Artificial Intelligence 201 (2013) 1–31

Contents lists available at SciVerse ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Belief functions on distributive lattices ☆

Chunlai Zhou

Department of Computer Science and Technology, School of Information, Renmin University, Beijing, 100872, China

ARTICLE INFO

Article history:
Received 18 August 2012
Received in revised form 7 May 2013
Accepted 10 May 2013
Available online 15 May 2013

Keywords:
Dempster–Shafer theory
Möbius transforms
Distributive lattices
de Morgan lattices
First degree entailments

ABSTRACT

The Dempster–Shafer theory of belief functions is an important approach to deal with uncertainty in AI. In the theory, belief functions are defined on *Boolean* algebras of events. In many applications of belief functions in real world problems, however, the objects that we manipulate is no more a Boolean algebra but a *distributive lattice*. In this paper, we employ Birkhoff's representation theorem for finite distributive lattices to extend the Dempster–Shafer theory to the setting of distributive lattices, which has a mathematical theory as attractive as in that of Boolean algebras. Moreover, we use this more general theory to provide a framework for reasoning about belief functions in a deductive approach on non-classical formalisms which assume a setting of distributive lattices. As an illustration of this approach, we investigate the theory of belief functions for a simple epistemic logic the first-degree-entailment fragment of relevance logic **R** by providing an axiomatization for reasoning about belief functions for this logic and by showing that the complexity of the satisfiability problem of a belief formula with respect to the class of the corresponding Dempster–Shafer structures is *NP*-complete.

1. Introduction

Dealing with uncertainty is a fundamental issue for Artificial Intelligence [31]. Numerous approaches have been proposed, including Dempster–Shafer theory of belief functions (also called Dempster–Shafer theory of evidence). Ever since the pioneering works by Dempster [10] and Shafer [52], belief functions were brought into a practically usable form by Smets [59] and have become a standard tool in Artificial Intelligence for knowledge representation and decision-making.

Dempster–Shafer belief functions on a finite frame of discernment S are defined on the power set of S, which is a *Boolean algebra*. They have an attractive mathematical theory and many intuitively appealing properties. Belief functions satisfy the three axioms which generalize the Kolmogorov axioms for probability functions. Interestingly enough, they can also be characterized in terms of *mass functions* m. Intuitively, for a subset event A, $m(A)$ measures the belief that an agent commits *exactly* to A, not the total belief that an agent commits to A. Shafer [52] showed that an agent's belief in A is the sum of the masses he has assigned to all the subsets of A. This characterization of belief functions through mass functions is simply an example of the well-known Inclusion–Exclusion principle in Enumerative Combinatorics [61] and hence has a strong combinatorial flavor. In this theory, mass functions are recognized as Möbius transforms of belief functions.

Dempster–Shafer theory of belief functions is closely related to other approaches dealing with uncertainty. It includes the Bayesian theory [51] as a special case. The first three rules of the Bayesian theory are simply those three axioms for probability functions. It is shown [52] that a belief function on S is *Bayesian* (also a probability function) if and only if

☆ This is an expanded and improved version of a preliminary paper with the same title that appears in *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* (AAAI-12), 1968–1975, Toronto, 2012. All the technical proofs here are not published in the AAAI paper. In some sense, the proof techniques developed in this paper are as important as the main results. In addition, Sections 3 and 4 in the conference paper are totally revised and replaced by more comprehensive sections, and some main results there get strengthened in this paper.

E-mail addresses: czhou@ruc.edu.cn, chunlai.zhou@gmail.com.

0004-3702/$ – see front matter
http://dx.doi.org/10.1016/j.artint.2013.05.003

its corresponding mass function assigns *positive* weights only to singletons. So a Bayesian belief function μ is more like a point function than a set function in its level of complexity in the sense that μ is determined by its values at *singletons* rather by its values at all events (its values at other non-singletons are 0). This implies that generally Bayesian belief functions are simpler and easier to describe than belief functions. As Shafer pointed out in Chapter 9 of his book [52], this simplicity makes the Bayesian belief functions awkward for the representation of evidence. In practice, the Bayesian approach is criticized for having difficulty in efficiently providing reasonable estimate of the probability of some events and for describing confidence by a single point rather than a range [18]. The main advantage of the Dempster–Shafer theory over the Bayesian theory is that it allows for a proper representation of ignorance under incomplete information and assigns a meaningful interval to an event as a representation of the uncertainty of the event. In this aspect, the theory of belief functions is equivalent to the approach adopted by Fagin and Halpern [18] dealing with uncertainty using inner and outer probability measures. In contrast with the Dempster–Shafer theory, probability theory does not assign a probability to every event [30] and probability measures are defined on a σ-algebra, which is a *subclass* of the power set of the space under consideration. *Non-measurable events*, those *without* probabilities, are usually considered as meaningless in probability theory. However, in modeling uncertainty in artificial intelligence, they are used to represent those events to which an agent has insufficient information to assign probabilities. A non-measurable event E is provided with inner probability measure (outer probability measure) which is the probability of the largest measurable event contained in E (the smallest measurable one containing E). The inner (outer) measure gives a lower bound (upper bound) on the agent's degree of beliefs in E. So inner probability measure (together with outer probability measure) induced by probability measure assigns an interval to every event as a representation of uncertainty of the event E, which is similar to belief/plausibility functions. Moreover, belief functions and inner probability measures are equivalent on *formulas* [18]. There is an immediate payoff to this view of Dempster–Shafer belief functions: a logic for reasoning about belief functions can be obtained from that for inner probability measures [19].

The first investigation of mathematical properties of belief functions on more general lattices was initiated by Barthélemy [4] with the combinatorial theory on lattices by Rota [47], which was motivated by possible applications of belief functions for *non-standard* representation of knowledge. Grabisch [26] continued along this direction and showed that such properties as Dempster's rule of combination and Smets's canonical decomposition [57] in the case of Boolean algebras can be transposed in *general lattice setting*. This generalized theory has been applied to many objects in real world problems that may not form a Boolean algebra. Let us give some examples: set-valued variables in *multi-label classification* [13,70], the set of partitions in *ensemble clustering* [39] and *bi-capacities* in cooperative game theory [29]. Because of its generality, however, Grabisch's theory loses many intuitively appealing properties in the Dempster–Shafer theory. For example, since a lattice does not necessarily admit a probability function [26], belief functions in general lattice settings fail to maintain a close connection with the Bayesian theory and therefore lack many of the desirable properties associated with this theory as in Dempster–Shafer theory [18]. Since pignistic probabilities are used for decision-making in the transferable belief model by Smets [65,58], it would be impossible to develop decision theory for general lattice structures along a similar line. In particular, most non-classical formalisms[1] in AI assume a setting of distributive lattices and hence it would be difficult to apply Grabisch's theory directly to obtain a deductive approach for reasoning about belief functions over these formalisms as developed in Section 5 of this paper. Our deductive approach *requires* a setting of distributive lattices. Moreover, in many real-world problems such as evidential reasoning on fuzzy events [68,55,56,66] and bipolar belief pairs on vague propositions [36] which do assume a setting of distributive lattices, belief functions are defined in *totally different* forms. It would be desirable to establish a *uniform* theory for all these applications of belief functions.

An optimal balance between utility and elegance of a theory of belief functions is achieved for *distributive lattices*, which is the main contribution of this paper. Not only does our approach for distributive lattices yield a mathematical theory as appealing as Dempster–Shafer theory, but also its applications extend to many *non-classical* formalisms of structures in Artificial Intelligence (quantum theory [67] is one of very few important exceptions). The main difficulty in the extension is how to characterize the class of belief functions *without* reference to mass assignments. Birkhoff's fundamental theorem for finite distributive lattices solves this problem. Through this characterization, many fundamental properties of belief functions in the Boolean case are also preserved in distributive lattices. Just as in the Dempster–Shafer theory, we obtain nice formulas for Möbius functions of distributive lattices. We show that, for any lattice (not necessarily distributive), a capacity is totally monotone iff its Möbius transform is non-negative, which answers an open question raised in [26] and further justifies the proposal by Barthélemy and Grabisch that evidential reasoning can be done naturally on lattices while probability theory can only live on distributive lattices. A more fundamental result is that there is a one-to-one correspondence from the perspective viewing belief functions as *lower envelopes* between belief functions on a distributive lattice L and those on its *generated Boolean algebra*, which is the smallest Boolean algebra of which L is a sublattice. We establish a close connection to the Bayesian theory by showing that a belief function *Bel* on a finite distributive lattice L is Bayesian (or a probability function) if and only if all *focal* elements i.e., those elements with positive weights assigned by its mass function are *join-irreducible* in L. For belief functions, join-irreducibles to distributive lattices are the same as singletons to Boolean algebras. Moreover, we prove by appealing to Smiley's theory on measurability in lattices [60] that belief functions on distributive lattices can be viewed as generalized probability functions in the sense that *Bel* is equivalent to the inner

[1] In this paper, being non-classical means being non-Boolean.

measure induced by some probability function on formulas. This perspective paves the way to our following deductive approach for reasoning about belief functions.

