



# Representing belief: beyond probability and logic

Didier Dubois, Henri Prade

## ► To cite this version:

Didier Dubois, Henri Prade. Representing belief: beyond probability and logic. The Handbook of Rationality, Part II: Theoretical rationality (Section 4: Probabilistic Reasoning), <https://mitpress.mit.edu/about>, 2021, 978-0262045070. hal-03545993

**HAL Id: hal-03545993**

**<https://ut3-toulouseinp.hal.science/hal-03545993>**

Submitted on 27 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Representing belief: beyond probability and logic

---

Didier Dubois and Henri Prade

## Abstract

This chapter surveys recent approaches to the representation of belief. There is a clash between the notion of degree of belief in the subjective probability tradition and the idea of certainty as accepted belief often couched in the language of logic, and especially modal logic. The attempt to consider degrees of certainty finds its origin in the works of Francis Bacon often opposed to the one of Pascal. However, by reviewing some more recent trends in the representation of uncertainty, such as possibility theory, ranking theory, evidence theory and imprecise probability, one may argue that these novel approaches try to bridge the gap between the two traditions, even if dropping some favourite properties on the way, such as the additivity of degrees of belief and the adjunction law for accepted beliefs.

## 1. Introduction

There are two traditions for representing belief, one based on probability, one based on modal logic. The probability tradition goes back to Pascal and other scientists of his time, who tried to measure our confidence in events (today we also speak of propositions in the logical sense). In those times, a distinction was made between so-called chances, and probabilities (Shafer, 1978). The notion of chance was dedicated to repeatable events and measured via a combinatorial count of the frequencies of occurrence assuming elementary events had equal chances to occur. The notion of probability was understood as a subjective notion, typically the confidence to be granted to a testimony. The latter notion is still prevalent in important texts from the XVIIIth century such as D'Alembert and Diderot's encyclopaedia<sup>1</sup>. The notion of subjective probability was to be caught up again in the XXth century by scholars like Ramsey (1926) and de Finetti (1937). They tried to justify why degrees of belief should be additive. However, additivity can only be obtained by introducing severe restrictions on the model of graded belief, namely, it essentially prevents to capture the notion of ignorance. The reader is referred to the rather recent compilation by Huber and Schmidt-Petri (2009) for a survey of approaches to degrees of belief.

In contrast, the modal logic tradition was developed in the XXth century, even if modal logic per se goes back to middle-ages philosophy, and even to ancient Greek philosophy. Hintikka (1962) proposed a so-called epistemic logic that tried to define a rational logical framework for the idea of knowledge, and its articulation with an all-or-nothing notion of belief. Namely knowledge is essentially true belief, while a simply believed proposition is not supposed to be actually true. In this setting, modeling ignorance is obvious as you just express that you believe neither a proposition nor its negation. Besides, one essential feature of this approach is that when you know that each of two propositions is true, you also know that their

---

<sup>1</sup> See the entry "Probabilité", accredited to Benjamin de Langes de Lubières (1714-1790), Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers. D. Diderot and J. Le Rond d'Alembert. <http://enccre.academie-sciences.fr/encyclopedia/page/v13-p403>

conjunction is true, which is sometimes called the *adjunction axiom* (Kyburg, 1997). This axiom is also supposed to hold for beliefs, with the understanding that modal beliefs represents *accepted beliefs*, that is, propositions taken for granted to the point of reasoning with them as if they were true.

There is a clash of intuitions between the two views, because for them to match, one needs to explain the connection between graded belief and belief *simpliciter*. One way is to add a threshold to graded beliefs and retain propositions above this threshold, but it is then almost impossible to preserve the adjunction axiom, if we model graded beliefs by probabilities. This is the point made by the Lottery Paradox (Kyburg, 1961), discussed later in this paper.

The aim of this chapter is to provide an overview of belief representations that relate and go beyond subjective probability and epistemic logic, in the sense that some of them are graded versions of the epistemic logic approach and all of them give up the additivity assumption of probability theory. Most of these approaches have been developed only since the middle of the XXth century with a sudden acceleration from 1975 on. This chapter should be of particular interest to psychologists, because they often stick to the alternative logic vs. probability, and are not always aware of the fact that there is a large area beyond.

## 2. Rationality axioms for set functions representing belief

In this section, we survey mathematical models of belief and axioms that have been proposed to represent them. We consider the most general approaches and show that accepting the rule of adjunction considerably narrows the expressive power of graded belief representations.

### 2.1 Degrees of belief

We assume throughout the paper that belief qualifies propositions, represented by subsets  $A$ ,  $B$ ,  $C$  ... of a set  $S$  of possible worlds or states of affairs. More precisely,  $A$  stands for the proposition that event  $A$  occurs, or equivalently that some entity of interest, say  $x$ , lies in  $A$ . We do not distinguish between logically equivalent propositions. Quantifying (the possible lack of) belief is one instance of expressing uncertainty for an agent.

The most usual representation of uncertainty consists in assigning, to each proposition or event  $A$ , a number  $g(A)$  in the unit interval. It evaluates the confidence (we deliberately use a more general term than "belief" or "plausibility") of the agent in the truth of proposition  $x \in A$ . This proposition can only be true or false by convention, even if the agent may ignore this truth-value. The following requirements sound natural:

$$g(\emptyset) = 0; g(S) = 1$$

as you cannot have confidence in contradictions, and you should believe in tautologies. Also natural is the monotonicity with respect to inclusion:

$$\text{If } A \subseteq B \text{ then } g(A) \leq g(B).$$

Indeed if  $A$  is more specific than  $B$  in the wide sense (in other words,  $A$  implies  $B$ ), a rational agent should not be more confident in  $A$  than in  $B$ : all things being equal, the more imprecise a proposition, the more certain it is. Under these properties, the function  $g$  is sometimes called a *capacity* (after Choquet (1953)), sometimes a *fuzzy measure* (after Sugeno (1977)). In order to stick to the uncertainty framework, it was also called a *confidence measure* (Dubois &

Prade, 1988). Such a set function represents the epistemic state of an agent, i.e., what an agent knows, or thinks he or she knows (irrespective of whether it is true or not)<sup>2</sup>.

An important particular case of confidence measure is the probability measure  $g = P$  which satisfies the *additivity* property

$$\text{If } A \cap B = \emptyset, \text{ then } P(A \cup B) = P(A) + P(B).$$

Probability measures have been very often used to represent degrees of belief, hence with a subjective flavour, as opposed to limits of frequencies of repeatable events, that account for random phenomena. Contrary to frequencies, belief may be attached to unique events. However, the additivity axiom sounds much more natural when modeling frequencies than for degrees of belief. Actually, in pioneering works of Bernoulli and Lambert, they were not supposed to be additive (Shafer, 1978).

