (see Chapter 4 in Volume 2). More specifically, the intended meaning of answer-set programs can be made more explicit through a translation in GPL (using a 3-level scale for the possibility distributions).

Lastly, in the same way as imprecise probabilities (see next chapter in this volume) are of interest, one may think of imprecise possibilities. In that respect, the following result is particularly worth noticing: any capacity (i.e., any monotonic increasing set function) on a finite domain can be characterized by a set of possibility mesures; then capacities offer a semantics to non regular modal logics (useful for the handling of paraconsistency) [Dubois et al, 2015b], and it may provide a unifying framework for multiple source information processing in the spirit of Belnap logic.

4.2.2 Inconsistency and Non Monotonic Reasoning

An important feature of possibilistic logic is its ability to deal with inconsistency. The inconsistency level $inc(B)$ of a possibilistic base B is defined as

$$inc(B) = \max\{\alpha \mid B \vdash (\bot, \alpha)\}.$$

No formula whose level is strictly greater than $inc(B)$ contributes to inconsistency. It can be shown that $1 - inc(B)$ is the height $h(\pi_B)$ of π_B, defined by $h(\pi_B) = \max_s \pi_B(s)$ (π_B being the possibility distribution induced by B). Moreover, $inc(B) = 0$ if and only if the set of logical formulas appearing in B, irrespective of the weights, is consistent in the classical sense. All the formulas in B whose level is smaller or equal to $inc(B)$ are ignored in the standard possibilistic inference mechanism; they are said to be "drowned". However, there exist other extensions of possibilistic inference that take into account formulas at the inconsistency level or below, especially those not involved in any inconsistent subset of formulas (called free formulas), see [Benferhat et al, 1999a] for a complete overview of these inferences.

The application of default rules having potential exceptions (for instance, "birds fly") to particular situations (e.g., "Tweety is a bird") about which information is incomplete, may lead to tentative conclusions (here, "Tweety flies") that become inconsistent with the new conclusions obtained when more information becomes available on such particular situations (e.g., "Tweety is a penguin"). The non monotonic nature of conditional qualitative possibility enables us to handle this problem. Indeed it allows $N(B \mid A) > 0$ and $N(\overline{B} \mid A \cap A') > 0$ to simultaneously hold, i.e., the arrival of the piece of information A' leads to reject a previously accepted proposition B in the context where we only knew A.

Indeed, a default rule "if A_i then generally B_j" can be represented by the possibilistic constraint $\Pi(B_j \cap A_i) > \Pi(\overline{B_j} \cap A_i)$ expressing that it is more possible to have B_i true than B_i false in the context where A_i is true. A base of default rules is then represented by a set of such constraints, which in turn determines a set of possibility measures that satisfy them. From such a rule base, two types of inference are natural in order to deduce new rules applicable to the situation where one *exactly* knows A (i.e., the rules of the form "if A then generally B", which will allow us to conclude B (tentatively) in this situation).

A first type of inference, which is cautious, requires that the inequality constraint $\Pi(A \cap B) \geq \Pi(A \cap \overline{B})$ associated with $B|A$ be satisfied by par *all* possibility measures that agree with the constraints (supposed to be consistent) associated with the set of default rules. A second, bolder, inference only considers the largest (the least specific) possibility distribution that is a solution of the latter constraints (it can be shown that this distribution is unique when it exists). It can be established that the first inference relation exactly corresponds the so-called preferential inference (system P [Kraus et al, 1990]) obeying basic postulates for non monotonic plausible inference (see Chapter 2 in this volume), while the second one is nothing but the "rational closure" inference of Lehmann and Magidor [1992]. These two types of inference can be justified also using other semantics such as conditional objects [Dubois and Prade, 1994], infinitesimal probabilities, systems Z and Z^+ [Pearl, 1990; Goldszmidt and Pearl, 1991], conditional modal logic [Boutilier, 1994], Halpern's plausibility measures [Halpern, 2001]; see [Benferhat et al, 1997] for an overview and references. There are also semantics in terms of big-stepped probabilities [Benferhat et al, 1999b], or conditional probabilities in De Finetti sense [Coletti and Scozzafava, 2002]. In this latter case the rule "if A then generally B" simply corresponds to a constraint $Prob(B|A) = 1$ where $Prob(B|A)$ still makes sense when $Prob(A) = 0$ (0 does not mean impossible here, but rather something as "negligible at first glance"), thanks to a prioritized handling of constraints induced by a partitioning of the set of interpretations [Biazzo et al, 2002]. The setting of possibilistic logic thus enables us to practically handle a form of default reasoning [Benferhat et al, 1998], as well as reasoning from qualitative uncertain information; il is even possible to combine both [Dupin de Saint-Cyr and Prade, 2008].

Belief revision theory [Gärdenfors, 1988] (see Chapter 14 in this volume), which is closely related to non monotonic reasoning, relies on the notion of epistemic entrenchment, used by the revision process for ordering the way pieces of information are called into question. It is interesting to note that an epistemic entrenchment relation is nothing but a qualitative necessity relation [Dubois and Prade, 1991] (whose unique counterpart on a totally ordered scale is a necessity measure [Dubois, 1986]). Moreover the possibilistic setting can make sense of the intuition that propositions in the belief base that are independent of the input information should remain after revision [Dubois et al, 1999]. Besides, updating and revision can be combined, just as in Kalman [1960] filtering, in the qualitative setting of possibilistic logic [Benferhat et al, 2000].

Let us also mention a model of causal *ascription* where an agent, in the presence of a sequence of events that took place, is supposed to assert causal relations between some of these events on the basis of his beliefs on the normal course of things [Bonnefon et al, 2008]. The normal course of things is represented by default rules (obeying system *P* postulates). In this approach, *causality* plays a role different from the one in the logics of action (see Chapter 15 in this volume) or in diagnosis (see Chapter 21 in this volume), where causality relations are supposed to be known. The possibilistic framework for causal ascription favors "abnormal" events as potential causes which may be adopted by the agent; a detailed comparison of this approach with the probabilistic modeling of causation can be found in [Bonnefon et al, 2012]. The prediction of the way people ascribe causality relations between reported events is not to be confused with actual causality judgements that get rid of spurious correlations by means of *interventions* in the sense of Pearl [2000] (such interventions can also be handled in the possibilistic setting [Benferhat and Smaoui, 2011]). The reader is referred to Chapter 9 in this volume for an overview of approaches to causality modeling.

4.2.3 Possibility Theory and Formal Concept Analysis

Formal concept analysis (FCA) is a knowledge representation formalism at the basis of a data mining methodology (see Chapters 12 and 13 of Volume 2). It provides a theoretical setting for learning hierarchies of concepts. Strong similarities between this representation framework and possibility theory have been pointed out in the last decade (and also to some extent with rough set theory [Pawlak and Skowron, 2007]). This is the reason for the presence of this – maybe unexpected – subsection in this chapter.