After establishing the mathematical theory for belief functions on distributive lattices, we use this more general theory to provide a framework for reasoning about belief functions in a deductive approach on non-classical formalisms which assume a distributive lattice. The integrating of belief functions and non-classical formalisms is intended to master two sources of ignorance. While belief functions take care of the limitation of the information that the agent has *at his disposal*, non-classical formalisms usually take care of the incompleteness or inconsistency in the knowledge-base due to imperfect data. So, the advantage of our framework over the well-known work by Fagin and Halpern [18] is that it provides a deductive machinery to reason about belief functions for these non-classical formalisms. As an illustration of this deductive approach, we deal with belief functions on two particular classes of distributive lattices: bilattices and de Morgan lattices, which are actually mathematical objects in reasoning under incomplete and inconsistent information. A well-known simple *non-Boolean* epistemic logic the first-degree-entailment fragment $\mathbf{R}_{fde}$ of relevance logic $\mathbf{R}$ [1] provides a complete deductive system for this type of non-classical information, which is used to deal with the famous *logical-omniscience problem* in the foundations of Knowledge Representation [20,37], and used for reasoning in the presence of inconsistency in knowledge base systems [38]. A sound and complete axiomatization is provided for the integration of belief functions and the non-classical logic $\mathbf{R}_{fde}$, and finally the complexity of the satisfiability problem of a belief formula with respect to the class of the corresponding Dempster–Shafer structures is shown to be *NP*-complete.

Although we have generalized *only some* important properties in the Dempster–Shafer theory to the setting of distributive lattices, we have no doubt that almost all other important properties in this theory can be transposed in a similar way through the techniques developed in this paper. Note that here we mainly focus on the perspective viewing belief functions as generalized probability functions, which is one of the two perspectives for belief functions pointed out by Halpern and Fagin [32]. The rest is organized as follows. In Section 2, we provide a preliminary background on lattice theory and belief functions on lattices, and answer an open question raised by Grabisch [26]. Section 3 is the main part of this paper. There we generalize the Dempster–Shafer theory on Boolean algebras to the setting of distributive lattices through Birkhoff's fundamental theory for finite distributive lattices, establish a close connection to the Bayesian theory and provide a perspective from which belief functions on distributive lattices can be regarded as generalized probability functions. In Section 4, we present the semantical framework for belief functions on both distributive bilattices and de Morgan lattices. In Section 5, we provide a sound and complete deductive system for reasoning about belief functions for the first-degree-entailment fragment of $\mathbf{R}$ and show that the complexity of the satisfiability problem of belief formulas with respect to the class of the corresponding Dempster–Shafer structures is *NP*-complete. And in the final section, we discuss related work and possible applications of our theory in this paper to knowledge representation and decision-making. Appendix A provides some properties of distributive lattices which are needed in other parts of this paper.

2. Belief function on lattices

We will first recall some basic definitions about lattices. Next Dempster–Shafer theory of belief functions on Boolean algebras will be generalized to this more general setting.

All posets and lattices occurring in this paper are supposed to be finite. All lattice-theoretical notation and terminology in this paper follows [61].

2.1. Lattices

A *partially ordered set* P (or *poset* for short) is a set (which by abuse of notation we also call P), together with a binary relation denoted $\leqslant$, satisfying the following three axioms:

1. For all $t \in P, t \leqslant t$ (*reflexivity*).
2. If $s \leqslant t$ and $t \leqslant s$, then $s = t$ (*antisymmetry*).
3. If $s \leqslant t$ and $t \leqslant u$, then $s \leqslant u$ (*transitivity*).

A subset I of P is called an *order ideal* (or *semi-ideal* or *down-set* or *decreasing subset*) if, for any $x, y \in L$, $x \in I$ provided that $x \leqslant y$ and $y \in I$. I is called a *principal order ideal* if $I = \{y \in L\colon\ y \leqslant x\}$ for some $x \in I$. Otherwise, it is called a *non-principal order ideal*. Dually, a subset F of L is called a *dual order ideal* (or *up-set* or *increasing subset* or *filter*) if, for any $x, y \in L$, $y \in F$ provided that $x \leqslant y$ and $x \in F$.

A strict partial ordering $<$ is defined from $\leqslant$ as $x < y$ if $x \leqslant y$ and $x \neq y$. x is said to *cover* y if $x > y$ and no other element z such that $x < z < y$. A *chain* is a poset in which any two elements are comparable. An *antichain* is a subset C' of L such that any two distinct elements of C' are incomparable. A subset C of a poset P is called a *chain* if C is a chain when regarded as an ordered set in its own right. The chain C of P is called *saturated* (also *maximal*) if there does not exist $z \in P - C$ such that $x < z < y$ for some $x, y \in C$ and such that $C \cup \{z\}$ is a chain. The *length* $l(C)$ of a finite chain is defined by $l(C) := |C| - 1$. The *height* (or rank) of a finite poset is $l(P) := max\{C\colon\ C \text{ is a chain of } P\}$. If every maximal chain of P has the same height n, then we say that P is *ranked*. In this case, the associated *rank function* $\rho : P \to \{0, 1, \dots, n\}$ is defined as: $\rho(x) = 0$ if x is a minimal element of P, and $\rho(y) = \rho(x) + 1$ if y covers x in P.

If s and t belong to a poset P, then an *upper bound* of s and t is an element $u \in P$ satisfying $u \geqslant s$ and $u \geqslant t$. A *least upper bound* (or *join* or *supremum*) of s and t is an upper bound u of s and t such that every upper bound v of s and t satisfies $v \geqslant u$. If a least upper bound exists, then it is clearly unique and is denoted $s \vee t$. Dually one can define the *greatest lower bound* (or *meet* or *infimum*) $s \wedge t$, when it exists. A *lattice* is a poset L for which every pair of elements has a least upper bound and greatest lower bound and is denoted $\langle L, \vee, \wedge \rangle$. When the context is clear, we simply write the lattice as L. For finite lattices, the greatest element (denoted $\top$) and the least element (denoted $\bot$) always exist. An element s of the lattice L is called *join-irreducible* if $s \neq \bot$ and one cannot write $s = t \vee u$ where $t < s$ and $u < s$; otherwise, it is called *join-reducible*. *Meet-irreducible* as well as *meet-reducible* is defined dually. A lattice $\langle L', \vee', \wedge' \rangle$ is called a *sublattice* of the lattice $\langle L, \vee, \wedge \rangle$ if $L' \subseteq L$ and, for any two elements $x, y \in L'$, $x \wedge' y = x \wedge y$ and $x \vee' y = x \vee y$.

For a given lattice L, a subset I of L is called an *ideal* if it is an order ideal with respect to the associated partial order and additionally, for any $x, y \in L$, $x \vee y \in I$. Dually, a subset F of L is called a *filter* if it is a dual order ideal (or filter) with respect to the associated partial order and additionally, for any $x, y \in F$, $x \wedge y \in F$. F is called a *prime filter* if it satisfies the following additional condition: for any $a, b \in L$, $a \in F$ or $b \in F$ whenever $a \vee b \in F$.

The map $h : \langle L, \wedge, \vee \rangle \to \langle L', \wedge', \vee' \rangle$ where L and L' are two lattices is called a *homomorphism* if, for any two elements $x, y \in L$, $h(x \wedge y) = h(x) \wedge' h(y)$ and $h(x \vee y) = h(x) \vee' h(y)$. If h is onto, it is called a *surjective homomorphism*. And if it is one-to-one and onto, then it is called an *isomorphism*. L can be *embedded* into L' if L is isomorphic to a sublattice of L' (denoted as $L \preccurlyeq L'$).

A lattice L is *distributive* if $(x \vee y) \wedge z = (x \wedge z) \vee (y \wedge z)$ holds for all $x, y, z \in L$. For any $x \in L$, we say that x has a *complement* in L if there exists $x' \in L$ such that $x \wedge x' = \bot$ and $x \vee x' = \top$. L is said to be *complemented* if every element has a complement. *Boolean lattices* (*algebras*) are distributive and complemented lattices. In a Boolean lattice, every element has a unique complement. According to the famous Stone representation theorem, every Boolean lattice is isomorphic to the *concrete* Boolean lattice $\langle 2^S, \subseteq \rangle$ for some set S. For the lattice $\langle 2^S, \subseteq \rangle$, we have $\vee = \cup$, $\wedge = \cap$, $\top = S$ and $\bot = \emptyset$.

A *de Morgan lattice* D is a bounded distributive lattice $\langle D, \vee, \wedge, \top, \bot \rangle$ with an *involution* $\neg$ which satisfies the following *de Morgan's laws*:

$$\neg(x \wedge y) = \neg x \vee \neg y \quad \text{and} \quad \neg\neg x = x \quad \text{for all } x, y \in D.$$

It follows immediately that $\neg(x \vee y) = \neg x \wedge \neg y$, $\neg\top = \bot$ and $\neg\bot = \top$. So $\neg$ is a *dual automorphism*. Note that in a de Morgan lattice, the following laws do not always hold:

$$\neg x \vee x = \top \quad \text{and} \quad x \wedge \neg x = \bot.$$

De Morgan lattices are important for the study of the mathematical aspects of fuzzy logic [69]. The standard fuzzy algebra $F = \langle [0, 1], \max(x, y), \min(x, y), 0, 1, 1 - x \rangle$ is an example of a de Morgan lattice. Moreover, de Morgan lattices are the algebraic semantics for the non-classical formalism relevance logic [16].

2.2. Belief functions on lattices

Let $(L, \leqslant)$ be a poset having a bottom element $\bot$ and a top one $\top$ and $\mathbb{R}$ be the real field. Without further notice, every function in this paper is meant to be a real-valued map. The *Möbius function* $\mu : L^2 \to \mathbb{R}$ of L is defined recursively by

$$\mu(x, y) = \begin{cases} 1 & \text{if } x = y, \\ -\sum_{x \leqslant t < y} \mu(x, t) & \text{if } x < y, \\ 0 & \text{if } x > y. \end{cases}$$

Note that μ solely depends on L.