To recover additivity, scientists of the mid-XXth century such as de Finetti (1937) have interpreted  $P(A)$  as the amount of money the agent is ready to pay for buying a lottery that earns 1\$ if the proposition  $A$  is true. Moreover it is assumed that the price is fair, namely the agent would accept to sell this ticket in the same conditions for the same price, and wants to avoid sure loss. The additivity axiom follows from this interpretation of degrees of belief. Hence probability measures should be used to represent degrees of belief. Such probabilities are said to be subjective (see chapter 1 of this section by Hájek & Staffel). Bayesian probabilists consider non-additive degrees of belief to be irrational since incurring sure money loss.

**Remark:** Supplying precise degrees of beliefs sounds cognitively very demanding. It seems more natural to only represent the relative strength of confidence of an agent between various propositions expressing his/her knowledge rather than trying to force him/her to deliver numerical evaluations. It is indeed easier to assert that a proposition is more credible than another, rather than assessing a degree of belief (whose meaning is not always simple to grasp), or even to guess a frequency for each of them. The idea of representing uncertainty by means of partial order relations over a set of events dates back to (Ramsey, 1926), (de Finetti, 1937), (Koopman, 1940). They tried to find an ordinal counterpart to subjective probabilities. Later, philosophers of logic such as David Lewis (1973) have considered other types of relations, including comparative possibilities in the framework of modal logic. We omit discussing this literature for the sake of brevity.

## 2.2 Accepted belief

Alternatively to measuring degrees of belief, one may want to consider propositions as beliefs accepted by an agent, if the latter is ready to reason as if such believed propositions were true. In particular if  $A$  and  $B$  are beliefs, the conjunction  $A \cap B$  is also a belief (a debated issue). But the empty set should not be a belief. Moreover if  $A$  is a belief and  $A \subseteq B$  then  $B$  is a belief. In other words, a set of beliefs should be consistent and deductively closed.

Consider a Boolean-valued set-function  $N$  from  $2^S$  to  $\{0, 1\}$  such that  $N(A) = 1$  if  $A$  is believed and 0 otherwise. It is clear that  $N$  is a confidence measure (monotonic under inclusion) that satisfies the law of minitivity:

---

<sup>2</sup> Halpern (2003) calls *plausibility measures* set-functions of this kind, not even assuming a total order on events (replacing  $[0, 1]$  by a partially ordered set). However this terminology may lead to confusion with Shafer's older plausibility functions (see Section 4.2).

$$N(A \cap B) = \min(N(A), N(B)).$$

This function is a special case of necessity measures (Dubois & Prade, 1988). It models the idea of certainty. It is easy to check that in the finite case there exists a non-empty subset  $E = \bigcap \{A: N(A) = 1\}$  and that  $N(A) = 1$  if and only if  $E \subseteq A$ . The set  $E$  contains all states the agent does not consider impossible, given his or her beliefs, which is the simplest representation of an agent's epistemic state<sup>3</sup>. In other words, function  $N$  describes propositions that can be proved from the epistemic state  $E$ .

### 2.3 Extracting beliefs from confidence measures

In order to relate accepted beliefs to graded beliefs, one should extract accepted beliefs from a confidence function  $g$ . A natural way of proceeding is to define a belief as a proposition in which an agent has enough confidence. So we should define a positive belief threshold  $\beta$  such that  $A$  is a belief if and only if  $g(A) \geq \beta > 0$ . However the closure of accepted beliefs under conjunction leads to endorse the following property

**Accepted belief postulate:** *If  $g(A) \geq \beta$  and  $g(B) \geq \beta$  then  $g(A \cap B) \geq \beta$*

This requirement is very strong. As the set of accepted beliefs should not include the empty set, it is clear that one should have that  $\min(g(A), g(B)) < \beta$  if  $A \cap B = \emptyset$ . Worse, if the postulate holds for any positive threshold  $\beta$ , then it is clear that  $g(A \cap B) \geq \min(g(A), g(B))$ , but as  $g$  is monotonic under inclusion, it enforces the equality, which comes down to the statement that a capacity represents accepted beliefs if and only if  $g(A \cap B) = \min(g(A), g(B))$ , that is  $g$  is a graded necessity measure, still denoted by  $N$ . Letting  $\iota$  from  $S$  to  $[0, 1]$  be the function defined by  $\iota(s) = N(S \setminus \{s\})$  (the degree of belief that the actual state of affairs is not  $s$ ), it is clear that  $N(A) = \min_{s \notin A} \iota(s)$ . The value  $1 - \iota(s)$  can be interpreted as the degree of plausibility  $\pi(s)$  of state  $s$ , where  $\pi$  is the membership function of a fuzzy epistemic state, usually called a possibility distribution (Zadeh, 1978). The set function  $\Pi$  from  $2^S$  to  $[0, 1]$  such that

$$\Pi(A) = 1 - N(A^c) = \max_{s \in A} \pi(s)$$

represents the degree of plausibility of  $A$ , measuring to what extent  $A$  is not totally ruled out by the agent. This setting is the one of *possibility theory* (Dubois & Prade, 1988). This means that possibility theory accounts for the notion of accepted belief.

Possibility theory was proposed by L. A. Zadeh (1978) in the late 1970's for representing uncertain pieces of information expressed by fuzzy linguistic statements, and later developed in an artificial intelligence perspective (Dubois & Prade, 1988, 1998). Formally speaking, the proposal is quite similar to the one made almost thirty years before by the economist G. L. S. Shackle (1949), who advocated and developed a non-probabilistic view of uncertainty based on the idea of degree of potential surprise. The degree of potential surprise attached to proposition  $A$  can be modelled as  $N(A^c)$ , namely the more you believe  $A^c$ , the more surprising you find the occurrence of  $A$ .

A necessity-like function was explicitly used by L. J. Cohen (1977) under the name "Baconian probability". It is related to the English philosopher Francis Bacon, while

---

<sup>3</sup> or doxastic state: we do not make the difference in this chapter.

probability was investigated by Pascal and followers, as pointed out by Cohen (1980). It is devoted to the idea of provability in contrast with probability, and it perfectly fits necessity measures. Especially if you can prove  $A$  with some confidence, you cannot at the same time claim you can prove its negation, which makes condition  $\min(g(A), g(A^c)) = 0$  natural. It is satisfied by necessity measures. So, condition  $N(A) > 0$  expresses that  $A$  is an accepted belief, the absolute value of  $N(A)$  expressing the strength of acceptance. Such Baconian probabilities, viewed as shades of certainty, are claimed to be more natural than probabilities for use in legal matters for instance. Deciding whether someone is guilty cannot be done using statistics, nor be based on betting probabilities: you must prove guilt using convincing dedicated arguments.