In FCA [Barbut and Montjardet, 1970; Ganter and Wille, 1999], one starts with a binary relation $\mathscr{R}$, called *formal context*, between a set of objects $\mathscr{O}$ and a set of properties $\mathscr{P}$; $x\mathscr{R}y$ means that x possesses property y. Given an object x and a property y, let $R(x) = \{y \in \mathscr{P} \mid x\mathscr{R}y\}$ be the set of properties possessed by object x and let $R(y) = \{x \in \mathscr{O} \mid x\mathscr{R}y\}$ be the set of objects having property y. In FCA correspondences are defined between the sets $2^{\mathscr{O}}$ and $2^{\mathscr{P}}$. These correspondences are Galois derivation operators. The Galois operator at the basis of FCA, here denoted by $(.)^{\Delta}$ (for a reason made clear in the following), enables us to describe the set of properties satisfied by *all* the objects in $X \subseteq \mathscr{O}$ as

$$X^{\Delta} = \{y \in \mathscr{P} \mid \forall x \in \mathscr{O}\ (x \in X \Rightarrow x\mathscr{R}y)\} = \{y \in \mathscr{P} \mid X \subseteq R(y)\} = \bigcap_{x \in X} R(x).$$

In a dual manner, the set of objects satisfying all the properties in Y is given by

$$Y^{\Delta} = \{x \in \mathscr{O} \mid \forall y \in \mathscr{P}\ (y \in Y \Rightarrow x\mathscr{R}y)\} = \{x \in \mathscr{O} \mid Y \subseteq R(x)\} = \bigcap_{y \in Y} R(y).$$

The pair of operators $((.)^{\Delta}, (.)^{\Delta})$ applied respectively to $2^{\mathscr{O}}$ and $2^{\mathscr{P}}$ constitutes a Galois connection that induces formal concepts. Namely, a *formal concept* is a pair (X,Y) such that

$$X^{\Delta} = Y \text{ and } Y^{\Delta} = X.$$

In other words, X is a maximal set of objects, and Y a maximal set of properties such that each object in X satisfies all the properties in Y. Then the set X (resp. Y) is called *extension* (resp. *intension*) of the concept. In an equivalent way, (X,Y) is a formal concept if and only if it is a maximal pair for the inclusion

$$X \times Y \subseteq \mathscr{R}.$$

The set of all formal concepts is naturally equipped with an order relation (denoted by $\preccurlyeq$) and defined by: $(X_1,Y_1) \preceq (X_2,Y_2)$ iff $X_1 \subseteq X_2$ (or $Y_2 \subseteq Y_1$). This set equipped with the order relation $\preccurlyeq$ forms a complete lattice. Then association rules between properties can be found by exploiting this lattice, see [Guigues and Duquenne, 1986; Pasquier et al, 1999].

Note that $X^{\Delta} = \bigcap_{x \in X} R(x)$ mirrors the definition of a guaranteed possibility measure $\Delta(F) = \min_{s \in F} \pi(s)$ (where π is a possibility distribution), changing L into 2^Y and π into a set-valued map ($R(x)$ is the set of properties satisfied by object x). On the basis of this parallel with *possibility theory*, other operators can be introduced [Dubois and Prade, 2012]. Namely, the possibility operator (denoted by $(.)^{\Pi}$) and its dual necessity operator (denoted by $(.)^{N}$), as well as the operator $(.)^{\nabla}$ dual to the operator $(.)^{\Delta}$ on which FCA is based. They are defined as follows:

- X^{Π} is the set of properties satisfied by at least one object in X:

$$X^{\Pi} = \{y \in \mathscr{P} \mid \exists x \in X, x\mathscr{R}y\} = \{y \in \mathscr{P} \mid X \cap R(y) \neq \emptyset\} = \bigcup_{x \in X} R(x);$$

- X^N is the set of properties that only the objects in X have:

$$X^N = \{y \in \mathscr{P} \mid \forall x \in \mathscr{O}\,(x\mathscr{R}y \Rightarrow x \in X)\} = \{y \in \mathscr{P} \mid R(y) \subseteq X\} = \bigcap_{x \notin X} \overline{R}(x),$$

where $\overline{R}(x)$ is the set of properties that x has not;
- X^{∇} is the set of properties that are not satisfied by at least one object outside X:

$$X^{\nabla} = \{y \in \mathscr{P} \mid \exists x \in \overline{X}, x\overline{\mathscr{R}}y\} = \{y \in \mathscr{P} | R(y) \cup X \neq \mathscr{O}\} = \bigcup_{x \notin X} \overline{R}(x).$$

The operators Y^{Π}, Y^N, Y^{∇} are obtained similarly. While the equalities $X^{\nabla} = Y$ and $Y^{\nabla} = X$ provide another characterization of usual formal concepts, it can be shown that pairs (X,Y) such that $X^N = Y$ and $Y^N = X$ (equivalently, $X^{\Pi} = Y$ and $Y^{\Pi} = X$) characterize independent sub-contexts (i.e., that have no object or property in common) inside the initial context [Dubois and Prade, 2012]. The pairs (X,Y) such that $X^N = Y$ and $Y^N = X$ are such that:

$$(X \times Y) \cup (\overline{X} \times \overline{Y}) \supseteq \mathscr{R}.$$

It can be checked that the four sets $X^{\Pi}, X^N, X^{\Delta}, X^{\nabla}$ are complementary pieces of information, all necessary for a complete analysis of the situation of X in the formal context $\mathscr{K} = (\mathscr{O}, \mathscr{P}, \mathscr{R})$. In practice, one supposes that both $R(x) \neq \emptyset$ and $R(x) \neq \mathscr{P}$, i.e., each object possesses at least one property in $\mathscr{P}$, but no object possesses all the properties in $\mathscr{P}$. Under this hypothesis of *bi-normalisation*, the following inclusion relation holds: $R^N(Y) \cup R^{\Delta}(Y) \subseteq R^{\Pi}(Y) \cap R^{\nabla}(Y)$, which is a counterpart of a relation that holds as well in possibility theory (provided that distributions π and $1 - \pi$ are both normalized).

Finally, let us also mention that there exists an extension of FCA to graded properties [Belohlavek, 2002], as well as an extension to formal contexts displaying incomplete or uncertain information [Burmeister and Holzer, 2005; Ait-Yakoub et al, 2017]. Another extension deals with the capability of associating objects no longer with simple properties, but with structured descriptions, possibly imprecise, or with logical descriptions, thanks to so-called *patron structures* [Ganter and Kuznetsov, 2001; Ferré and Ridoux, 2004]. They remain in agreement with the possibilistic paradigm [Assaghir et al, 2010].

4.3 Quantitative Possibility and Bridges to Probability

In the quantitative version of possibility theory, it is natural to relate possibility and probability measures. It can be done in several independent ways. In the following, we outline the three main ones: namely, a possibility distribution can be viewed as a likelihood function in non-Bayesian statistics, possibility (resp. necessity) degrees of events can be viewed either as upper (resp. lower) probability bounds, or as a suitable transformation of exponents of infinitesimal probabilities.