Proposition 2.1 *(Möbius inversion formula). (See Proposition 3.7.1 in [61].) Let P be a poset. Let f and g be two functions. Then*

$$g(t) = \sum_{s \leqslant t} f(s) \quad \text{for all } t \in P \tag{1}$$

if and only if

$$f(t) = \sum_{s \leqslant t} g(s)\mu(s, t) \quad \text{for all } t \in P, \tag{2}$$

where μ is the Möbius function of P.

The function f in the above proposition is called the *Möbius transform* of g. And the *co-Möbius transform*[2] of g, denoted by q, is defined as: for any $x \in P$,

[2] This notion was coined in [28].

$$q(x) = \sum_{y \geqslant x} f(y).$$

Definition 2.2. Given a lattice $\langle L, \leqslant \rangle$, a function f on L is called a *capacity* if it satisfies the following three conditions:

1. $f(\bot) = 0$;
2. $f(\top) = 1$;
3. $x \leqslant y$ implies $f(x) \leqslant f(y)$.

A function $bel : L \to [0, 1]$ is called a *belief function* if $bel(\top) = 1$, $bel(\bot) = 0$ and its Möbius transform m is nonnegative. m is also called the *mass function* or *mass assignment* of f. For each element $a \in L$, the quantity $m(a)$ is intended to measure the belief that one commits *exactly* to a, not the total belief that one commits to a. To obtain the measure of the total belief committed to a, one must add to $m(a)$ the quantities $m(b)$ for all elements that are strictly smaller than a:

$$bel(a) = \sum_{b \leqslant a} m(b).$$

An element $a \in L$ is called a *focal element* of L if $m(a) > 0$. The co-Möbius transform $q : L \to [0, 1]$ of bel is called the *commonality function* associated to bel.

Note that any belief function is a monotonic function by nonnegativity of m, and hence a capacity.

Example 2.3. In the above definition, if L is a Boolean algebra, then bel on L is defined in the same way as in the Dempster–Shafer theory [52]. Let Ω be a finite space. In this case, a function $m : 2^{\Omega} \to [0, 1]$ is a mass allocation function if $m(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$. A belief function on Ω is a function $bel : 2^{\Omega} \to [0, 1]$ generated by a mass allocation function as follows:

$$bel(A) := \sum_{B \subseteq A} m(B), \quad A \subseteq \Omega.$$

Note that $bel(\emptyset) = 0$ and $bel(\Omega) = 1$. The Möbius function $\mu : 2^{\Omega} \times 2^{\Omega} \to [0, 1]$ is

$$\mu(A, B) = \begin{cases} (-1)^{|B \setminus A|} & \text{if } A \subseteq B, \\ 0 & \text{otherwise.} \end{cases}$$

m is the Möbius transform of bel and is expressed as the following formula:

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} bel(B).$$

And the associated *plausibility function* pl is defined as

$$pl(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - bel(\Omega \setminus A), \quad A \subseteq \Omega.$$

Definition 2.4. Let K denote the set $\{1, 2, \ldots, k\}$. Given a lattice $\langle L, \leqslant \rangle$, a function f on L is called a *k-monotone* whenever for each $(x_1, \ldots, x_k) \in L^k$, we have

$$f\Bigl(\bigvee_{1 \leqslant i \leqslant k} x_i\Bigr) \geqslant \sum_{J \subseteq K, J \neq \emptyset} (-1)^{|J|+1} f\Bigl(\bigwedge_{j \in J} x_j\Bigr). \tag{3}$$

A capacity is *totally monotone* if it is k-monotone for every $k \geqslant 2$. A k-monotone function f is called a *k-valuation* if the above inequality degenerates into the following equality:

$$f\Bigl(\bigvee_{1 \leqslant i \leqslant k} x_i\Bigr) = \sum_{J \subseteq K, J \neq \emptyset} (-1)^{|J|+1} f\Bigl(\bigwedge_{j \in J} x_j\Bigr). \tag{4}$$

It is an *∞-valuation* if it is a k-valuation for each integer k. f is called a *probability function* if it is both a capacity and an ∞-valuation.

Lemma 2.5. *If L is distributive, then the following equality is sufficient for an ∞-valuation f:*

$$f(a \wedge b) + f(a \vee b) = f(a) + f(b) \quad \textit{for all } a, b \in L. \tag{5}$$

Proof. We prove by induction on the number of disjuncts.

- Base case: $k = 1, 2$. The equality (4) for ∞-valuation follows immediately from (5).
- Induction case: Assume that the equality (4) holds for $k = n$. Now we need to show that it holds for $k = n + 1$.

$$\begin{aligned}
f\Big(\bigvee_{1\leqslant i\leqslant n+1} x_i\Big) &= f\Big(\bigvee_{1\leqslant i\leqslant n} x_i \vee x_{n+1}\Big)\\
&= f\Big(\bigvee_{1\leqslant i\leqslant n} x_i\Big) + f(x_{n+1}) - f\Big(\bigvee_{1\leqslant i\leqslant n} (x_{n+1}\wedge x_i)\Big) \quad (\text{according to (5)})\\
&= \sum_{\emptyset\neq J\subseteq\{1,\ldots,n\}} (-1)^{|J|+1} f\Big(\bigwedge_{j\in J} x_j\Big) + f(x_{n+1}) - f\Big(\bigvee_{1\leqslant k\leqslant n} (x_{n+1}\wedge x_i)\Big)\\
&= \sum_{\emptyset\neq J\subseteq\{1,\ldots,n\}} (-1)^{|J|+1} f\Big(\bigwedge_{j\in J} x_j\Big) + f(x_{n+1}) - \sum_{\emptyset\neq J\subseteq\{1,\ldots,n\}} (-1)^{|J|+1} f\Big(x_{n+1}\wedge\bigwedge_{j\in J} x_j\Big)\\
&= f(x_1) + \cdots + f(x_n) + f(x_{n+1})\\
&\quad + \Big[\sum_{\substack{J\subseteq\{1,\ldots,n\}\\ |J|\geqslant 2}} (-1)^{|J|+1} f\Big(\bigwedge_{j\in J} x_j\Big) + \sum_{\substack{n+1\in J\subseteq\{1,\ldots,n+1\}\\ |J|\geqslant 2}} (-1)^{|J|+1} f\Big(\bigwedge_{j\in J} x_j\Big)\Big]\\
&= \sum_{\emptyset\neq J\subseteq\{1,\ldots,n,n+1\}} (-1)^{|J|+1} f\Big(\bigwedge_{j\in J} x_j\Big).
\end{aligned}$$

Note that both the third and fourth equalities follow from the Induction Hypothesis.

So we have finished the proof that the equality (4) holds. □

Before we show Theorem 2.8, we need a lemma, which is obvious in measure theory and follows directly from the above lemma. Define a function $\mu : \mathcal{P}(X) \to \mathbb{R}$ on the power set $\mathcal{P}(X)$ of an arbitrary finite set X to be *additive* if $\mu(\emptyset) = 0$, and $\mu(A) + \mu(B) = \mu(A \cap B) + \mu(A \cup B)$.

Lemma 2.6. *If μ is additive, then*

$$\mu\Big(\bigcup_{i=1}^{n} R_i\Big) = \sum_{\emptyset\neq J\subseteq\{1,\ldots,n\}} (-1)^{|J|+1} \mu\Big(\bigcap_{i\in J} R_i\Big). \tag{6}$$

Proof. The lemma follows from Lemma 2.5 and the fact that $\mathcal{P}(X)$ with the usual set operations is a Boolean algebra and hence is a distributive lattice. This is also well-known as the *principle of inclusion and exclusion* in probability theory (formula (2.9) on page 24 in [8]). □

The following proposition [4] tells us that every belief function is totally monotone.

Proposition 2.7. *Let $f : L \to [0, 1]$ be a capacity and m be its Möbius transform. If f is a belief function, then it is totally monotone.*

Shafer proved that the converse is also true for any belief function on *Boolean algebras* (Theorem 2.1 in [52]). We show that actually it holds *generally* for any lattice, which answers an open question raised in [26].

Theorem 2.8. *Let L be a lattice and $f : L \to [0, 1]$ be a capacity on L and m be its Möbius transform. The following two statements are equivalent*:

- *m is nonnegative*;
- *f is totally monotone.*

The essential part of the following proof is based on a structurally inductive construction of the Möbius transform m from f: $m(a) = f(a) - \sum_{b<a} m(b)$.