About a decade later, in the late 1980's, W. Spohn (1988) introduced the notion of *ordinal conditional functions*, now called *ranking functions*, as a basis for a dynamic theory of epistemic states. Ranking functions  $\kappa$  are a variant of potential surprise taking values on the non-negative integers, that is  $N_\kappa(A^c) = 2^{-\kappa(A)}$  is a degree of potential surprise. The theory of ranking functions (see Chapter 5.3 by Spohn et al. of this volume) and possibility theory can be developed in parallel<sup>4</sup>, even if they were independently devised.

Subjective probability has been justified by Savage (1954) from first principles in the setting of decision under uncertainty. This result has been very important to popularize probability as the natural way of representing degrees of belief. As it turns out, a similar approach has been carried out for possibility theory in the same act-based setting as Savage, albeit assuming a finite state space  $S$ . See (Dubois, Prade, & Sabbadin, 2001) for a detailed account. The approach leads to qualitative counterparts of expected utility and extends the Wald pessimistic and optimistic criteria to possibility distributions. They may be considered not enough discriminant and can be refined using special forms of expected utility that encode lexicographic refinements of min and max (leximin and leximax) (Fargier & Sabbadin, 2005). See (Dubois, Fargier, Prade, & Sabbadin, 2009) for a survey (see also Chapter 8.4 by Hill of this volume).

## 2.4 Probability and accepted beliefs

It is clear that set functions representing accepted beliefs and probability measures supposed to capture rational degrees of belief are at odds with one another. If we apply the threshold method to a probability function in order to recover accepted beliefs, then we fail to satisfy conjunctive closure since however large the threshold  $\beta$  is:

$$P(A) > \beta \text{ and } P(B) > \beta \text{ do not imply } P(A \cap B) > \beta \text{ for all } A, B.$$

This state of fact has been especially pointed out by Kyburg (1961, 1997). He proposed the lottery paradox. If  $S$  contains the set of possible equiprobable outcomes of a chance game like buying a lottery ticket where only one wins, then, by making  $S$  big enough, we can make the probability of losing when betting on  $s$ , namely,  $(|S|-1)/|S|$ , as high as possible. But if you buy all tickets you are sure to win. This example questions the cogency of accepted belief understood as high enough degree of belief. As a consequence, Kyburg proposed to give up the idea that accepted beliefs are deductively closed and constructed a specific logic accounting for this standpoint (Kyburg & Teng, 2012).

---

<sup>4</sup> Up to the presence or not of technical assumptions like well-ordering in the infinite setting.

Another way out of the lottery paradox is suggested in (Dubois, Fargier & Prade, 2004) by restricting the set of probabilities. We look for probability measures that respect the adjunction rule, namely for any context  $C$ ,

$$P(A|C) > P(A^c|C) \text{ and } P(B|C) > P(B^c|C) \text{ imply } P(A \cap B|C) > P(A^c \cup B^c|C) \text{ for all } A, B.$$

This is the adjunction rule for  $\beta = 0.5$ . Such probability functions do exist and they are so-called *big-stepped* probabilities (Benferhat, Dubois & Prade, 1999) or yet *atomic bound* probabilities (Snow, 1999); see also Leitgeb (2014)'s stability theory. They have probability masses of the form  $p_1 > p_2 > \dots > p_n$ , where  $p_i = P(\{s_{ij}\})$ , such that for all  $i < n$ ,  $p_i > p_{i+1} + p_{i+2} + \dots + p_n$ , i.e., they are the discrete counterpart of exponential probability functions. The same kind of result can be obtained for any acceptance threshold. This result suggests that the notion of accepted belief is more consistent with probability theory for distributions that are strongly biased towards some specific outcome (the opposite of uniform distributions used in the lottery paradox), that is, those for which common sense beliefs can be entertained.

### 3. Qualitative possibility theory: reasoning with defeasible accepted beliefs

As mentioned before, there are two different traditions for doxastic reasoning: modal logic that captures accepted beliefs by modalities, and set-functions that use probability. The fact that there is a class of set-functions that also captures accepted beliefs suggests a bridge between the two traditions. Let us also mention a third tradition that models partial ignorance using a truth-functional many-valued logic, e.g., Kleene logic that expresses partial knowledge about atomic propositions only. It severely limits its representation power (Ciucci & Dubois, 2013). We claim that qualitative possibility and necessity functions, valued in a bounded chain, offer a unified framework for these traditions. In particular, in this approach, accepted beliefs are defeasible, and it has close connections with non-monotonic reasoning, counterfactual reasoning and belief revision (see chapters 5.2 by Rott, and 6.1 by Starr).

#### 3.1 Boolean set-functions and modalities: a simple epistemic logic

Consider a propositional language  $\mathcal{L}$  where well-formed formulas  $p, q, \dots$  encode propositions. In artificial intelligence, a set of formulas in propositional logic is often called a *belief base*, and when deductively closed, a *belief set*. In the syntax of propositional logic, you can express the fact that a proposition is believed, but there is no way to express that it is not believed. All you can express is that its negation is believed. Introducing a belief modality  $\Box$  in front of  $p$  improves the situation as we can now distinguish between  $\neg\Box p$  ( $p$  is not believed) and  $\Box\neg p$  (its negation is believed). The standard approach to model accepted beliefs is indeed modal logic, after (Hintikka, 1962). Note that the meaning of modality  $\Box$  is described by the axioms ruling the logic. It can range from a very loose interpretation ("the agent has received information that  $p$  is true") to a very strong one ("the agent knows that  $p$ , in the sense of true belief").

The usual axioms of doxastic logic are those of the KD45 modal logic (see Chapter 5.1 by van Ditmarsch in this volume), which presupposes a modal language  $\mathcal{M}$  that extends  $\mathcal{L}$  and allows for nested modalities (the symbol  $\Rightarrow$  represents material implication):

(PL) All axioms of propositional logics for  $\mathcal{M}$ -formulas.