4.3.1 Possibility Distributions as Likelihood Functions

The idea of casting likelihood functions inside the framework of possibility theory was suggested by Smets [1982], but it has roots in considerations relating statistical inference and consonant belief functions (another name for necessity measures) in Shafer [1976]'s book; see also [Denœux, 2014] on this topic. The connection was formalized in [Dubois et al, 1997], and further studied in the coherence framework of De Finetti in [Coletti and Scozzafava, 2003]. Consider an estimation problem where the value of a parameter $\theta \in \Theta$ that governs a probability distribution $P(\cdot \mid \theta)$ on S is to be determined from data. Suppose the obtained data is described by the information item A. The function $\ell(\theta) = P(A \mid \theta), \theta \in \Theta$ is not a probability distribution, it is a *likelihood* function: a value θ is all the more plausible as $P(A \mid \theta)$ is greater, while this value can be ruled out if $P(A \mid \theta) = 0$ (in practice, less that a small relevance threshold). Such a function is often renormalized so that is maximal value is 1, since a likelihood function is defined up to a positive multiplicative constant. There are some good reasons why one may see $\ell(\theta)$ as a degree of possibility of θ, and let $\pi(\theta) = P(A \mid \theta)$ (up to renormalizing). First, it can be checked that, in the absence of prior probability on Θ, $\forall B \subseteq \Theta$, $P(A \mid B)$ is upper and lower bounded as follows:

$$\min_{\theta \in B} P(A \mid \theta) \leq P(A \mid B) \leq \max_{\theta \in B} P(A \mid \theta)$$

It suggests that we can apply the maxitivity axiom to get an optimistic estimate of $P(A \mid B)$ from $\{P(A \mid \theta), \theta \in B\}$. However, insofar as $\ell(b)$ is the likelihood of $\theta = b$, and we extend it to all subsets B of Θ, we should have that $\ell(B) \geq \ell(b)$, for all $b \in B$. Hence, in the absence of prior probability, we can identify $\ell(B)$ as a possibility measure with distribution $\pi(\theta) = P(A \mid \theta)$ [Coletti and Scozzafava, 2003]. Considering the lower bound of $P(A \mid B)$ would yield a guaranteed possibility measure.

However, note that under this view, possibility degrees are known in relative values, which means that not all basic notions of possibility theory apply (e.g., comparing the informativeness of π and π' using fuzzy set inclusion, by checking if $\pi \leq \pi'$ becomes questionable).

4.3.2 Possibility as Upper Probability

Alternatively, possibility degrees valued on $[0,1]$ viewed as an absolute scale can be exactly defined as upper probability bounds as Zadeh [1978] had the intuition from the start. The generation process can be described as follows: consider an increasing sequence of nested sets $E_1 \subset E_2 \subset, \ldots, \subset E_k$ and let $\alpha_1 \leq \alpha_2 \leq, \ldots, \leq \alpha_k \in [0,1]$, such that α_i is a lower bound on $P(E_i)$. This type of information is typically provided by an expert estimating a quantity v by means of set E_k with confidence α_k that E_k contains v. Consider the probability family $\mathscr{P} = \{P : P(E_i) \geq \alpha_i, \forall i = 1, \ldots, k\}$. It is easy to check [Dubois and Prade, 1992] that the function $P_*(A) = \inf_{P \in \mathscr{P}} P(A)$ is a necessity measure and the function $P^*(A) = \sup_{P \in \mathscr{P}} P(A)$ is a possibility measure induced by the possibility distribution:

$$\forall s \in S, \quad \pi(s) = \min_{i=1,\ldots,k} \max(E_i(s), 1 - \alpha_i). \tag{16}$$

where $E_i(s) = 1$ if $s \in E_i$ and 0 otherwise. See [de Cooman and Aeyels, 1999] for the extension of this result to infinite settings. Each pair (E_k, α_k), made of a set and its confidence level is encoded by the possibility distribution $\max(E_i(s), 1 - \alpha_i)$, where $1 - \alpha_i$ is an upperbound on the probability that $v \notin E_k$. Equation (16) just performs the conjunction of these local distributions. It is clear that π is a very concise encoding of the probability family $\mathscr{P}$. Conversely, the (convex set) of probability measures encoded by a possibility distribution π can be retrieved as

$$\mathscr{P}(\pi) = \{P, P(A) \leq \Pi(A), \forall A \text{ measurable}\} = \{P, P(A) \geq N(A), \forall A \text{ mesurable}\},$$

and it can be checked that $\Pi(A) = \sup_{P \in \mathscr{P}(\pi)} P(A)$. In the case where the sets E_i are not nested, the above formula (which is in agreement with possibilistic logic semantics of Section 4.2.1) only yields an approximation of the probability family $\mathscr{P}$; better approximations can be obtained by means of pairs of possibility distributions enclosing $\mathscr{P}$ [Destercke et al, 2008]. This view of possibility measures cast them in the landscape of imprecise probability theory studied in the next chapter of this volume.

Nested shortest dispersion intervals can be obtained from a given probability distribution (or density) p, letting $E_\alpha = \{s : p(s) \geq \alpha\}$, and $\alpha_\alpha = P(E_\alpha)$. The obtained possibility distribution, that covers p as tightly as possible, is called optimal probability-possibility transform of p [Dubois et al, 2004] and is instrumental for comparing probability distributions in terms of their peakedness or entropies (by comparing their possibility transforms in terms of relative specificity) [Dubois and Hüllermeier, 2007].

4.3.3 Possibility as Infinitesimal Probability

Ranking functions, originally called ordinal conditional functions (OCF), have been proposed by Spohn [1988, 2012] to represent the notion of belief in a setting that is basically equivalent to possibility theory, but for the direction and nature of its value scale. Each state of the world $s \in S$ is assigned a degree $\kappa(s)$ not in $[0,1]$, but in the set of non-negative integers $\mathbb{N}$, (sometimes even ordinals). The convention for ranking functions is opposite to the one in possibility theory, since the smaller $\kappa(s)$ the more possible s. It is more in agreement with a degree of potential surprise suggested by Shackle [1961]: $\kappa(s) = +\infty$ means that s is impossible, while $\kappa(s) = 0$ means that nothing opposes to s being the true state of the world. Set functions similar to possibility distribution are then built in the same style as Shackle [1961]:

$$\kappa(A) = \min_{s \in A} \kappa(s) \text{ and } \kappa(\emptyset) = +\infty.$$

More specifically, Spohn [1990] interprets $\kappa(A)$ as the integer exponent of an infinitesimal probability $P(A) = \varepsilon^{\kappa(A)}$, which is indeed in agreement with the union-minitivity property $\kappa(A \cup B) = \min(\kappa(A), \kappa(B))$ of ranking functons.

Conditioning is defined by Spohn [1988] as follows:

$$\kappa(s \mid B) = \begin{cases} \kappa(s) - \kappa(B) & \text{si } s \in B \\ +\infty & \text{sinon} \end{cases}$$

It is obvious that $\kappa(s \mid B)$ is the exponent of the infinitesimal conditional probability $P(s \mid B) = \varepsilon^{\kappa(s)} / \varepsilon^{\kappa(B)}$.

Casting ranking functions in possibility theory is easy, due to the following transformations [Dubois and Prade, 1991] :

$$\pi_\kappa(s) = 2^{-\kappa(s)}, \Pi_\kappa(A) = 2^{-\kappa(A)}.$$

As a consequence possibility distributions π_κ and functions Π_κ take values on a subset of rational numbers in $[0,1]$. Function Π_κ is indeed a possivility measure since

$$\Pi_\kappa(A \cup B) = 2^{-\min(\kappa(A), \kappa(B))} = \max(\Pi_\kappa(A), \Pi_\kappa(B)).$$

Moreover, for the conditional ranking function one obtains $\forall s$,

$$\pi_{\kappa(s|B)} = 2^{-\kappa(s)+\kappa(B)} = \frac{2^{-\kappa(s)}}{2^{-\kappa(B)}} = \frac{\pi_\kappa(s)}{\Pi_\kappa(B)},$$

which is the product-based conditioning of possibility theory. The converse (logarithmic) transformation of a possibility distribution into a ranking function is only possible if it maps real numbers to non-negative integers. More on the comparison between possibility theory and ranking functions can be found in [Dubois and Prade, 2016].