Proof. Here we only need to show that, if f is totally monotone, then its Möbius transform m is nonnegative, since the other direction is simply Proposition 2.7. That is to say, any totally monotone capacity on a lattice is a belief function. Given a totally monotone capacity f on L, we are computing inductively $m : L \to \mathbb{R}$ as follows:

- For the minimal element $\perp$, $m(\perp) = 0$;
- For any non-minimal element $a \in L$,

$$m(a) = f(a) - \sum_{b<a} m(b).$$

Note that the range of m is $\mathbb{R}$. It is easy to see that, for such defined m, $f(a) = \sum_{b \leqslant a} m(b)$ for all $a \in L$. It remains to show the following claim:

Claim 1. *$m(a) \geqslant 0$ for all $a \in L$.*

It is easy to see that, when a is the minimal element, $m(a) \geqslant 0$. Now assume that a is a non-minimal element. We need to show that $m(a) \geqslant 0$. Let $A := \{x\colon\ x \in L, x < a\}$. For each $b < a$, I_b denotes the set $\{x\colon\ x \in L, x \leqslant b\}$. It is easy see that $A = \bigcup_{b<a} I_b$. So

$$\begin{aligned}
\sum_{x\in A} m(x) &= \sum_{J\subseteq A,\, J\neq\emptyset} (-1)^{|J|+1} \Bigg(\sum_{x\in(\bigcap_{b\in J} I_b)} m(x) \Bigg) \quad \text{(Lemma 2.6)}\\
&= \sum_{J\subseteq A,\, J\neq\emptyset} (-1)^{|J|+1} \Bigg(\sum_{x\in(I_{\bigwedge\{x:\ x\in J\}})} m(x) \Bigg)\\
&= \sum_{J\subseteq A,\ J\neq\emptyset} (-1)^{|J|+1} f\Big(\bigwedge\{x\colon\ x \in J\} \Big)\\
&= \sum_{J\subseteq A,\ J\neq\emptyset} (-1)^{|J|+1} f\Big(\bigwedge_{x\in J} x \Big).
\end{aligned}$$

The first equality comes from Lemma 2.6 if we define an additive set function $\mu(X) := \sum_{x\in X} m(x)$ for all $X \subseteq L$. In other words, $\sum_{b\in\{x\in L:\ x<a\}} m(b) = \sum_{J\subseteq\{x\in L:\ x<a\}, J\neq\emptyset} (-1)^{|J|+1} f(\bigwedge_{x\in J} x)$. Moreover, since f is a totally monotonic capacity,

$$\begin{aligned}
f(a) &\geqslant f\Big(\bigvee_{b<a} b \Big)\\
&\geqslant \sum_{\emptyset\neq J\subseteq\{x\in L:\ x<a\}} (-1)^{|J|+1} f\Big(\bigwedge_{x\in J} x \Big)\\
&= \sum_{b\in\{x\in L:\ x<a\}} m(b).
\end{aligned}$$

Note that the second inequality comes from the total monotonicity of f. It follows that

$$m(a) = f(a) - \sum_{b\in\{x\in L:\ x<a\}} m(b) \geqslant 0. \qquad \square$$

So any probability function is also a belief function. From Proposition A.1 in Appendix A, we know that a belief function can be defined on any lattice while probability functions can live only on distributive lattices. In Example 2 in [27], Grabisch constructed a "counterexample" to show that Theorem 2.8 does not hold. But, his counterexample does not work because 6-monotonicity there is equivalent to $1 \geqslant 6\beta - 6\alpha$ not to $1 \geqslant 6\beta - 9\alpha$ as provided by Grabisch.

Example 2.9. Here we adapt an example of belief functions on set-valued variables [13] from [12] to illustrate the theory in this section. Let $\Theta = \{a, b, c, d\}$ be the set of possible faults of a given system. For any $A, B \subseteq \Theta$, $[A, B]$ denotes the interval $\{C \subseteq \Theta\colon\ A \subseteq C \subseteq B\}$. The interval means that the faults in B may be present but those in A are definitely present in the system. Let Ω be the set of all these intervals in Θ. Ω is a lattice associated with the following two operations: for any two $[A_1, A_2], [B_1, B_2] \in \Omega$,

$$[A_1, A_2] \wedge [B_1, B_2] = \begin{cases} [A_1 \cup B_1, A_2 \cup B_2] & \text{if } A_1 \cup B_1 \subseteq A_2 \cap B_2,\\ \emptyset_\Omega & \text{otherwise.} \end{cases}$$

and

$$[A_1, A_2] \vee [B_1, B_2] = [A_1 \cap B_1, A_2 \cup B_2]$$

where $\emptyset_\Omega$ is the empty set of Ω while $\emptyset_\Theta$ denotes the empty set of Θ.

From different pieces of evidence, we may derive the representation of degrees of beliefs in different events as the following mass assignment:

$$m\big(\big[\{a,c\},\{a,b,c\}\big]\big) = 0.56, \qquad m\big(\big[\{c\},\{a,b,c\}\big]\big) = 0.24,$$
$$m\big(\big[\{a,c\},\{a,c\}\big]\big) = 0.14, \qquad m\big(\big[\{d\},\{a,c,d\}\big]\big) = 0.06.$$

It is easy to check that $\langle \Omega, \wedge, \vee \rangle$ is a lattice but not distributive and even the focal elements don't satisfy the law of distributivity. For example,

$$\big[\{a,c\},\{a,b,c\}\big] \vee \big(\big[\{c\},\{a,b,c\}\big] \wedge \big[\{d\},\{a,c,d\}\big]\big)$$
$$\neq \big(\big[\{a,c\},\{a,b,c\}\big] \vee \big[\{c\},\{a,b,c\}\big]\big) \wedge \big(\big[\{a,c\},\{a,b,c\}\big] \vee \big[\{d\},\{a,c,d\}\big]\big).$$

From the above mass assignment, we may derive the degree of belief in the proposition that fault d is not present:

$$bel\big(\big[\emptyset_\Theta, \overline{\{d\}}\big]\big) = bel\big(\big[\emptyset_\Theta, \{a,b,c\}\big]\big) = 0.56 + 0.24 + 0.14 = 0.94.$$

3. Belief functions on distributive lattices

In this section, we study belief functions on distributive lattice from two perspectives. The first one treats mass functions on distributive lattices as the Möbius transforms of their corresponding belief functions, and characterizes belief functions without reference to mass assignments. A more general result is that there is a one-to-one correspondence between belief functions on a distributive lattice L and those on its generated Boolean algebra, which is the smallest Boolean algebra of which L is a sublattice. We rely on the techniques from Birkhoff's representation theorem for finite distributive lattices (Theorem A.2 in Appendix A). The second one regards belief functions as generalized probability functions. More specifically, we investigate the condition when a belief function is Bayesian or a probability function and the perspective from which belief functions can be regarded as inner probability functions. So, in the general setting of distributive lattices, a key part of the theory of belief functions is firmly rooted in probability theory. The immediate payoff to this view is a logic of reasoning about belief functions, which will be provided in Section 5.

3.1. Möbius transforms of belief functions

Given a poset P, $J(P)$ denotes the lattice of order ideals of P with the ordinary union and intersection (as subsets of P). So $J(P)$ is distributive. Conversely, from Theorem A.2 (Birkhoff's fundamental theorem for finite distributive lattices) in Appendix A, we know that for any finite distributive lattice L, there is a unique (up to isomorphism) finite poset P for which $L \cong J(P)$. Usually P is chosen to be the poset of join-irreducibles in L. Note that all valid propositions for $J(P)$ are *dually* valid for $F(P)$, which denotes the lattice of filters of P with the ordinary union and intersection. If we replace in this section all posets $\langle P, \leqslant \rangle$ by its dual $\langle P, \leqslant^\partial \rangle$ (where $x \leqslant^\partial y$ iff $y \leqslant x$) and all $\leqslant$-order ideals by $\leqslant^\partial$ filters, the dual forms of these propositions remain valid.

The following two propositions provide formulas for Möbius functions and Möbius transforms in distributive lattices.

Proposition 3.1. *(See Example 3.9.6 in [61].) The Möbius function of the distributive lattice $L = J(P)$ where P is a poset is: for any $I, I' \in J(P)$,*

$$\mu\big(I, I'\big) = \begin{cases} (-1)^{|I' \setminus I|} & \text{if } [I, I'] \text{ is a Boolean algebra}, \\ 0 & \text{otherwise}, \end{cases}$$

where $[I, I']$ denotes the interval $\{K \in J(P): I \subseteq K \subseteq I'\}$.

From this proposition, we immediately obtain a nice formula for Möbius transforms.

Theorem 3.2. *Let $L = J(P)$ be a distributive lattice for some poset P. Suppose $Bel : L \to [0,1]$ is the belief function given by the mass assignment $m : L \to [0,1]$. Then*

$$m(A) = \sum_{\substack{[B,A] \text{ is a} \\ \text{Boolean algebra}}} (-1)^{|A \setminus B|} Bel(B)$$

for all $A \in J(P)$.

In order to compare the following proposition to Theorem 2.1 in [52], we prove the following proposition from the above two theorems about Möbius transforms along the same line as Shafer's proof of Theorem 2.1 in [52]. The relationship to Shafer's original proof on Boolean algebras can be made through the following fact (Section 3.4. in [61]): for any two elements $I, I' \in J(P)$ for some poset P, $[I, I']$ is a Boolean algebra iff $I' \setminus I$ is an antichain.

The following theorem is also the distributive version of Theorem 2.8.

Theorem 3.3. *Given a distributive lattice $L = J(P)$, a monotonic function $Bel : L \to [0, 1]$ is a belief function iff it satisfies the following conditions*:

1. $Bel(\bot) = 0$;
2. $Bel(\top) = 1$;
3. *For every positive number n and every collection $A_1, \ldots, A_n$ of elements of L,*

$$Bel(A_1 \cup \cdots \cup A_n) \geqslant \sum_{\substack{I \subseteq \{1,\ldots,n\},\\ I \neq \emptyset}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right).$$

Proof. The left-to-right direction is simply Proposition 2.7. Now we mainly focus on the right-to-left direction. Assume that Bel satisfies the above three conditions and, for any $A \in L$, $Bel(A) = \sum_{B \subseteq A} m(B)$. According to Theorem 3.2, we know its Möbius transform:

$$m(A) = \sum_{\substack{[B,A] \text{ is a}\\ \text{Boolean algebra}}} (-1)^{|A \setminus B|} Bel(B).$$

In order to show that Bel is a belief function, it suffices to show that $m(A) \geqslant 0$ for all $A \in L = J(P)$. Give an element $A \in J(P)$, define A_{max} to be the set of maximal elements of A in P. It is easy to see that

- $A \setminus A_{max}$ is an order ideal in P and hence is an element of $J(P)$ and
- A_{max} is an antichain in P.