(K)]  $\Box(p \Rightarrow q) \Rightarrow (\Box p \Rightarrow \Box q)$

- (D)  $\Box p \Rightarrow \neg \Box \neg p$   
(S4)  $\Box p \Rightarrow \Box \Box p$   
(S5)  $\neg \Box p \Rightarrow \Box \neg \Box p$

Inference rules are modus ponens and necessitation (if  $p$  is a theorem, deduce  $\Box p$ ). The semantics is in terms of accessibility relations  $R \subseteq S \times S$ . The satisfaction of  $\Box p$  at a state  $s$  based on relation  $R$  is defined by  $sR \subseteq [p]$  where  $sR = \{s' : (s, s') \in R\}$  and  $[p]$  denotes the set of states where  $p$  is true.

This approach can be simplified in order to relate doxastic logic with the representation of accepted belief in the previous section:

- KD45 uses a complex language, while we can restrict it by putting modality  $\Box$  only in front of propositions in  $\mathcal{L}$ .
- Axioms S4 and S5, which are often called positive and negative introspection, seem to be there basically to make complex formulas of the language  $\mathcal{M}$  equivalent to simpler ones without nested modalities (even if they have a philosophical meaning).
- It may sound counterintuitive to evaluate doxastic formulas on a real state of affairs; it is more natural to do it on epistemic or doxastic states (non-empty subsets of  $S$ ).

The propositional language  $\mathcal{L}_\Box$ , whose atomic variables are of the form  $\Box p$ ,  $p \in \mathcal{L}$ , is the simplest language for an epistemic or doxastic logic. Note that it is disjoint from  $\mathcal{L}$  in the sense that it cannot express objective formulas. Then we only keep axioms K, D, and necessitation is modelled by an axiom saying that  $\Box p$  is valid whenever  $p$  is a tautology of the propositional calculus. This system is called MEL (minimal epistemic logic) (Banerjee & Dubois, 2014).

The semantics of MEL is in terms of simple epistemic states  $E \subseteq S$ , and the satisfaction of  $\Box p$  by  $E$  is then expressed by  $E \subseteq [p]$ , i.e.,  $p$  is true in all states that are not considered impossible for the agent, or alternatively  $p$  is believed, i.e.,  $N([p]) = I$  for the Boolean necessity measure equivalent to  $E$ . Indeed axiom K ensures that  $\Box(p \wedge q)$  is logically equivalent to  $\Box p \wedge \Box q$ , which is the axiom of necessity measures. Axiom (D) expresses that  $N(A) = I$  implies  $\Pi(A) = I$ . It ensures consistency (models of MEL are then non-empty subsets of  $S$ ).

In other words the logic MEL bridges the gap between epistemic logic (dropping the idea of introspection) and the representation of belief by means of set functions, identifying the latter with modalities. Especially the modality  $\Diamond p = \neg \Box \neg p$  corresponds to possibility measures.

### 3.2 The logic of graded acceptance

The logic of acceptance MEL can be extended to graded necessity functions  $N$  using a simple multimodal logic. If  $L$  is a finite totally ordered scale of necessity degrees (understood as describing a gradation in accepted beliefs), we can expand the modal language of the MEL logic to allow for several belief modalities denoted by  $\Box^\lambda$ , for  $\lambda \in L$ ,  $\lambda \neq 0$ , where the sentence  $\Box^\lambda p$  encodes the statement  $N([p]) \geq \lambda$ , and the  $\Box$  modality in MEL corresponds to  $\Box^1$  (expressing full belief). The language of this logic is thus a propositional language  $\mathcal{L}_\Box^L$  where atoms are of the form  $\Box^\lambda p$ ,  $p \in \mathcal{L}$ , for  $\lambda > 0 \in L$ . Note that the formula  $\neg \Box^\lambda p$  stands for  $N([p]) < \lambda$ , which, due to the assumed finiteness of  $L$  can be expressed as  $\Pi([p]) \geq 1 - s(\lambda)$ , which is encoded as  $\Diamond^{s(\lambda)} \neg p$ , for the lower value  $s(\lambda)$  next to  $\lambda$  in the scale  $L$ .



The sub-logic obtained by fixing value  $\lambda$ , is a copy of the logic MEL (it satisfies axiom K, D and necessitation axiom). There is also the weakening axiom:  $\Box^\lambda p \Rightarrow \Box^\mu p$  if  $\mu \leq \lambda$ , and axiom (D) is valid in a stronger form  $\Box^\lambda p \Rightarrow \Diamond^1 p$ . The semantics is in terms of L-valued possibility distributions  $\pi$  representing gradual epistemic states, and  $\pi$  satisfies  $\Box^\lambda p$  if and only if  $N([p]) \geq \lambda$  where  $N$  is based on  $\pi$ . Soundness and completeness of this logic, called GPL (generalized possibilistic logic), has been proved (Dubois, Prade & Schockaert, 2017).

This logic is very expressive and enables to reason about ignorance and defeasible beliefs. It can encode several non-monotonic formalisms especially:

- The older standard possibilistic logic (Dubois, Lang & Prade, 1994) is obtained by restricting the language  $\mathcal{L}_\Box$  to using only conjunctions of atomic epistemic statements  $\Box^\lambda p$ , which are written  $(p, \lambda)$  in the original syntax. This logic is non-monotonic (Dubois & Prade, 1998).
- Conditional logics with statements of relative belief of the form  $N(p) > N(q)$  can be encoded by GPL formulas of the form  $\bigvee_{\{\lambda \neq 0\}} \Box^\lambda p \wedge \neg \Box^\lambda q$ .
- System P of Kraus, Lehmann and Magidor (1990) where conditional statements express the plausible inference of  $q$  from  $p$  are modelled by the constraint  $N(p \Rightarrow q) > N(p \Rightarrow \neg q)$ .
- Answer set programs can be expressed in GPL using a three-valued scale  $L = \{1 > \lambda > 0\}$ , i.e., two necessity modalities: a strong one  $\Box^1$  and a weak one  $\Box^\lambda$ .

In summary, GPL can be viewed as the logic of qualitative Baconian probabilities.

## 4. Non-adjunctive settings for rational degrees of belief

If we consider that belief should come in degrees, and if, like Kyburg, we reject the adjunction rule, we are left with non-classical logics where deduction is not closed under conjunction, like for instance the logic of risky knowledge (Kyburg and Teng, 2012). Or we give up the Boolean framework of logic altogether, and concentrate on the properties of set functions that can model degrees of belief. The natural question is then whether degrees of belief should be additive at all. Fifty years ago, the answer was yes, of course. Since then, a number of proposals have emerged, from which it follows that such additivity should not be taken for granted. This section discusses some reasons for questioning additive beliefs and focuses on some approaches to graded belief that are deliberately non-additive.