Note that this approach is often presented as qualitative while it is a numerical one. In some applications, or when modeling expert opinions, it may be more convenient to describe degrees of (dis)belief by means of integers rather than by real numbers in $[0,1]$. However it is easier to introduce intermediary grades with a continuous scale. The integer scale of ranking functions has been used recently by Kern-Isberner and Eichhorn [2014] to encode non-monotonic inferences and applied in [Eichhorn and Kern-Isberner, 2015] to belief networks.

5 The Cube of Opposition: A Structure Unifying Representation Frameworks

Many knowledge representation formalisms, although they look quite different at first glance and aim at serving diverse purposes, share a common structure where involutive negation plays a key role. This structure can be summarized under the form of a square or a cube of opposition. This in particular true for frameworks able to represent incomplete information. It can be observed that the properties of non empty intersection and of inclusion related by negation are at the basis of possibility theory, formal concept analysis, as well as rough set theory. It is still true for belief functions presented in the next chapter in this volume. This section first introduces the square and the cube of opposition, and indicates the formalisms to which it applies.

The traditional square of opposition [Parsons, 2008], which dates back to Aristotle's time, is built with universally and existentially quantified statements in the following way. Consider four statement of the form (**A**): "all *P*'s are *Q*'s", (**O**): "at least one *P* is not a *Q*", (**E**): "no *P* is a *Q*", and (**I**): "at least one *P* is a *Q*". They can be displayed on a square whose vertices are traditionally denoted by the letters **A**, **I** (affirmative half) and **E**, **O** (negative half), as pictured in Figure 1 (where $\overline{Q}$ stands for "not *Q*").

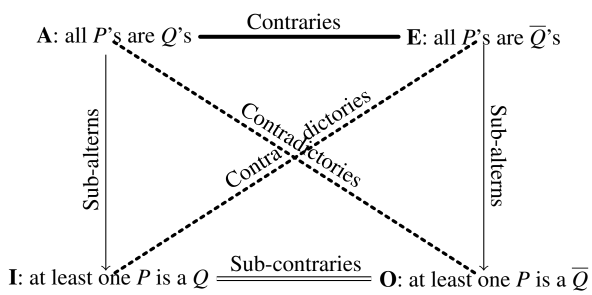

Fig. 1 Square of opposition

As can be checked, noticeable relations hold in the square provided that there a non empty set of *P*'s to avoid existential import problems:

1. **A** and **O** (resp. **E** and **I**) are the negation of each other;
2. **A** entails **I**, and **E** entails **O** (it is assumed that there is at least one *P*);
3. **A** and **E** cannot be true together;
4. **I** and **O** cannot be false together.

Another classical example of such a square is obtained with modal logic operators by taking **A** as $\Box p$, **E** as, $\Box\neg p$, **I** as $\Diamond p$, and **O** as $\Diamond\neg p$. This structure largely forgotten with the advent of modern logic after G. Boole, was rediscovered by Blanché [1966] and then by Béziau [2003] who both advocate its interest. In particular, Blanché noticed that adding two vertices **U** and **Y** defined respectively as the disjunction of **A** and **E**, and as the conjunction of **I** and **O**, leads to a hexagon that includes three squares of opposition in the above sense. Such a hexagon is obtained each time we start with three mutually exclusive statements, such as **A**, **E**, and **Y**, and it turns out that this structure is often encountered when representing relationships between concepts on the same domain (e.g., deontic notions such as permission, obligation, interdiction, etc.).

Switching to first order logic notations (e.g., **A** becomes $\forall x, P(x) \to Q(x)$), and negating the predicates, i.e., changing P into $\neg P$, and Q in $\neg Q$ leads to another similar square of opposition **aeoi**, where we also assume that the set of "not-P's" is non-empty. Altogether, we obtain eight statements that may be organized in what may be called a *cube of opposition* [Reichenbach, 1952]. The front facet and the back facet of the cube are traditional squares of opposition, and the two facets are related by entailments.

Such a structure can be extended to graded notions [Ciucci et al, 2016], using an involutive negation such as $1-(\cdot)$, and where the mutual exclusiveness of **A** and **E** translates into a sum of degrees less or equal to 1, while entailments are translated by inequalities between degrees (in agreement with residuated implications). An example of a graded cube is given by possibility theory. Indeed, assuming a normalized possibility distribution $\pi : S \to [0,1]$, and also assuming that $1-\pi$ is normalized (i.e., $\exists s \in S, \pi(s) = 0$), we obtain a cube of opposition on Fig. 2, linking $\Pi(A)$, $N(A)$, $\Delta(A)$, $\nabla(A)$, $\Pi(\overline{A})$, $N(\overline{A})$, $\Delta(\overline{A})$, and $\nabla(\overline{A})$. The front and back facets form two squares of opposition, while the side facets express a different property, namely inequalities such as $\min(\Pi(A), \nabla(A)) \geq \max(N(A), \Delta(A))$. Since these set functions rely on ideas of graded inclusion and degrees of non-empty intersections, the fact that they fit with a graded structure of cube of opposition should not be too surprizing.

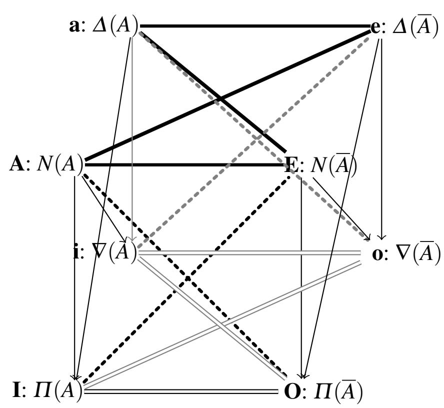

Fig. 2 Cube of opposition of possibility theory

In fact, the structure of cube of opposition is quite general. As noticed in [Ciucci et al, 2016], any binary relation R on a Cartesian product $X \times Y$ (one may have

$Y = X$) gives birth to a cube of opposition, when applied to a subset. Indeed, we assume $R \neq \emptyset$. Let $R(x) = \{y \in Y \mid (x,y) \in R\}$. $\overline{R}$ denotes the complementary relation ($(x,y) \in \overline{R}$ iff $(x,y) \notin R$), and R^t the transposed relation ($(y,x) \in R^t$ if and only if $(x,y) \in R$); let $R(y) = \{x \in X \mid (x,y) \in R\} = R^t(y)$. Moreover, it is assumed that $\forall x,\ R(x) \neq \emptyset$, which means that the relation R is *serial*, namely $\forall x, \exists y$ such that $(x,y) \in R$. Similarly, R^t is also supposed to be serial, i.e., $\forall y,\ R(y) \neq \emptyset$, as well as $\overline{R}$ and its transpose, i.e. $\forall x,\ R(x) \neq Y$ and $\forall y,\ R(y) \neq X$.