Moreover, A_{max} is the maximal subset B of A such that

- $A \setminus B$ is an order ideal in P and hence is an element of $J(P)$ and
- B is an antichain in P.

That is to say, if B is an order ideal in P and $A \setminus B$ is an antichain in P, then $A \setminus B \subseteq A_{max}$. Suppose that $A_{max} = \{\theta_1, \ldots, \theta_n\}$. If $n = 1$ and $A_{max} = \{a_{max}\}$, then

$$\begin{aligned} m(A) &= \sum_{\substack{[B,A] \text{ is a}\\ \text{Boolean algebra}}} (-1)^{|A \setminus B|} Bel(B) \\ &= Bel(A) + \sum_{\substack{A \setminus B \subseteq A_{max}\\ A \setminus B \neq \emptyset}} (-1)^{|A \setminus B|} Bel(B) \\ &= Bel(A) - Bel\big(A \setminus \{a_{max}\}\big) \\ &\geqslant 0. \end{aligned}$$

The second equality comes from the fact immediately before this theorem and the last inequality follows from the monotonicity. Without loss of generality, we assume that $n \geqslant 2$. Let $K = \{1, 2, \ldots, n\}$ and $A_i = A \setminus \{\theta_i\}$. If $A \setminus B = \{\theta_{i_1}, \ldots, \theta_{i_k}\}$, then $B = A_{i_1} \cap \cdots \cap A_{i_k}$. It is easy to see that, for any two different A_i's, their union is equal to A. It follows from Theorem 3.2 that

$$\begin{aligned} m(A) &= Bel(A) - \sum_{\substack{I \subseteq K,\\ I \neq \emptyset}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right) \\ &\geqslant 0. \end{aligned}$$

Similarly the equality follows from the fact that $[B, A]$ is a Boolean algebra iff $A \setminus B$ is an antichain. And the last inequality follows from the third condition of total monotonicity in the assumption. □

This theorem gives another characterization of belief functions on distributive lattices *without* reference to mass functions. So it plays an important role in our axiomatization of reasoning about belief functions. If L is Boolean, then the formulas in Proposition 3.1 and Theorem 3.2 are the same as those corresponding formulas in Example 2.3.

3.2. Belief functions on distributive lattices and Boolean algebras

In this subsection, we discuss some direct connections between belief functions on distributive lattices to those on Boolean algebras. Given a distributive lattice $L = J(P)$ for some poset P, we know that L as a lattice is a sublattice of the powerset of P with the usual set operations, which is a Boolean algebra. So, for any distributive lattice L, the set of Boolean algebras of which L is a sublattice is not empty, i.e., $\{B\colon\ L$ is a sublattice of B and B is a Boolean algebra$\} \neq \emptyset$. Also the intersection of any finite number of Boolean algebras is a Boolean algebra.

Definition 3.4. A Boolean algebra B_L is *generated* by the distributive lattice L if $B_L = \bigcap\{B\colon\ L$ is a sublattice of B and B is a Boolean algebra$\}$.

If $L = J(P)$ is a distributive lattice for some poset P, then the Boolean algebra B_L generated by L is the powerset of P with the usual set operations (Theorem 2.1. in [41]). A more general result than this lemma but in a different form follows from Theorem 1 in [45]. For any two Boolean algebras B_1 and B_2, $B_1 \preccurlyeq B_2$ iff $|B_1| \leqslant |B_2|$. So, for any distributive lattice L, the generated Boolean algebra B_L is the *smallest* Boolean algebra into which L can be embedded.

Lemma 3.5. *Let L be a finite distributive lattice and B_L be the Boolean algebra generated by L. If μ is a probability function on L, then μ has a unique extension of probability function on B_L. Conversely, if μ is a probability function on B_L, then the restriction of μ into L is also a probability function on L.*

Proof. Since the second part is clear, we focus on the first one. Assume that $L = J(P)$ for some poset P and μ is a probability function on L. According to the above discussion, we know that 2^P with the usual set operations is the Boolean algebra generated by $J(P)$. In order to extend the probability function on $L = J(P)$ to 2^P, it suffices to know probabilities on singletons. For any $x \in P$, define $\bar{\mu}(\{x\}) := \mu((x]) - \mu((x] \setminus \{x\})$ where $(x] := \{y \in L\colon\ y \leqslant x\}$. Note that the last two probabilities are well-defined because both $(x]$ and $(x] \setminus \{x\}$ are elements of $J(P)$. Probabilities of all other subsets of P can be obtained easily from the additivity property. Since the extended probability function $\bar{\mu}$ on B_L totally depends on the function μ on L, this extension is unique. □

In the following, we use $\bar{\mu}$ to denote this unique extension on B_L for any probability function μ on L. So there is a one-to-one correspondence between probability functions on a distributive lattice and those on its generated Boolean algebra. Surprisingly, this property is also shared by belief functions.

Lemma 3.6. *Let L be a finite distributive lattice and B_L be the Boolean algebra generated by L. For any belief function Bel on B_L, the restriction $Bel\restriction_L$ of Bel to L is a belief function on L.*

Proof. The proof follows from Theorem 3.3. It is easy to check that $Bel\restriction_L$ satisfies the following properties:

1. $Bel\restriction_L(\bot) = 0$;
2. $Bel\restriction_L(\top) = 1$;
3. For every positive number n and every collection $x_1, \ldots, x_n$ of elements of L,

$$Bel\restriction_L(x_1 \vee \cdots \vee x_n) \geqslant \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} Bel\restriction_L\Big(\bigwedge_{i \in I} x_i\Big).$$

These properties are inherited from the generated Boolean algebra B_L. □

Lemma 3.7. *Let L be a finite distributive lattice and B_L be the Boolean algebra generated by L. Any belief function Bel on L can be extended to a belief function Bel^* on B_L in the sense that, for any $x \in L$, $Bel^*(x) = Bel(x)$.*

Proof. Here we will appeal to the fundamental theorem for belief functions on distributive lattices Theorem 3.3. For a belief function Bel on L, there is a mass assignment m on L such that, for any $x \in L$, $m(x) \geqslant 0$ and $Bel(x) = \sum_{y \leqslant x} m(y)$. Now we define a mass assignment m^* on B_L which is an extension of m on L. Define

$$m^*(x) = \begin{cases} m(x) & \text{if } x \in L, \\ 0 & \text{if } x \notin L. \end{cases}$$

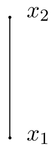

Fig. 1. Hasse diagram of P_0.

According to Theorem 3.3, there is a belief function Bel^* on B_L: $Bel^*(x) := \sum_{y \leqslant x} m^*(y)$ which is an extension of Bel on L in the sense that, for any element $x \in L$, $Bel(x) = Bel^*(x)$. $\square$

However, the extension of belief functions on distributive lattice to their generated Boolean algebras are generally not unique. Consider a simple poset $P_0 = \langle \{x_1, x_2\}, \{(x_1, x_1), (x_2, x_2), (x_1, x_2)\} \rangle$. The Hasse diagram of this poset is illustrated in Fig. 1.

Now we define a belief function Bel on $L_0 := J(P_0)$ as follows: $m(\{x_1\}) = \frac{1}{2} = m(\{x_1, x_2\})$. So $Bel(\{x_1\}) = \frac{1}{2}$, and $Bel(\{x_1, x_2\}) = 1$. In addition to the type of extension Bel^* in Lemma 3.7, there is at least another *different* extension Bel' on the generated Boolean algebra B_{L_0} with the following mass assignment m': $m'(\{x_1\}) = \frac{1}{2}, m'(\{x_2\}) = \frac{1}{4}, m'(\{x_1, x_2\}) = \frac{1}{4}$.

But, in the following, we show that Bel^* in Lemma 3.7 is *unique* from the perspective viewing belief functions as *lower envelopes*. For any belief function Bel on a distributive lattice L, we define

$$\mathcal{P}_{Bel} = \{\mu\colon\ \mu \text{ is a probability function on } L \text{ and } \mu \geqslant Bel\}$$

where $\mu \geqslant Bel$ means that, for all $x \in L$, $\mu(x) \geqslant Bel(x)$.

Lemma 3.8. *For any $x \in L$, $Bel(x) = \inf\{\mu(x)\colon\ \mu \in \mathcal{P}_{Bel}\}$.*

Proof. It suffices to show that, for each $x \in L$, there is a probability function $\mu \in \mathcal{P}_{Bel}$ such that $\mu(x) = Bel(x)$. But this follows directly from Lemmas 3.5 and 3.7 and Dempster's similar result on Boolean algebras [10]. Given $x \in L$ and a belief function Bel on L, we consider x as an element in B_L and Bel as Bel^* on B_L. According to Dempster's result on Boolean algebras [10], we know that there is a probability function $\mu' \in \mathcal{P}_{Bel^*}$ (i.e., for every element $a \in B_L$, $\mu'(a) \geqslant Bel^*(a)$) such that $\mu'(x) = Bel^*(x)$. It follows that $\mu' \upharpoonright_L (x) = Bel(x)$. $\square$

Recall that $\bar{\mu}$ denotes the unique extension of a probability function μ on L to B_L (Lemma 3.5). The extension Bel^* defined in Lemma 3.7 is *unique* in the following sense.