### 4.1 Can a single probability distribution capture any epistemic state?

The so-called *Bayesian* approach to subjective probability theory posits a uniqueness principle as a preamble to any kind of uncertainty modelling: any state of knowledge is representable by a single probability distribution (see for instance (Lindley, 1982)). Note that indeed, if, following the fair bet procedure of de Finetti, an agent decides to directly assign subjective probabilities via buying prices to all possible outcomes in some game of chance, the coherence principle forces this agent to define a unique probability distribution in this case.

Yet another mathematical attempt to justify probability theory as the only reasonable belief measure is the one of R. T. Cox (1946). 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 degree of belief,  $A, B$  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)  $g(A^c|B) = n(g(A|B))$ ,  $B \neq \emptyset$ , where  $A^c$  is the complement of  $A$ ;
- iii) The 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 that  $g(A|B)$  must be isomorphic to a conditional probability measure.

This result 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 revise numerical degrees of beliefs (Horvitz, Heckerman & Langlotz, 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 (1999) has shown that the result does not hold on finite spaces, and needs an additional technical condition 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, which forbids the representation of uncertainty due to partial ignorance as seen later on. The above comments seriously weaken the alleged universality of Cox results.

Applying the Bayesian credo as recalled above, justified via the avoidance of Dutch books or by obedience to Cox axioms, forces the agent to use a single probability measure as the universal tool for representing uncertainty whatever its source. This stance leads to serious representation difficulties already pointed more than forty 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. One may indeed admit that additive degrees of beliefs are justified if they reflect extensive statistical evidence. But what if this information is not available?

Take the example of die tossing. The uniform probability assignment models the assumption that the die is fair. But if the agent assigns equal prices to bets assigned to all facets of the die, 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, (s)he could test it hundreds of times prior to placing the bets, or from counting cases)? 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 and handles the two situations in the same way.

Next, the choice of a set of mutually exclusive outcomes depends on the chosen language, e.g., the one used by the information source. However, 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 representation of the state space may conflict with a uniform one on another encoding of the same state space for the same problem, while in case of ignorance this is the only available model. 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 ( $v = li$ ), or not ( $v = \neg li$ ), then the agent proposes  $P_1(li) = P_1(\neg li) = 1/2$  on  $S_1 = \{li, \neg li\}$ . However, it makes sense to distinguish between animal life ( $ali$ ), and vegetal life only ( $vli$ ), which leads to the state space  $S_2 = \{ali, vli, \neg li\}$ . The ignorant Bayesian agent is then bound to propose  $P_2(ali) = P_2(vli) = P_2(\neg li) = 1/3$ . As " $li$ " is the disjunction of " $ali$ " and " $vli$ ", the distributions  $P_1$  and  $P_2$  are not compatible with each other, while they are supposed to represent ignorance. Another example comes from noticing that expressing ignorance about a real-valued quantity by means of a uniform distribution for  $x \in [a, b]$ , a positive interval, is not compatible with a uniform distribution on  $y = \log(x) \in [\log(a), \log(b)]$ , while the agent has the same level of ignorance about  $x$  and  $y$ .

Finally, Ellsberg (1961)'s paradox (see Chapter 8.2 by Peterson, in this volume) showed quite early that when expressing preferences between gambles consisting of drawing balls from an urn, the content of which is ill-known, many experiments have shown that people tend to systematically violate Savage axioms (especially the sure thing principle), because they tend to be pessimistic about rewarding events of unknown probability. One way of accounting for the results of these experiments is to give up additive beliefs.

The above limitations of expressive power of single probability distributions have motivated the emergence of other approaches to uncertainty representations. Some of them, as seen above, give up the numerical setting of degrees of belief and use ordinal or qualitative structures, like qualitative possibility theory. Another option is to tolerate incomplete information in the probabilistic approach, which leads to different mathematical models of various levels of generality. They are reviewed in the rest of this chapter.

## 4.2 Shafer belief functions and the merging of uncertain testimonies

The theory of evidence by Shafer (1976) can be viewed as a specific interpretation of Dempster (1967)'s upper and lower probability framework for handling imprecise statistical information, or as the revival of the concept of probability (as opposed to chance) invented in the XVIIth century by Bernoulli and later by Lambert, around the problem of representing and merging unreliable testimonies (Shafer, 1978). Shafer's book is clearly in the latter trend.

The main issue is first to model an unreliable testimony. Suppose that a witness claims that proposition  $E$  is true but the receiver only partially believes this statement, considering that the witness is reliable with probability  $p$ . So  $p$  can be viewed as the degree of belief of the receiver in proposition  $E$  due to the unreliability of the witness. The additivity issue is raised by the question: to what proposition should the complementary weight  $1 - p$  be assigned? The regular probabilist would assign it to the opposite proposition  $E^c$ . But if the testimony  $E$  is interpreted as  $E^c$  it means that the witness lies, i.e., says  $E$  when knowing it is false. There is another option: namely due to the incompetence of the witness, there is a probability  $1 - p$  that the testimony is just useless. In the latter case, the probability  $1 - p$  is assigned not to  $E^c$ , but to the whole of  $S$ , i.e., to the state of ignorance. Namely, with probability  $p$  the receiver *knows* that  $E$  is true, and he/she knows nothing with probability  $1 - p$ , which is modelled by a basic assignment  $m$  from the power set  $2^S$  to  $[0, 1]$  such that  $m(E) = p$  and  $m(S) = 1 - p$ . A belief function that models such a simple unreliable testimony  $E$  is called a *simple support function*.

More generally, consider a process whose outcomes are set-valued (i.e., imprecise), and uncertain (there is a probability value attached to each outcome). This is modelled by a more general basic assignment  $m$  from  $2^S$  to  $[0, 1]$ , such that  $m(\emptyset) = 0$ , and  $\sum_{E \subseteq S} m(E) = 1$ . Epistemic states  $E$  with  $m(E) > 0$  are called focal sets. The degree of belief in a proposition  $A$ , and its dual plausibility degree are then defined by

$$Bel(A) = \sum\{m(E): E \subseteq A, E \neq \emptyset\}; \quad Pl(A) = \sum\{m(E): E \cap A \neq \emptyset\} \quad (3)$$

It is clear that the belief function  $Bel$  is non-additive, e.g.,  $Bel(A) + Bel(A^c) \leq 1$ , and the degree of plausibility is  $Pl(A) = 1 - Bel(A^c)$ . In the case of a simple support function, observe that when  $A \neq S$ ,  $Bel(A) = p$  if  $E \subseteq A$ , and  $Bel(A) = 0$  otherwise.