Let T be a subset of Y and $\overline{T}$ its complement. We assume $T \neq \emptyset$ and $T \neq Y$. The composition is defined in the usual way $R(T) = \{x \in X \mid \exists t \in T, (x,t) \in R\}$. From the relation R and the subset T, one can define the four following subsets of X (and their complements):

$$R(T) = \{x \in X \mid T \cap R(x) \neq \emptyset\} \tag{17}$$

$$\overline{R(\overline{T})} = \{x \in X \mid R(x) \subseteq T\} \tag{18}$$

$$\overline{\overline{R}(T)} = \{x \in X \mid T \subseteq R(x)\} \tag{19}$$

$$\overline{R}(\overline{T}) = \{x \in X \mid T \cup R(x) \neq X\}. \tag{20}$$

These four subsets and their complements can be nicely organized into a cube of opposition (Fig.3). Some of the required conditions for the cube hold thanks to seriality (which plays the role of normalization in possibility theory).

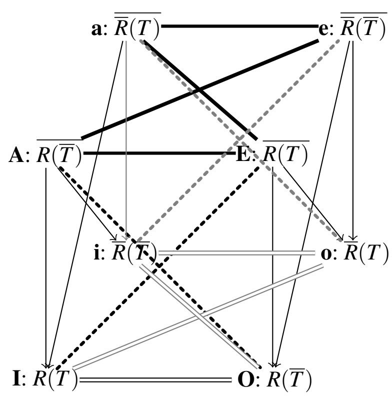

Fig. 3 Cube induced by a relation R and a subset T

The front facet of the cube fits well with the modal logic reading of the square where R is viewed as an accessibility relation, and T as the set of models of a proposition p. Indeed, $\Box p$ (resp. $\Diamond p$) is true in world x means that p is true at *every* (resp. at *some*) possible world accessible from x; this corresponds to $\overline{R(\overline{T})}$ (resp. $R(T)$) which is the set of worlds where $\Box p$ (resp. $\Diamond p$) is true.

Other than the semantics of modal logics, there are a number of AI formalisms that exploit a relation and to which the cube of opposition of Fig. 3 applies: formal concept analysis, as seen in Section 4.2.3, rough sets induced by an equivalence relation (see Section 2.6), or abstract argumentation based on an attack relation between arguments. Graded squares or cubes also apply to belief functions and to upper and lower probabilities [Pfeifer and Sanfilippo, 2017] presented in the next chapter in this volume, [Dubois et al, 2015a], as well as to aggregation functions such as Sugeno integrals used in multiple criteria aggregation and qualitative decision theory, or yet Choquet integrals [Dubois et al, 2017b], both presented in Chapter 16 in this volume.

This common structure is deeply related to the interplay of three negations as revealed by the relational cube. In contrast the square and the cube collapse to a segment in case of probabilities since they are autodual.

The cube of opposition lays bare common features underlying many knowledge representation formalisms. It exhibits fruitful parallelisms between them, which may even lead to highlight some missing components present in one formalism and currently absent from another.

6 Conclusion

In this chapter, we have tried to show that while probability theory properly captures uncertainty due to the randomness of precisely observed phenomena, the representation of uncertainty due to incomplete information requires a different setting having roots in classical and modal logics, where incompleteness is a usual feature. The corresponding uncertainty framework is possibility theory, which allows for a qualitative representation of uncertainty as well as a quantitative one. It has been shown that numerical possibility theory is appropriate provided that the available information items, although imprecise, are consonant, i.e., do not contradict each other. The handling of imprecise and possibly conflicting information items require joint extensions of probability and quantitative possibility theory studied in the next chapter.

References

Abadi M, Halpern JY (1994) Decidability and expressiveness for first-order logics of probability. Inf Comput 112(1):1–36

Ait-Yakoub Z, Djouadi Y, Dubois D, Prade H (2017) Asymmetric composition of possibilistic operators in formal concept analysis: Application to the extraction of attribute implications from incomplete contexts. Int J of Intelligent Systems 32(12):1285–1311

Assaghir Z, Kaytoue M, Prade H (2010) A possibility theory-oriented discussion of conceptual pattern structures. In: Deshpande A, Hunter A (eds) Proc. Int. Conf. on

Scalable Uncertainty Management (SUM'10), Toulouse, Sept. 27-29, Springer, no. 6379 in LNCS, pp 70–83
Bacchus F (1991) Representing and Reasoning With Probabilistic Knowledge: A Logical Approach to Probabilities. MIT Press, Cambridge, Ma
Banerjee M, Dubois D (2014) A simple logic for reasoning about incomplete knowledge. Int J Approx Reasoning 55:639–653
Barbut M, Montjardet B (1970) Ordre et Classification: Algèbre et Combinatoire. Hachette
Belhadi A, Dubois D, Khellaf-Haned F, Prade H (2013) Multiple agent possibilistic logic. J of Applied Non-Classical Logics 23(4):299–320
Belnap ND (1977a) How a computer should think. In: Contemporary Aspects of Philosophy, Oriel Press, Boston, pp 30–56
Belnap ND (1977b) A useful four-valued logic. In: Epstein G (ed) Modern Uses of Multiple-valued Logic, Reidel, pp 8–37
Belohlavek R (2002) Fuzzy Relational Systems. Foundations and Principles. Kluwer
Ben Amor N, Benferhat S, Dubois D, Mellouli K, Prade H (2002) A theoretical framework for possibilistic independence in a weakly ordered setting. Int J of Uncertainty, Fuzziness and Knowledge-Based Systems 10(2):117–155
Ben Amor N, Dubois D, Gouider H, Prade H (2017) Possibilistic preference networks. Information Sciences to appear
Benferhat S, Kaci S (2003) Logical representation and fusion of prioritized information based on guaranteed possibility measures: Application to the distance-based merging of classical bases. Artificial Intelligence 148:291–333
Benferhat S, Smaoui S (2011) Inferring interventions in product-based possibilistic causal networks. Fuzzy Sets and Systems 169(1):26–50
Benferhat S, Dubois D, Prade H (1997) Nonmonotonic reasoning, conditional objects and possibility theory. Artificial Intelligence 92:259–276
Benferhat S, Dubois D, Prade H (1998) Practical handling of exception-tainted rules and independence information in possibilistic logic. Applied Intelligence 9:101–127
Benferhat S, Dubois D, Prade H (1999a) An overview of inconsistency-tolerant inferences in prioritized knowledge bases. In: Dubois D, Prade H, Klement E (eds) Fuzzy Sets, Logic and Reasoning about Knowledge, Applied Logic Series, vol 15, Kluwer, Dordrecht, pp 395–417
Benferhat S, Dubois D, Prade H (1999b) Possibilistic and standard probabilistic semantics of conditional knowledge bases. J of Logic and Computation 9:873–895
Benferhat S, Dubois D, Prade H (2000) Kalman-like filtering in a possibilistic setting. In: Horn W (ed) Proc.14th Europ.Conf. on Artificial Intelligence (ECAI'00), Berlin, Aug. 20-25, pp 8–12
Benferhat S, Dubois D, Garcia L, Prade H (2002) On the transformation between possibilistic logic bases and possibilistic causal networks. Int J of Approx Reasoning 29:135–173