Theorem 3.9. *For any $x \in B_L$, $Bel^*(x) = \inf\{\bar{\mu}(x)\colon\ \mu \in \mathcal{P}_{Bel}\}$.*

Proof. Without loss of generality, we assume that $L = J(P)$ and $B_L = 2^P$ for some poset P. From the above lemma, we know that, for any $A \in L$, $Bel^*(A) = \inf\{\bar{\mu}(A)\colon\ \mu \in \mathcal{P}_{Bel}\}$. It suffices to show the following claim:

Claim 2. *For any $A \in B_L \setminus L$, there is a probability function $\mu \in \mathcal{P}_{Bel}$ such that $\bar{\mu}(A) = Bel^*(A)$.*

Denote $A_* = \bigcup\{I \in J(P)\colon\ I \subseteq A\}$, $A^* = \bigcap\{I \in J(P)\colon\ I \supseteq A\}$ and $A_- = A^* \setminus A$. Note that $A_* \in J(P), A^* \in J(P)$, $A_* \subsetneq A \subsetneq A^*$ and hence $A_- \neq \emptyset$. It follows from Lemma 3.8 that there is a probability function $\mu \in \mathcal{P}_{Bel}$ on $J(P)$ such that $Bel^*(A) = Bel(A_*) = \mu(A_*)$.

If $\bar{\mu}(A) = Bel^*(A)(= \mu(A_*))$, then we are done. If $\bar{\mu}(A) > Bel^*(A)(= \bar{\mu}(A_*) = \mu(A_*))$, then we need to construct a new probability function μ_A such that $\mu_A \upharpoonright_L \in \mathcal{P}_{Bel}$ and $\mu_A(A) = \mu(A_*)$ in order to prove the claim. Recall that, for each $a \in P$, $(a]$ denotes the set $\{b \in P\colon\ b \leqslant a\}$. For each $a \in A \setminus A_*$, $(a] \cap A_- \neq \emptyset$. For, otherwise, since $(a] \subseteq A^*$, $A_* \subsetneq (a] \cup A_* \subseteq A$, which contradicts the fact that A_* is the largest element in L that is a subset of A. Let f_A denote a function from $A \setminus A_*$ to A_- such that, for each $a \in A \setminus A_*$, $f_A(a) \in (a] \cap A_-$.

Now we transpose the mass of $\bar{\mu}$ at each $a \in A \setminus A_*$ to $f_A(a)$ in $(a] \cap A_-$. We define a new probability function μ_A on B_L with the mass assignment as follows: for any $a \in P$,

$$\mu_A(\{a\}) = \begin{cases} \bar{\mu}(\{a\}) & \text{if } a \in A_* \cup (P \setminus A^*), \\ 0 & \text{if } a \in A \setminus A_*, \\ \bar{\mu}(\{a\}) + \sum_{\substack{f_A(b)=a \\ b \in A \setminus A_*}} \bar{\mu}(\{b\}) & \text{if } A_- \ni a = f_A(b) \text{ for some } b \in A \setminus A_*, \\ \bar{\mu}(\{a\}) & \text{otherwise.} \end{cases}$$

The transposition defined in μ_A can be illustrated in Fig. 2.

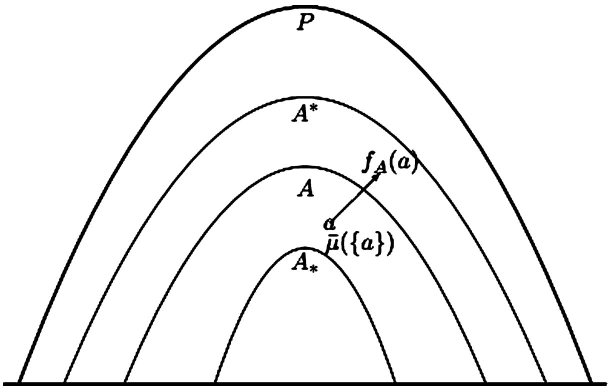

Fig. 2. The transposition defined in μ_A.

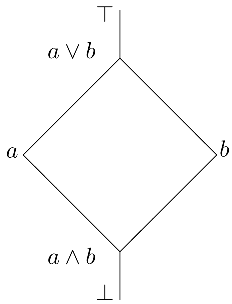

Fig. 3. Hasse diagram of a distributive lattice L_0 as a counterexample.

It is easy to see that $\sum_{a \in P} \mu_A(\{a\}) = 1$ and hence it defines a probability function on B_L, which is still denoted as μ_A. Note that $\mu_A(A) = \bar{\mu}(A_*) = Bel^*(A)$ and the only elements in P where the masses of μ_A may be smaller than those of $\bar{\mu}$ are those elements in $A \setminus A_*$. In order to show that μ_A is the probability function that we want, we only need to show that, for all $I \in J(P)$, $\mu_A(I) \geqslant Bel(I)$. For every $I \in J(P)$, if $a \in I$ and $a \in A \setminus A_*$, then $f_A(a) \in (a] \subseteq I$. So

$$\sum_{a \in I, a \in A \setminus A_*} \bigl(\mu_A\bigl(\{a\}\bigr) + \mu_A\bigl(\{f_A(a)\}\bigr)\bigr) = \sum_{a \in I, a \in A \setminus A_*} \bigl(\bar{\mu}\bigl(\{a\}\bigr) + \bar{\mu}\bigl(\{f_A(a)\}\bigr)\bigr).$$

This equality implies that, for each $I \in J(P)$, $\mu_A(I) \geqslant \bar{\mu}(I)$. Since $\mu \in \mathcal{P}_{Bel}$ and $\bar{\mu}$ is the extension of μ, $\mu_A(I) \geqslant Bel(I)$. So we have shown that $\mu_A(A) = Bel^*(A)$ and $\mu_A \restriction_L \in \mathcal{P}_{Bel}$. $\square$

3.3. *Bayesian belief functions*

In this part, we investigate *Bayesian* belief functions and the condition in the setting of distributive lattices when a belief function is Bayesian. Shafer showed that a belief function on a Boolean algebra is Bayesian if and only if all of its focal elements are singletons (Theorem 2.8 in [52]). We have a similar property for distributive lattices. This property can be summarized in a simple comparison: join-irreducibles to distributive lattices are the same as singletons to Boolean algebras.

Definition 3.10. Given a distributive lattice $L = \langle L, \leqslant \rangle$, a belief function $Bel : L \to [0, 1]$ is called *Bayesian* if,

$$Bel(a \vee b) + Bel(a \wedge b) = Bel(a) + Bel(b) \quad \text{whenever } a, b \in L. \tag{7}$$

From Lemma 2.5, for any distributive lattice $L = \langle L, \leqslant \rangle$, a belief function $Bel : L \to [0, 1]$ is Bayesian if and only if it is a probability function.

Remark 3.11. The condition (7) can't be replaced with another simpler but weaker *additivity* condition:

$$Bel(a \vee b) = Bel(a) + Bel(b) \quad \text{whenever } a, b \in L \text{ such that } a \wedge b = \perp. \tag{8}$$

The counterexample is illustrated in Fig. 3.

It is easy to see that the above lattice L_0 is a distributive lattice. Define a function $Bel : L_0 \to [0,1]$ as follows:

$$Bel(\bot) = 0, \qquad Bel(a) = \frac{2}{5} = Bel(b), \qquad Bel(a \wedge b) = \frac{1}{5}, \qquad Bel(a \vee b) = \frac{4}{5}, \qquad Bel(\top) = 1.$$

The condition (8) hold vacuously for the above defined *Bel*. However, $Bel(a \vee b) + Bel(a \wedge b) \neq Bel(a) + Bel(b)$. This implies that the condition (7) does not hold for *Bel*.

Definition 3.12. Let *Bel* be a belief function on a distributive lattice L with the set I_L of join-irreducibles in L. The *inner belief function* Bel° of *Bel* is defined as follows:

$$Bel^\circ(a) = \sum_{x \in I_L,\ x \leqslant a} m(x).$$

In other words, $Bel^\circ(a)$ is the sum of all mass assignments on join-irreducibles which are less than or equal to a. So $Bel^\circ(a) \leqslant Bel(a)$.

Lemma 3.13. *Let Bel be a belief function on a distributive lattice L with the set I_L of join-irreducibles in L. Then the above defined inner belief function Bel° is an ∞-valuation on L.*

Proof. Let *Bel* be a belief function on a distributive lattice L with the set I_L of join-irreducibles in L and Bel° be the corresponding inner belief function. For any elements $a_1, \dots, a_n \in L$,

$$\begin{aligned}
Bel^\circ(a_1 \vee \cdots \vee a_n) &= \sum_{\substack{x \text{ is join-irreducible} \\ x \leqslant \bigvee_{i \in \{1,\dots,n\}} a_i}} m(x) \\
&= \sum_{\substack{x \text{ is join-irreducible} \\ x \leqslant a_i \text{ at least for some } i\ (1 \leqslant i \leqslant n)}} m(x) \\
&= \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} (-1)^{|I|+1} \Bigg[\sum_{\substack{x \text{ is join-irreducible} \\ x \leqslant \bigwedge_{i \in I} a_i}} m(x) \Bigg] \\
&= \sum_{\emptyset \neq I \subseteq \{1,\dots,n\}} (-1)^{|I|+1} Bel^\circ\Bigg(\bigwedge_{i \in I} a_i \Bigg).
\end{aligned}$$

The second equality follows from the fact that, for elements $x, x_1, \dots, x_n$ in a distributive lattice, if $x \leqslant \bigvee_{1 \leqslant i \leqslant n} x_i$ and x is join-irreducible, then $x \leqslant x_i$ at least for some x_i $(1 \leqslant i \leqslant n)$. And the third one comes from Lemma 2.6 if we define an additive set function $\mu(X) := \sum_{x \in X} m(x)$ for all $X \subseteq L$. For example, $\mu(\{x \in L\colon x \text{ is join-irreducible and } x \leqslant a_i\}) = \sum_{\substack{x \text{ is join-irreducible} \\ x \leqslant a_i}} m(x)$.