It is important to point out that belief functions generalize probabilities (recovered when  $\forall E, m(E) > 0$  implies that  $E$  is a singleton), Boolean necessity measures, (recovered when  $m(E) = 1$  for some epistemic state  $E$ ) and also graded necessity measures (recovered focal sets are all nested, e.g., a simple support function). A degree of belief  $Bel(A)$  clearly evaluates the probability of proving  $A$  from the available information. A plausibility degree  $Pl(A)$  evaluates the probability that  $A$  is not logically incompatible with the available information. To use Cohen (1977) terminology, belief functions join the probable and the provable, or put Pascal and Bacon together. But doing so, belief functions are no longer additive, nor do they respect adjunction.

The major problem addressed by XVII and XVIIIth centuries pioneers is the merging of such testimonies. They proposed special cases of what is now known as Dempster's rule of combination. Let  $m_1$  and  $m_2$  be two mass assignments coming from independent sources. The result of the combination is a mass assignment  $m_1 \otimes m_2$  defined by:

$$\forall A \subseteq S, m_1 \otimes m_2 (A) = \sum\{m_1(A_1) \cdot m_2(A_2) : A = A_1 \cap A_2\} / K$$

where  $K = \sum\{m_1(A_1) \cdot m_2(A_2) : A_1 \cap A_2 \neq \emptyset\}$  and  $m_1 \otimes m_2 (\emptyset) = 0$ .

It consists in intersecting any two overlapping focal sets, each coming from a distinct source, computing the probability of obtaining each subset  $A$  via such an intersection, and renormalizing the obtained mass assignment as some pairs of focal sets may be conflicting. In the case of merging two simple support functions focusing on the same set  $E$  where  $Bel_1(E) = m_1(E) = p_1$  and  $Bel_2(E) = m_2(E) = p_2$ , the resulting belief in  $E$  is  $Bel_1 \otimes Bel_2(E) = p_1 + p_2 - p_1 p_2$ , which operates a reinforcement of the belief in  $E$ , a result already suggested by pioneers of belief functions in the XVII-XVIIIth centuries. This combination rule assumes that sources of information are independent, which makes the reinforcement effect plausible.

In the book of Shafer, a major question is whether all belief functions can be expressed as the result of merging independent simple testimonies in the form of simple support functions. It turns out that only a subclass of belief functions, called *separable*, can be generated in this way. Later on, Smets (1995) has tried to extend the notion of simple support function so that to cover all belief functions. An extensive presentation of the theory of evidence as a theory of rational belief is proposed by Haenni (2009).

Finally, decision criteria under uncertainty, when the latter is described by a belief function, are studied by Smets and Kennes (1994) and Jaffray (1989). The former propose to define a so-called *pignistic* probability measure from a belief function (generalizing Laplace principle of insufficient reason), and apply the expected utility criterion. This probability measure coincides with the well-known Shapley value in game theory (Shapley, 1953), and, in some sense projects the provable on the probable. Jaffray (1989) proposes and axiomatizes an extension of Hurwicz criterion.

#### 4.3 Imprecise Probabilities, Desirability, and Generalized Betting Theory

The alternative approach to the modelling of degrees of belief consists in revisiting de Finetti's approach to subjective probability, dropping the constraint that the price proposed by a gambler for buying a lottery ticket should be fair. This view was pioneered by C. Smith (1961), P.M. Williams (1975), R. Giles (1982), and more extensively developed by P. Walley (1991).

In this approach, the agent offers buying prices of gambles. A gamble is a function  $f$  from  $S$  to the real line that expresses losses ( $f(s) < 0$ ) or gains ( $f(s) > 0$ ) according to what is the actual state of affairs  $s$ . The gamble associated to an event is its characteristic function. The agent is not committed to selling such gambles at the same prices as their buying prices. The approach relies on so-called *desirable gambles* (Walley, 1991) that the agent would agree to buy for a positive price. The set of desirable gambles contains at least all positive gambles. Moreover, the sum of two desirable gambles is considered desirable, and a desirable gamble remains desirable when multiplied by a positive constant. The maximal price at which the agent accepts to buy a gamble is the maximal value  $\alpha$  such that  $f - \alpha$  is desirable. It is called the *lower prevision* of a gamble  $f$ . It can be shown that given a set of gambles  $\mathcal{G}$  and their lower previsions  $LP(f)$ , there is a convex set of probabilities  $\mathcal{P}_{\mathcal{G}}$ , called *credal set*, such that  $LP(f) = \inf\{E_P(f) : P \in \mathcal{P}_{\mathcal{G}}\}$  is the lower expectation of  $f$  according to  $\mathcal{P}_{\mathcal{G}}$ , for all  $f \in \mathcal{G}$ , where  $E_P(f)$  is the expectation of  $f$  with respect to probability  $P$ . One important point is that any convex set of probabilities can be represented by lower previsions on some family of gambles.

In this setting, the upper prevision  $UP(f)$  of a gamble  $f$  is provably equal to  $-LP(-f)$ . The upper prevision  $UP(f)$  is thus the minimal selling price of  $f$ . If the credal set attached to a set of gambles and its lower previsions is empty, then the proposal is inconsistent and the agent incurs a sure loss after buying and resolving these gambles. Avoiding sure loss means that  $UP(f) \geq LP(f)$  for all gambles  $f$ .

Moreover, due to the interaction between gambles, it may be that the consistent buying prices proposed by the agent for gambles  $f_i \in \mathcal{G}$  are too low and could be raised without altering the credal set. A set of buying prices  $bp(f_i)$ ,  $f_i \in \mathcal{G}$  is said to be *coherent* if and only if  $LP(f_i) = bp(f_i)$ ,  $\forall f_i \in \mathcal{G}$ . In other words, a set of buying prices for a set of gambles  $\mathcal{G}$  is coherent if and only if for any  $f_i \in \mathcal{G}$ , it holds that

$$\inf\{E_P(f_i) : E_P(f_j) \geq bp(f_j), \forall f_j \in \mathcal{G}\} = bp(f_i).$$

Under this approach the degree of belief in proposition  $A$  is a coherent lower probability  $P_*(A) = \inf\{P(A) : P \in \mathcal{P}\} = LP(1_A)$ , the lower prevision of its characteristic function, where  $\mathcal{P}$  is the credal set induced by the lower prevision  $LP(f)$  on some gambles.