Benferhat S, Dubois D, Kaci S, Prade H (2008) Modeling positive and negative information in possibility theory. Inter J of Intelligent Systems 23:1094–1118
Béziau JY (2003) New light on the square of oppositions and its nameless corner. Logical Investigations 10:218–233
Biazzo V, Gilio A, Lukasiewicz T, Sanfilippo G (2002) Probabilistic logic under coherence, model-theoretic probabilistic logic, and default reasoning in system P. J of Applied Non-Classical Logics 12(2):189–213
Blanché R (1966) Structures Intellectuelles. Essai sur l'Organisation Systématique des Concepts. Vrin, Paris
Bolt JH, van der Gaag LC, Renooij S (2005) Introducing situational signs in qualitative probabilistic networks. Int J Approx Reasoning 38:333–354
Bonnefon JF, Da Silva Neves R, Dubois D, Prade H (2008) Predicting causality ascriptions from background knowledge: model and experimental validation. Int J Approx Reasoning 48:752–765
Bonnefon JF, Da Silva Neves R, Dubois D, Prade H (2012) Qualitative and quantitative conditions for the transitivity of perceived causation - theoretical and experimental results. Ann of Math and Artif Intell 64:311–333
Boutilier C (1994) Modal logics for qualitative possibility theory. Int J Approx Reasoning 10(2):173–201
Buchanan BG, Shortliffe EH (eds.) (1984) Rule-Based Expert Systems. Addison-Wesley, Reading, Mass.
Burmeister P, Holzer R (2005) Treating incomplete knowledge in formal concepts analysis. In: Ganter B (ed) Formal Concept Analysis : Foundations and Applications, LNCS, vol 3626, Springer, pp 114–126
Cayrac D, Dubois D, Prade H (1996) Handling uncertainty with possibility theory and fuzzy sets in a satellite fault diagnosis application. IEEE Trans on Fuzzy Systems 4(3):251–269
Cayrol M, Farreny H, Prade H (1982) Fuzzy pattern matching. Kybernetes 11(2):103–116
Cheeseman P (1988) An inquiry into computer understanding. Computational Intelligence 4:58–66, with comments by R. Aleliunas, A. Bundy, N. C. Dalkey, A. P. Dempster, D. Dubois and H. Prade, M. L. Ginsberg, R. Greiner, P. J. Hayes, D. Israel, L. Kanal and D. Perlis, H. Kyburg, D. McDermott, D. L. McLeish, C. G. Morgan, E. Neufeld and D. Poole, J. Pearl, L. Rendell, E. H. Ruspini, L.K. Schubert, G. Shafer, D. J. Spiegelhalter, R. R. Yager, L. A. Zadeh (67–128), and a reply by P. Cheeseman (129–142)
Choquet G (1953) Theory of capacities. Annales de l'Institut Fourier 5:131–295
Ciucci D, Dubois D (2013) A modal theorem-preserving translation of a class of three-valued logics of incomplete information. J of Applied Non-Classical Logics 23(4):321–352
Ciucci D, Dubois D, Prade H (2016) Structures of opposition induced by relations - the Boolean and the gradual cases. Ann Math Artif Intell 76(3-4):351–373
Coletti G, Scozzafava R (2002) Probabilistic Logic in a Coherent Setting. Kluwer Acad. Publ.

Coletti G, Scozzafava R (2003) Coherent conditional probability as a measure of uncertainty of the relevant conditioning events. In: Nielsen TD, Zhang NL (eds) Proc. 7th Europ. Conf. Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'03), Springer, LNCS, vol 2711, pp 407–418

Coletti G, Vantaggi B (2006) Possibility theory: Conditional independence. Fuzzy Sets and Systems 157(11):1491–1513

Coletti G, Vantaggi B (2009) T-conditional possibilities: Coherence and inference. Fuzzy Sets and Systems 160(3):306–324

Cox RT (1946) Probability, frequency, and reasonable expectation. Am Jour Phys 14:1–13

de Cooman G (1997) Possibility theory. Part I: Measure- and integral-theoretic ground- work; Part II: Conditional possibility; Part III: Possibilistic independence. Int J of General Syst 25:291–371

de Cooman G, Aeyels D (1999) Supremum preserving upper probabilities. Information Sciences 118(1-4):173–212

De Finetti B (1936) La logique de la probabilité. In: Congrès International de Philosophie Scientifique, Hermann et Cie, Paris, pp 1–9

De Finetti B (1974) Theory of Probability. Wiley and Sons, New York

Denœux T (2014) Likelihood-based belief function: justification and some extensions to low-quality data. Int J Approx Reasoning 55(7):1535–1547

Destercke S, Dubois D, Chojnacki E (2008) Unifying practical uncertainty representations: I. Generalized p-boxes. II. Clouds. Int J of Approx Reasoning 49:649–663, 664–677

Domotor Z (1985) Probability kinematics - Conditional and entropy principles. Synthese 63:74–115

Dubois D (1986) Belief structures, possibility theory and decomposable confidence measures on finite sets. Computers and Artificial Intelligence 5(5):403–416

Dubois D (2012) Reasoning about ignorance and contradiction: many-valued logics versus epistemic logic. Soft Comput 16(11):1817–1831

Dubois D, Hüllermeier E (2007) Comparing probability measures using possibility theory: A notion of relative peakedness. Int J of Approx Reasoning 45:364–385

Dubois D, Prade H (1982) A class of fuzzy measures based on triangular norms. a general framework for the combination of uncertain information. Int J of General Syst 8(1):43–61

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

Dubois D, Prade H (1989) Handling uncertainty in expert systems: pitfalls, difficulties, remedies. In: Hollnagel E (ed) The Reliability of Expert Systems, Ellis Horwood, Chichester, U.K., pp 64–118

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

Dubois D, Prade H (1992) Putting rough sets and fuzzy sets together. In: Slowinski R (ed) Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Acad. Publ., pp 203–232

Dubois D, Prade H (1994) Conditional objects as nonmonotonic consequence relationships. IEEE Trans on Systems, Man and Cybernetics 24(12):1724–1740

Dubois D, Prade H (1996) What are fuzzy rules and how to use them. Fuzzy Sets and Systems 84:169–185

Dubois D, Prade H (1998) Possibility theory: Qualitative and quantitative aspects. In: Gabbay DM, Smets P (eds) Quantified Representation of Uncertainty and Imprecision, Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol 1, Kluwer Acad. Publ., pp 169–226

Dubois D, Prade H (2001) Possibility theory, probability theory and multiple-valued logics : A clarification. Ann of Math and Artif Intell 32:35–66

Dubois D, Prade H (2004) Possibilistic logic: a retrospective and prospective view. Fuzzy Sets and Systems 144:3–23

Dubois D, Prade H (2009) Formal representations of uncertainty. In: Bouyssou D, Dubois D, Pirlot M, Prade H (eds) Decision-making - Concepts and Methods, Wiley, chap 3, pp 85–156

Dubois D, Prade H (2012) Possibility theory and formal concept analysis: Characterizing independent sub-contexts. Fuzzy Sets and Systems 196:4–16

Dubois D, Prade H (2014) Possibilistic logic – An overview. In: Siekmann JH (ed) Computational Logic, Handbook of the History of Logic, vol 9, Elsevier, pp 283–342