So indeed Bel° is an ∞-valuation on L. $\square$

Note that Bel° is not necessarily a probability function on L because $Bel^\circ(\top)$ may not be equal to 1. The following theorem is a partial converse to this lemma and is also the distributive counterpart of Lemma 2.7 in [52].

Theorem 3.14. *A belief function Bel on a distributive lattice $L = J(P)$ where P is a poset is Bayesian iff its corresponding mass assignment m is given by $m(A') = Bel(A')$ for any minimal join-irreducibles A' in $J(P)$ and $m(A) = 0$ for all join-reducibles $A \in J(P)$.*

Proof. Let *Bel* be a belief function on a distributive lattice $L = J(P)$ where P is a poset and m be its corresponding mass assignment. First note that

1. $I \in J(P)$ is join-irreducible in the lattice $J(P)$ iff I is a principal order ideal in P;
2. Minimal join-irreducible elements are atoms, hence singletons. And A' is a minimal join-irreducible in $J(P)$ iff $A' = \{\theta\}$ for some minimal element $\theta \in P$.

Assume that *Bel* is also a probability function on $J(P)$.

Claim 3. *For any minimal element $\theta \in P$, $Bel(\{\theta\}) = m(\{\theta\})$.*

It follows directly from the fact that, for any minimal element $\theta \in P$, $\{\theta\}$ is a singleton principal order ideal of P, i.e., $\{\theta\} \in J(P)$ that $Bel(\{\theta\}) = m(\{\theta\})$. Next we show

Claim 4. *$m(A) = 0$ for all join-reducibles A in L, which are non-principal order ideals of P.*

We prove this by induction on the rank of A, which is actually the number $|A|$ of elements in A (according to Theorem A.4 in Appendix A). Note that the lowest rank of a join-reducible element in L is 2.

- Base case: $rank(A) = 2$ and A is a non-principal order ideal of P. It follows that $A = \{a_1, a_2\}$ for some two distinct minimal elements a_1 and a_2 in P. Since Bel is a probability function, $Bel(\{a_1, a_2\}) = Bel(\{a_1\}) + Bel(\{a_2\})$. On the other hand, since Bel is a belief function, $Bel(\{a_1, a_2\}) = m(\{a_1, a_2\}) + m(\{a_1\}) + m(\{a_2\})$. From the above Claim 3, it follows that $m(\{a_1, a_2\}) = 0$.
- Induction case: Suppose that $m(A') = 0$ for all non-principal order ideals A' of rank $\leqslant k$. Given a non-principal order ideal A of rank $k + 1$, we need to show that $m(A) = 0$. First note that, for any order ideal B (not necessarily non-principal) of rank $\leqslant k$,

$$\begin{aligned}
Bel(B) &= \sum_{C \subseteq B, C \in J(P)} m(C) \\
&= \sum_{\substack{C \subseteq B,\ C \in J(P) \\ C \text{ is join-irreducible}}} m(C) \quad \text{(from the induction assumption)} \\
&= Bel^{\circ}(B).
\end{aligned} \tag{9}$$

Since A is a non-principal order ideal in P, it must be a union of some different principal order ideals $(a_1], \ldots, (a_n]$ for some $a_1, \ldots, a_n \in P$ where $n \geqslant 2$ and $(a_i] = \{x \in L\colon\ x \leqslant a_i\} (1 \leqslant i \leqslant n)$.

$$\begin{aligned}
Bel(A) &= m(A) + \sum_{B \subsetneq A, B \in J(P)} m(B) \\
&= m(A) + \sum_{\substack{B \subsetneq A \\ B \text{ is a principal order ideal}}} m(B) \quad \text{(Induction hypothesis)} \\
&= m(A) + \sum_{\substack{B \subseteq A \\ B \text{ is a principal order ideal}}} m(B) \quad (A \text{ is non-principal}) \\
&= m(A) + \sum_{\substack{B \subseteq A \\ B \text{ is join-irreducible in } J(P)}} m(B) \quad \text{(Note at the beginning of the proof)} \\
&= m(A) + Bel^{\circ}(A) \quad \text{(Definition 3.12)} \\
&= m(A) + \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} Bel^{\circ}\left(\bigcap_{i \in I} (a_i]\right) \quad \text{(Lemma 3.13)} \\
&= m(A) + \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} (a_i]\right) \quad \left(\text{(9) and } rank\left(\bigcap_{i \in I} (a_i]\right) \leqslant k\right) \\
&= m(A) + Bel(A) \quad \text{(the assumption that } Bel \text{ is a probability function).}
\end{aligned}$$

It follows that $m(A) = 0$. So we have finished the induction case and hence the claim.

By summarizing the above inductive proof, we prove the left-to-right direction. Now we go for the other direction.

Assume that the mass assignment m that corresponds to Bel is given by $m(\{\theta\}) = Bel(\{\theta\})$ for any minimal elements θ in P and $m(A) = 0$ for all non-principal order ideals $A \in J(P)$. It immediately follows that, for any $A' \in J(P)$, $Bel(A') = Bel^{\circ}(A')$. So indeed Bel is an ∞-valuation and hence a probability function on $J(P)$. □

Corollary 3.15. *A belief function Bel on a distributive lattice is Bayesian (or a probability function) iff all its focal elements are join-irreducibles.*

So a Bayesian belief function (or probability function) μ on a distributive lattice $L = J(P)$ for some poset P is more like a point function than a set function in its level of complexity in the sense that μ is determined by its values at *elements of*

P (which correspond to principal order ideals in P) rather by its values at all *order ideals* of P (all values at non-principal order ideals are 0). This implies that generally probability functions are simpler and easier to describe than belief functions.

The Bayesian theory adopts a rule of conditioning for updating beliefs, which is a special case of Dempster's rule of combination in the Dempster–Shafer theory. Since this rule is related to another view of belief functions as evidence, we choose to leave it for another occasion.

3.4. *Belief functions as inner probability functions*

We have provided the condition when a belief function is Bayesian. In the remainder of this section, we explore one perspective from which a belief function is regarded as a *generalized* probability function in the sense that a belief function can be regarded as an inner probability function of a certain structure. The significance of this perspective is that the satisfiability problem of a given belief formula with respect to the class of belief-function structures is equivalent to that with respect to the class of probability-function structures. So we can translate the complexity of satisfiability problem with respect to the class of belief-function structures into that of satisfiability problem with respect to the class of probability structures, which follows directly from the well-known result in [18] with respect to the class of Boolean algebras as well as the representation theorem for distributive lattices (Theorem A.2 in Appendix A). In Section 5, we show that the complexity of deciding whether a belief formula is satisfiable with respect to the class of belief-function structures on distributive lattices (especially on those structures for many non-classical logics) is *NP*-complete. Moreover, this perspective provides an alternative representation of belief functions to the well-known one by Shafer through the composition of probability functions with $\cap$-homomorphisms [53].

Unlike belief functions, probability functions in classical probability theory and measure theory [30] are usually not defined on all events. Let's consider a probability function Pr on the structure called *measurable space* $\langle S, \mathcal{A}\rangle$ where S is a finite sample space and $\mathcal{A}$ is an algebra on S and is a subclass of 2^S. Pr is defined on $\mathcal{A}$ but not on 2^S. In other words, not every subset of 2^S, which is called an event, is assigned a probability. Only those in $\mathcal{A}$ are assigned a probability and are called "measurable". Non-measurable events are usually considered mathematically meaningless. However, the notion of non-measurability is a desirable feature in reasoning with probabilities [18]. It provides a means to model uncertainty under incomplete information. The example in Section 1 in [11] and Example 2.3 in [18] give a motivation for this notion in dealing with uncertainty.

In order to generalize the above idea to the setting of distributive lattices, we need first to define the notion of *measurability*. Let L be a distributive lattice and Bel be a belief function on L. Define

$$L_{Bel} := \big\{a \in L\colon\ Bel(a \vee b) = Bel(a) + Bel(b) - Bel(a \wedge b) \text{ for every } b \in L\big\}.$$

Every element $a \in L_{Bel}$ is called *Bel-measurable* or simply measurable when the context is clear. This is a natural generalization of a classic approach to measure extension due to Caratheodory [8]. From Theorem 1 in [60], the following theorem follows immediately:

Theorem 3.16. *L_{Bel} is a sublattice of L and hence is distributive. Moreover, $Bel \restriction_{L_{Bel}}$ is a probability function on L_{Bel}.*

Let L' be a sublattice of L. So L' is distributive. If μ is a probability function on L', then, for each element $x \in L$, we can define

- $\mu_*(x) = \sup\{\mu(y)\colon\ y \in L', y \leqslant x\}$;
- $\mu^*(x) = \inf\{\mu(y)\colon\ y \in L', y \geqslant x\}$.

The above defined two functions μ_* and μ^* on L are called *inner and outer probability functions* on L respectively. Note that they depend on the choice of L'. The probability function μ assigns a precise probability to each element in the *sublattice* L' and is not well-defined at other elements in L. So it is actually a *partial* function on L. In contrast, both μ_* and μ^* which are induced by μ, are *total* on L. Intuitively, for each element $a \in L$, $\mu_*(a)$ is the probability of the biggest element in L' which is less than or equal to a; dually, $\mu^*(a)$ is the probability of the smallest element in L' which is greater than or equal to a. In other words, given the probability function on the sublattice L', μ_* and μ^* give the *best* bounds on the "true" probability at each element $a \in L$.

The following proposition reveals a relation of inner probability functions to belief functions.