Some remarks are in order to position this approach with respect to other rational approaches to degrees of belief:

- The epistemic state of the agent is here represented by a credal set  $\mathcal{P}$ , but there is no ill-known probability inside. In particular, the interval  $[P_*(A), P^*(A)]$ , where  $P^*(A) = 1 - P_*(A^c)$  is not supposed to contain an ill-known subjective (nor objective) probability of  $A$ . Just as for belief functions, degrees of belief are precise and modeled by coherent lower probabilities.
- Mathematically, belief functions are a special case of coherent lower probabilities. They are super-additive set-functions at any order, while lower probabilities from any credal set only satisfy the inequality  $P_*(A \cup B) \geq P_*(A) + P_*(B)$  when  $A$  and  $B$  are mutually exclusive. In particular, the mass function recomputed from  $P_*$  instead of  $Bel$  in (3) (called the Moebius transform of  $P_*$ ) exists, is unique but not necessarily positive (Chateauneuf & Jaffray, 1989).
- An attempt to justify belief functions as the only rational approach to degrees of belief under a betting framework in the style of Walley (and not as unreliable testimonies) was recently published by Kerkvliet and Meester (2018).

The gamble approach leads to a decision rule that is specific to the imprecise probability setting, namely a gamble  $f$  is preferred to a gamble  $g$  if and only the gamble  $h = f - g$  is desirable, i.e., if the lower expectation of this gamble with respect to the corresponding credal set  $\mathcal{P}$  is positive. It gives a partial ordering on gambles. It implies that  $LP(f) \geq LP(g)$ . The latter inequality solves the Ellsberg paradox, in contrast with  $LP(f - g) \geq 0$ .

#### 4.4 Quantitative possibility in the setting of imprecise probability

It is natural to reconsider graded possibility and necessity measures in the setting of belief functions and imprecise probabilities. In fact they are at the crossroads of all non-additive approaches to uncertainty and may be interpreted in various ways:

- Necessity measures are a special case of belief functions. Their characteristic property is to have nested focal sets. In other words, they model coherent arguments in favor or disfavor of propositions. They are the only family of belief functions that obey the adjunction rule. Note that the weaker Baconian condition  $\min(Bel(A), Bel(A^c)) = 0$  for all  $A \subseteq S$  corresponds to overlapping (consistent) focal sets.
- As a consequence, necessity measures also stand for coherent lower probabilities. However, they correspond to a very cautious type of betting behavior, such that if the buying price for gambling on  $A$  is positive, then the agent feels obliged to sell this gamble at the maximal price (Giles, 1982).
- One may borrow the operational semantics of the Bayesians to derive personal possibility and necessity degrees. If we adopt the framework of belief functions for representing an agent's knowledge and accept the idea that a belief function induces a pignistic probability for making decisions, then we may reverse this process. Given a subjective probability reflecting fair prices of random events, one may look for the least informative belief function that induces this subjective probability. It can be proved that it is always a necessity measure (Dubois, Prade & Smets, 2008).
- Necessity functions induced on the unit interval by a suitable transformation of a Spohn ranking function (see Chapter 5.3 by Spohn et al. of this volume) have nothing to do with

lower probabilities. Basically, as shown in (Spohn, 1990) they are more related to powers of infinitesimal probabilities, for which the additivity axiom degenerates in the minitivity axiom.

- Yet another interpretation of possibility theory is in terms of likelihood. In statistical inference, given a parametric probabilistic model  $P(\cdot|\theta)$  where  $\theta \in \Theta$  is the parameter of the model, the probability  $P(R|\theta)$  based on data set  $R$  is not the probability of  $\theta$  based on  $R$ , only its likelihood. It represents a looser degree  $lik(\theta) = P(R|\theta)$  of confidence in  $\theta$  for the observer having received evidence  $R$ . Advocates of the likelihood approach (Edwards, 1972) are completely reluctant to attach prior probabilities to values of  $\theta$  basically because this quantity is not observable and is just a model artefact. Rather, it is natural to try and define the likelihood  $lik(A)$  for any  $A \subseteq \Theta$ , from the values  $lik(\theta)$ ,  $\theta \in A$ . It has been shown that the only meaningful definition is  $lik(A) = \max\{P(R|\theta) : \theta \in A\}$  (Dubois, Moral & Prade, 1997) (Dubois, 2006). Hence a likelihood function can be interpreted as a possibility measure, in the absence of prior probabilities. However, this kind of possibility measure is defined only up to a multiplicative constant, a specific feature that makes likelihood theory yet another kind of possibility theory.

## 5. Conclusion

This chapter has proposed a survey of various approaches to the notion of belief reflecting the progress made in the recent 50 years. It seems that the frontal opposition between degrees of belief and accepted beliefs, i.e., Pascal and Bacon traditions, may be alleviated to some extent if we give up the requirement that degrees of belief should be additive. There is a range of mathematical models between probability and modal logics, some of which retain the adjunction rule of Baconian probabilities. Some approaches blend the two traditions, and are consistent with the requirement that you cannot accept beliefs in a proposition and beliefs in its contrary. The Baconian tradition also comes close to the issue of formal argumentation, on which there is an abundant literature to-day (see (Haenni, 2009) for its connection with Shafer belief functions). Argumentation can be viewed as a rational approach to handle inconsistency in reasoning, due to conflicting pieces of information. One may argue that Baconian probabilities (in the form of, e.g., necessity functions, ranking functions and the like) represent imprecise but conflict-free information, while probabilities capture precise but conflicting observations. The new theories of belief deal with both imprecise and conflicting information and seem to bridge the gap between the two traditions of belief representation. One may then consider belief in a more dynamical setting, where starting with more or less probable conflicting evidence, one proceeds towards the provable via a suitable deliberation process involving argumentation.

## References

Banerjee, M., & Dubois, D. (2014). A simple logic for reasoning about incomplete knowledge. *Int. J. of Approximate Reasoning*, 55, 639-653.

Benferhat, S., Dubois, D., & Prade, H. (1999). Possibilistic and standard probabilistic semantics of conditional knowledge bases. *J. of Logic and Computation*, 9, 873-895.

Chateauneuf, A., Jaffray, J-Y (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17(3), 263-283.

Cheeseman, P. (1988) An inquiry into computer understanding (with comments and a rejoinder). *Computational Intelligence*, 4, 58-142.

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.

Cohen, L. J. (1977). *The Probable and The Provable*. Clarendon Press, Oxford, UK.