Dubois D, Prade H (2016) Qualitative and semi-quantitative modeling of uncertain knowledge - A discussion. In: Beierle C, Brewka G, Thimm M (eds) Computational Models of Rationality, College Publications, pp 280–296

Dubois D, Prade H (eds.) (2008) Bipolar Representations of Information and Preference. Part 1A & Part 1B: Cognition and Decision; Part 2: Reasoning and Learning. Special issue, Int. J. of Intelligent Systems, 23 (8,9,10), Wiley

Dubois D, Godo L, de Mántaras RL, Prade H (1993) Qualitative reasoning with imprecise probabilities. J Intell Inf Syst 2(4):319–363

Dubois D, Lang J, Prade H (1994) Possibilistic logic. In: Gabbay D, Hogger C, Robinson J, Nute D (eds) Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3, Oxford Univ. Press, pp 439–513

Dubois D, Moral S, Prade H (1997) A semantics for possibility theory based on likelihoods. J Math Anal Appl 205:359–380

Dubois D, Fariñas del Cerro L, Herzig A, Prade H (1999) A roadmap of qualitative independence. In: Dubois D, Prade H, Klement E (eds) Fuzzy Sets, Logics and Reasoning about Knowledge, Applied Logic series, vol 15, Kluwer Acad. Publ., Dordrecht, pp 325–350

Dubois D, Hajek P, Prade H (2000a) Knowledge-driven versus data-driven logics. J of Logic, Language, and Information 9:65–89

Dubois D, Pap E, Prade H (2000b) Hybrid probabilistic-possibilistic mixtures and utility functions. In: Fodor J, De Baets B, Perny P (eds) Preferences and Decisions under Incomplete Knowledge, Physica-Verlag, pp 51–73

Dubois D, Grabisch M, de Mouzon O, Prade H (2001) A possibilistic framework for single-fault causal diagnosis under uncertainty. Int J of General Syst 30(2):167–192

Dubois D, Foulloy L, Mauris G, Prade H (2004) Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. Reliable Computing 10:273–297

Dubois D, Esteva F, Godo L, Prade H (2007) Fuzzy-set based logics an history-oriented presentation of their main developments. In: Gabbay DM, Woods J (eds) Handbook of the History of Logic, vol 8, Elsevier, pp 325–449

Dubois D, Fodor J, Prade H (2010) Conditional measures: An alternative to Cox functional equation. In: Cintula P, Klement EP, Stout LN (eds) Proc. 31st Linz Seminar on Fuzzy Set Theory, Linz, Austria, pp 43–46

Dubois D, Prade H, Schockaert S (2012) Stable models in generalized possibilistic logic. In: Brewka G, Eiter T, McIlraith SA (eds) Proc.13th Int. Conf. Principles of Knowledge Representation and Reasoning (KR'12), Morgan Kaufmann, pp 519–529

Dubois D, Prade H, Rico A (2015a) The cube of opposition: A structure underlying many knowledge representation formalisms. In: Yang Q, Wooldridge M (eds) Proc. 24th Int. Joint Conf. on Artificial Intelligence, IJCAI'15, AAAI Press, pp 2933–2939

Dubois D, Prade H, Rico A (2015b) Representing qualitative capacities as families of possibility measures. Int J Approx Reasoning 58:3–24

Dubois D, Prade H, Rico A (2017a) Graded cubes of opposition and possibility theory with fuzzy events. Int J Approx Reasoning 84:168–185

Dubois D, Prade H, Rico A (2017b) Organizing families of aggregation operators into a cube of opposition. In: Kacprzyk J, Filev D, Beliakov G (eds) Granular, Soft and Fuzzy Approaches for Intelligent Systems, Springer Verlag, pp 27–45

Dubois D, Prade H, Schockaert S (2017c) Generalized possibilistic logic: Foundations and applications to qualitative reasoning about uncertainty. Artificial Intelligence 252:139–174

Dupin de Saint-Cyr F, Prade H (2008) Handling uncertainty and defeasibility in a possibilistic logic setting. Int J Approx Reasoning 49:67–82

Eichhorn C, Kern-Isberner G (2015) Using inductive reasoning for completing ocf-networks. J Applied Logic 13(4):605–627

Ferré S, Ridoux O (2004) Introduction to logical information systems. Information Process & Manag 40(3):383–419

Fine T (1983) Theories of Probability. Academic Press, New York

Gaifman H, Snir M (1982) Probabilities over rich languages, testing and randomness. J Symbolic Logic 47(3):495–548

Ganter B, Kuznetsov SO (2001) Pattern structures and their projections. In: Delugach HS, Stumme G (eds) Proc. 9th Int. Conf. on Conceptual Structures (ICCS'01), Springer, LNCS, vol 2120, pp 129–142

Ganter B, Wille R (1999) Formal Concept Analysis. Mathematical Foundations. Springer-Verlag

Gärdenfors P (1988) Knowledge in Flux. Modeling the Dynamics of Epistemic States. MIT Press (2nd ed., College Publications, 2008)

Ginsberg ML (1990) Bilattices and modal operators. J of Logic and Computation 1:1–41

Goldszmidt M, Pearl J (1991) System Z^+: A formalism for reasoning with variable-strength defaults. In: Proc. 9th National Conf. on Artificial Intelligence (AAAI'91), vol 1, pp 339–404

Greco S, Inuiguchi M, Slowinski R (2006) Fuzzy rough sets and multiple-premise gradual decision rules. Int J Approx Reasoning 41(2):179–211

Grzymala-Busse JW (1988) Knowledge acquisition under uncertainty - a rough set approach. J of Intell and Robotic Syst 1(1):3–16

Guigues JL, Duquenne V (1986) Familles minimales d'implications informatives résultant d'un tableau de données binaires. Mathématiques et Sciences Humaines 95:5–18

Hájek P (1998) The Metamathematics of Fuzzy Logics. Kluwer Academic

Halpern JY (1990) An analysis of first-order logics of probability. Artificial Intelligence 46:311–350

Halpern JY (1999a) A counterexample to theorems of Cox and Fine. J Artif Intell Res (JAIR) 10:67–85

Halpern JY (1999b) Technical addendum, Cox's theorem revisited. J Artif Intell Res (JAIR) 11:429–435

Halpern JY (2001) Plausibility measures: A general approach for representing uncertainty. In: Nebel B (ed) Proc. 17th Int. Joint Conf. on Artificial Intelligence (IJCAI'01), Morgan Kaufmann, pp 1474–1483

Halpern JY (2003) Reasoning About Uncertainty. MIT Press, Cambridge, Ma

Halpern JY, Pucella R (2002) A logic for reasoning about upper probabilities. J Artif Intell Res (JAIR) 17:57–81

Halpern JY, Pucella R (2006) A logic for reasoning about evidence. J Artif Intell Res (JAIR) 26:1–34

Hempel CG (1945) Studies in the logic of confirmation, I and II. Mind pp 1–26, 97–121

Higashi M, Klir GJ (1982) Measures of uncertainty and information based on possibility distributions. Int J General Syst 8:43–58

Hong T, Tseng L, Wang S (2002) Learning rules from incomplete training examples by rough sets. Expert Syst Appl 22(4):285–293