Cohen, L. J. (1980). Some historical remarks on the Baconian conception of probability. *Journal of the History of Ideas*, 41, 219-231.

Cox R. T. (1946). Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14, 1-13.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Poincaré*, 7, 1-68.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Statistics*, 28, 325-339.

Dubois, D. (2006). Possibility theory and statistical reasoning. *Computational Statistics & Data Analysis*, 51(1), 47-69.

Dubois, D., Fargier, H. & Prade, H. (2004). Ordinal and probabilistic representations of acceptance. *J. Artificial Intelligence Research*, 22, 23-56.

Dubois, D., Fargier, H., Prade, H. & Sabbadin R. (2009). A survey of qualitative decision rules under uncertainty. In D. Bouyssou, D. Dubois, M. Pirlot, & H. Prade (Eds.), *Decision-making Process - Concepts and Methods*, chapter 11 (pp. 435-473). London: ISTE & Wiley.

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

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

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

Dubois, D. & Prade, H. (1998). Possibility theory: Qualitative and quantitative aspects. In D. M. Gabbay, Ph. Smets, (Eds.), *Quantified Representation of Uncertainty and Imprecision*, vol. 1 of the *Handbook of Defeasible Reasoning and Uncertainty Management Systems* (pp. 169-226). Dordrecht, The Netherlands: Kluwer Academic Publications.

Dubois, D., Prade, H. & Sabbadin R. (2001). Decision-theoretic foundations of qualitative possibility theory. *Europ. J. of Operational Research*, 128(3), 459-478.



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

Dubois, D., Prade, H. & Smets P. (2008). A definition of subjective possibility. *Int. J. Approximate Reasoning*, 48:352-364, 2008.

Edwards A. W. F. (1972). *Likelihood*. Cambridge, UK: Cambridge Univ. Press.

Ellsberg, D. (1961). Risk, ambiguity and Savage axioms. *Quarterly Journal of Economics*, 75(4), 643-669.

Fargier, H. & Sabbadin, R. (2005). Qualitative decision under uncertainty: back to expected utility. *Artificial Intelligence*, 164(1-2), 245-280.

Giles, R. (1982). Foundations for a theory of possibility. In M. Gupta, E. Sanchez, (Eds.). *Fuzzy Information and Decision Processes* (pp. 183-195). Amsterdam: North-Holland.

Haenni, R. (2009) Non-additive degrees of belief. In F. Huber, C. Schmidt-Petri (Eds). *Degrees of Belief* (pp. 121-160). Berlin: Springer.

Halpern, J. Y. (1999). A counterexample to theorems of Cox and Fine. *J. Artificial Intelligence Research*, 10, 67-85. Technical addendum: Cox's theorem revisited. *ibidem*, 11, 429-435.

Halpern, J. Y. (2003). *Reasoning about Uncertainty*. Cambridge, Mass.: MIT Press.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca, N.Y.: Cornell University Press.

Horvitz E., Heckerman D. & Langlotz C. (1986). A framework for comparing alternative formalisms for plausible reasoning. In T. Kehler (Ed.), *Proc. 5th Nat. Conf. on Artificial Intelligence (AAAI'86)*, Philadelphia, Vol. 1 (pp. 210-214). San Mateo, Ca.; Morgan Kaufmann.

Huber, F. & Schmidt-Petri, C. (Eds.). (2009). *Degrees of Belief*. Berlin: Springer.

Jaffray, J.Y. (1989) Linear utility theory for belief functions. *Operations Research Letters*, 8, 107-112.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge, UK: Cambridge Univ. Press.

Kerkvliet, T. & Meester, R. (2018). A behavioral interpretation of belief functions. *Journal of Theoretical Probability*. To appear.

Koopman, B.O. (1940). The bases of probability. *Bulletin American. Mathematical. Society*, 46, 763-774.

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

Kyburg, H. (1961) *Probability and the Logic of Rational Belief*. Middletown, CT: Wesleyan University Press, 1961.

Kyburg, H. (1997). The rule of adjunction and reasonable inference. *Journal of Philosophy*, 94, 109-125.

Kyburg, H. & Teng C. M. (2012). The logic of risky knowledge, reprised. *Int. J. Approx. Reasoning*, 53, 274-285.

Leitgeb, H. (2014). The stability theory of belief. *Philosophical Review*, 123(2), 131-171.

Lewis, D.K. (1973). *Counterfactuals*. London: Basil Blackwell (2nd edition, Worcester, UK: Billing and Sons Ltd., 1986).

Lindley, D. V. (1982). Scoring rules and the inevitability of probability. *Int. Statistics Review*, 50, 1-26.

Paris, J. (1994). *The Uncertain Reasoner's Companion*. Cambridge, UK: Cambridge University Press.

Ramsey, F. P. (1926). Truth and probability. Republished in J. Kyburg, H.E. Smokler (Eds.), *Studies in Subjective Probability* (pp. 23-52). Huntington, N.Y.: Krieger Pub. Co, 1980.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

Shackle, G. L. S. (1949). *Expectation in Economics*. Cambridge, UK: Cambridge University Press.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, N.J.: 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.

Shapley, L. S. (1953). A value for n-person games. In Kuhn and Tucker, (Eds.), *Contributions to the Theory of Games*, II (pp. 307–317). Princeton, N.J : Princeton University Press.

Smets P. (1995). The canonical decomposition of a weighted belief. *Proc. Int. Joint Conf. on Artificial Intelligence* (pp. 1896-1901), San Mateo, Ca.: Morgan Kaufman.

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

Smith, C. A. B. (1961). Consistency in statistical inference and decision. *J. of the Royal Statistical Society, B-23*, 1-37.

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 W. L. Harper, B. Skyrms, (Eds.). *Causation in Decision, Belief Change, and Statistics*, vol. 2, (pp. 105-134). Dordrecht: Kluwer.

Spohn, W. (1990) A general, nonprobabilistic theory of inductive reasoning. In R. D. Shachter, T. S. Levitt, L. N. Kanal, J. F. Lemmer, (Eds.). *Uncertainty in Artificial Intelligence* vol. 4, (pp. 149-158). Amsterdam: North Holland.

M. Sugeno. (1977). Fuzzy measures and fuzzy integrals - A survey. In M. M. Gupta, G. N. Saridis, B. R. Gaines, (Eds.). *Fuzzy Automata and Decision Processes* (pp. 89-102). Amsterdam: North Holland.

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

Williams, P. M. (1975) Notes on conditional previsions (unpublished research report). *Int. J. Approximate reasoning*, 44, 366-383, 2007.

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