Horvitz E, Heckerman D, Langlotz C (1986) A framework for comparing alternative formalisms for plausible reasoning. In: Kehler T (ed) Proc. 5th Nat. Conf. on Artificial Intelligence. Vol. 1, Morgan Kaufmann, pp 210–214

Jaeger M (2001) Automatic derivation of probabilistic inference rules. Int J Approx Reasoning 28:1–22

Jaeger M (2006) Probabilistic role models and the guarded fragment. Int J of Uncertainty, Fuzziness and Knowledge-Based Systems 14(1):43–60

Jaynes ET (1979) Where do we stand on maximum entropy. In: Levine I, Tribus M (eds) The Maximum Entropy Formalism, MIT Press, pp 15–118

Jaynes ET (2003) Probability Theory: The Logic of Science. Cambridge Univ. Press, preprint version, 1996

Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Engin, Trans of ASME, series D 82:35–45

Kern-Isberner G, Eichhorn C (2014) Structural inference from conditional knowledge bases. Studia Logica 102(4):751–769

Kleene SC (1952) Introduction to Metamathematics. North Holland

Klement E, Mesiar R, Pap E (2000) Triangular norms. Kluwer Academic Pub., Boston

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

Kyburg, Jr HE, Teng CM (2012) The logic of risky knowledge, reprised. Int J Approx Reasoning 53:274–285

Lehmann DJ, Magidor M (1992) What does a conditional knowledge base entail? Artificial Intelligence 55:1–60

Lindley DV (1982) Scoring rules and the inevitability of probability. Int Statistics Rev 50:1–26

Liu W (2001) Propositional, Probabilistic and Evidential Reasoning: Integrating Numerical and Symbolic Approaches. Physica Verlag, Springer

Lucas P, van der Gaag L (1991) Principles of Expert Systems. Addison-Wesley

Marchioni E, Godo L (2004) A logic for reasoning about coherent conditional probability: A modal fuzzy logic approach. In: Alferes JJ, Leite JA (eds) Proc. 9th Europ. Conf. on Logics in Artificial Intelligence (JELIA'04), Springer, LNCS, vol 3229, pp 213–225

Martin T (2006) Logique du probable de Jacques Bernoulli à J.-H. Lambert. Journ@l Electronique d'Histoire des Probabilités et de la Statistique 2(1b):http://www.jehps.net/Novembre2006/Martin3.pdf

Milch B, Russell SJ (2007) First-order probabilistic languages: Into the unknown. In: Muggleton S, Otero RP, Tamaddoni-Nezhad A (eds) Revised Selected Papers of the16th Int. Conf. on Inductive Logic Programming (ILP'06), Springer, LNCS, vol 4455, pp 10–24

Mongin P (1994) Some connections between epistemic logic and the theory of nonadditive probability. In: Humphreys P (ed) Patrick Suppes: Scientific Philosopher. Vol.1: Probability and Probabilistic Causality, Springer, Synthese Library Vol. 234, pp 135–171

von Neumann J, Morgenstern O (1944) Theory Games and Economic Behavior. Princeton University Press, Princeton, NJ

Nilsson NJ (1993) Probabilistic logic revisited. Artificial Intelligence 59:39–42

Paris J (1994) The Uncertain Reasoner' Companion. Cambridge University Press

Parsons S (2001) Qualitative Approaches for Reasoning Under Uncertainty. MIT Press

Parsons T (2008) The traditional square of opposition. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy, Stanford University

Pasquier N, Bastide Y, Taouil R, Lakhal L (1999) Efficient mining of association rules using closed itemset lattices. Inf Syst 24:25–46

Pawlak Z (1991) Rough Sets. Theoretical Aspects of. Reasoning about Data. Kluwer Acad. Publ., Dordrecht

Pawlak Z, Skowron A (2007) 1. Rudiments of rough sets. 2. Rough sets: Some extensions. 3. Rough sets and Boolean reasoning. Information Sciences 177(1):3–73

Pearce D (2006) Equilibrium logic. Ann of Math and Artif Intell 47:3–41

Pearl J (1988) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publ.

Pearl J (1990) System Z: A natural ordering of defaults with tractable applications for default reasoning. In: Proc. of Theoretical Aspects of Reasoning about Knowledge (TARK90), pp 121–135

Pearl J (2000) Causality. Models, Reasoning and Inference. Cambridge University Press

Pfeifer N, Sanfilippo G (2017) Probabilistic squares and hexagons of opposition under coherence. Int J Approx Reasoning 88:282–294

Reichenbach H (1952) The syllogism revised. Philosophy of Science 19(1):1–16

Renooij S, van der Gaag L (1999) Enhancing QPNs for trade-off resolution. In: Laskey KB, Prade H (eds) Proc. 15th Conf. on Uncertainty in Artificial Intelligence (UAI '99), Morgan Kaufmann, pp 559–566

Renooij S, van der Gaag LC (2008) Enhanced qualitative probabilistic networks for resolving trade-offs. Artif Intell 172:1470–1494

Ruspini EH (1970) Numerical methods for fuzzy clustering. Information Sciences 2:319–350

Schweizer B, Sklar A (1963) Associative functions and abstract semi-groups. Publ Math Debrecen 10:69–180

Shackle GLS (1961) Decision, Order and Time in Human Affairs. (2nd edition), Cambridge University Press, UK

Shafer G (1976) A Mathematical Theory of Evidence. Princeton Univ. Press

Shafer G (1978) Non-additive probabilities in the work of Bernoulli and Lambert. Arch for History of Exact Sciences 19 (4):309–370

Shortliffe EH (1976) Computer-based Medical Consultations MYCIN. Elsevier

Smets P (1982) Possibilistic inference from statistical data. In: Proc. 2nd World Conf. on Mathematics at the Service of Man, Las Palmas, pp 611–613

Snow P (1999) Diverse confidence levels in a probabilistic semantics for conditional logics. Artificial Intelligence 113:269–279

Spohn W (1988) Ordinal conditional functions: a dynamic theory of epistemic states. In: Harper WL, Skyrms B (eds) Causation in Decision, Belief Change, and Statistics, vol 2, Kluwer, pp 105–134

Spohn W (1990) A general, nonprobabilistic theory of inductive reasoning. In: Shachter RD, Levitt TS, Kanal LN, Lemmer JF (eds) Uncertainty in Artificial Intelligence 4, North Holland, Amsterdam, pp 149–158

Spohn W (2012) The Laws of Belief: Ranking Theory and Its Philosophical Applications. Oxford Univ. Press

Sugeno M (1977) Fuzzy measures and fuzzy integrals - A survey. In: Gupta MM, Saridis GN, Gaines BR (eds) Fuzzy Automata and Decision Processes, North Holland, Amsterdam, pp 89–102

Touazi F, Cayrol C, Dubois D (2015) Possibilistic reasoning with partially ordered beliefs. J of Applied Logic 13(4, Part 3):770 – 798

Zadeh LA (1965) Fuzzy sets. Information Control 8:338–353

Zadeh LA (1968) Probability measures of fuzzy events. J Math Anal Appl 23(2):421–427

Zadeh LA (1975) The concept of a linguistic variable and its application to approximate reasoning. Information Sciences 8:199–249

Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems 1:3–28