Lemma 3.17. *The above defined μ_* is a belief function on L.*

Proof. Let L be a distributive lattice, L' be a sublattice and μ be a probability function on L'. It follows that there is a mass assignment $m_{L'}$ that corresponds to μ such that, for any element $a \in L'$, $\mu(a) = \sum_{b \leqslant a, b \in L'} m_{L'}(b)$. Now we define a mass assignment m_L on L as follows:

$$m_L(x) = \begin{cases} m_{L'}(x) & \text{if } x \in L', \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that $m_L(a) \geqslant 0$ for all $a \in L$. In other words, m_L is an *extension* of $m_{L'}$ to L.

Claim 5. *For the above defined μ_*, $\mu_*(a) = \sum_{b\in L', b\leqslant a} m_L(b)$ for all $a \in L$.*

For any $a \in L$, let $a_{L'} := \bigvee_{\substack{y\leqslant a\\ y\in L'}} y$. The proof of the claim proceeds as follows:

$$\begin{aligned}
\mu_*(a) &= \sup\bigl\{\mu(y):\ y \in L', y \leqslant a\bigr\} \quad \text{(by definition)}\\
&= \mu\Bigl(\bigvee_{\substack{y\leqslant a\\ y\in L'}} y\Bigr) \quad \bigl(\text{since } a_{L'} \in L'\bigr)\\
&= \sum_{\substack{x\in L'\\ x\leqslant a_{L'}}} m_{L'}(x)\\
&= \sum_{\substack{x\in L'\\ x\leqslant a}} m_{L'}(x)\\
&= \sum_{\substack{x\in L\\ x\leqslant a}} m_L(x) \quad (m_L \text{ is an extension of } m_{L'}).
\end{aligned}$$

This is to say, μ_* is a belief function on L that corresponds to the mass assignment m_L. $\square$

Corollary 3.18. *If Bel is a belief function on a distributive lattice L, then $Bel \geqslant (Bel\restriction_{L_{Bel}})_*$ in the sense that $Bel(a) \geqslant (Bel\restriction_{L_{Bel}})_*(a)$ for all $a \in L$.*

Lemma 3.17 says that every inner probability function is a belief function. But the converse does not quite hold. Let *Bel* be a belief function on the powerset of $S = \{s_1, s_2\}$ such that $Bel(\{s_1\}) = \frac{1}{2}$ and $Bel(\{s_2\}) = 0$. It is easy to check that *Bel* is not an inner probability function. So the inequalities in the above corollary may be strict. Still we can show that any belief function on a distributive lattice L is *equivalent to* an inner probability function in some *logical* sense. In order to show this equivalence, we need the following *positive* language which is the standard language for propositional logic without negation.

We start with a fixed infinite set $P := \{p_1, p_2, \ldots\}$ of propositional letters. We also use $p, q, \ldots$ to denote propositional letters. The set of formulas φ is built from propositional letters as usual by connectives $\vee$ and $\wedge$. Equivalently, a formula φ is formed by the following syntax:

$$\varphi := p \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2$$

where p is a propositional letter. We define Φ_0 the set of formulas in this syntax. Here we also use $\wedge$ and $\vee$ as connectives in logic. The context will determine whether they are used as lattice operations or as logical connectives.

Definition 3.19. A *probability structure* is a tuple $M = \langle P, L, L', \mu, v\rangle$ where P is a poset, both L and L' are distributive lattices, $L = J(P)$, L' is a sublattice of L, μ is a probability measure on L', and v associates each p with an element in L. v can be easily extended to a homomorphism from Φ_0 to L:

1. $v(\varphi_1 \wedge \varphi_2) := v(\varphi_1) \wedge v(\varphi_2)$;
2. $v(\varphi_1 \vee \varphi_2) := v(\varphi_1) \vee v(\varphi_2)$.

L_0 is called a *basis* of M if the elements in L_0 are disjoint from each other and L' consists precisely of all finite unions of elements in L_0. For any formula φ, $v(\varphi) \in L$. Generally, it is not true that $v(\varphi) \in L'$; in other words, $v(\varphi)$ is not necessarily "measurable".

Definition 3.20. A *DS structure* is a tuple $D = \langle P, L, Bel, v\rangle$ where P is a poset, L is a distributive lattice, $L = J(P)$, *Bel* is a belief function on L and v maps each propositional letter p to an element in L. Similarly, v can be easily extended to a homomorphism from Φ_0 to L as in the above definition.

We call a probability structure $M = \langle P, L, L', \mu, v\rangle$ and a *DS* structure $D = \langle P_D, L_D, Bel, v_D\rangle$ *equivalent* if

$$\text{for any formula } \varphi \in \Phi_0, \quad \mu_*\bigl(v(\varphi)\bigr) = Bel\bigl(v_D(\varphi)\bigr).$$

Theorem 3.21. *For any probability structure $M = \langle P, L, L', \mu, \nu\rangle$, there is an equivalent DS-structure.*

Proof. This follows directly from Lemma 3.17. $\square$

Now we prove the converse to this proposition.

Theorem 3.22. *For every DS structure $D = \langle P_D, L_D, Bel, \nu_D\rangle$, there is an equivalent probability structure $M = \langle P, L, L', \mu, \nu\rangle$.*

Here we simulate the proof in [18] to give a detailed proof, which can also be used to easily show the following corollaries.

Proof. Let $D = \langle P_D, L_D, Bel, \nu_D\rangle$ be a *DS*-structure and m be the corresponding mass assignment of *Bel*. According to the definition, $L_D = J(P_D)$.

The probability structure $M := \langle P, L, L', \mu, \nu\rangle$ is defined as follows. Let P be the set $\{(A, s)\colon\ s \in A, A \in J(P_D)\}$ with the identity relation. Let $A^* = \{(A, s)\colon\ s \in A\}$ for any $A \in J(P)$. It is easy to see that, if $A \neq B$, then $A^* \cap B^* = \emptyset$. In addition, $\bigcup_{A\in J(P_D)} A^* = P$. We take $\{A^*\colon\ A \in J(P_D)\}$ to be the "basis" for the distributive lattice L' in the sense that L' is the set of all finite unions of elements A^*. It is easy to see that L' with the usual set operations is indeed a distributive lattice and is actually a sublattice of $L := \langle 2^P, \cap, \cup\rangle$. It is interesting to note that both L and L' with the set complement are Boolean algebras.

Define $\mu(A^*) = m(A)$ for any $A \in J(P_D)$. Then we extend μ to all of L' by additivity. So μ is a probability function on L'. Define a valuation to L as follows: $\nu(p) = \{(B, s)\colon\ s \in \nu_D(p) \cap B, B \in J(P_D)\}$. We can show by induction on the complexity of φ that, for all formulas $\varphi \in \Phi_0$, $\nu(\varphi) = \{(B, s)\colon\ s \in \nu_D(\varphi) \cap B, B \in J(P_D)\}$.

Now we will show that such defined probability structure $M = \langle P, L, L', \mu, \nu\rangle$ is equivalent to the belief structure D. For each formula φ and $A \in J(P_D)$, we have $A^* \subseteq \nu(\varphi)$ iff $(A, s) \in \nu(\varphi)$ for all $s \in A$ iff $s \in \nu_D(\varphi)$ for all $s \in A$ iff $A \subseteq \nu_D(\varphi)$. So the largest "measurable" set (simply the largest element in L') contained in $\nu(\varphi)$ is $\bigcup_{A\subseteq \nu_D(\varphi)} A^*$. It follows that

$$\begin{aligned}
\mu_*\big(\nu(\varphi)\big) &= \mu\Bigg(\bigcup_{A\subseteq \nu_D(\varphi)} A^*\Bigg) \\
&= \sum_{A\subseteq \nu_D(\varphi)} \mu\big(A^*\big) \\
&= \sum_{A\subseteq \nu_D(\varphi)} m(A) \\
&= Bel\big(\nu_D(\varphi)\big).
\end{aligned}$$

So we have shown that M and D are equivalent. $\square$

Corollary 3.23. *For every DS-structure $D = \langle P_D, L_D, Bel, \nu_D\rangle$ on a distributive lattice L_D, there is an equivalent probability structure $M = \langle P, L, L', \mu, \nu\rangle$ such that both L and L' are Boolean algebras.*

Proof. This follows directly from the above construction of L and L'. $\square$

These equivalence results turn out to be crucial to show the complexity of reasoning about belief functions over some non-Boolean structures (Lemma 5.5 in Section 5). By summarizing the results in the above two theorems, we conclude that belief functions and inner probability functions are equivalent on distributive lattices if we view them both as functions on *formulas* rather on sets. Just as in [18], there is an immediate payoff to this view of belief functions as inner probability functions: a logic for reasoning about belief functions in Section 5 is obtained from that for inner probability functions [19] with minor modifications. The following proposition is simply a corollary of the above theorem, which is a partial converse to Lemma 3.17. Also it provides an alternative representation of belief functions to the well-known one by Shafer through the composition of probability functions with $\cap$-homomorphisms in [53].

Corollary 3.24. *Given a DS structure $\langle P_D, L_D, Bel_D, \nu_D\rangle$ defined on a finite distributive lattice $L_D (= J(P_D))$, there is a probability structure $\langle P, L, L', \mu, \nu\rangle$ and a surjection $f : P \to P_D$ such that, for each $x \in L_D$, we have $Bel(x) = \mu_*(f^{-1}(x))$.*

Proof. Here we choose L and L' in the proof of Theorem 3.22. A function $f : P \to P_D$ is defined as: $f((A, s)) = s$ for $(A, s) \in P$ such that $s \in A$ and $A \in J(P_D)$. f is a surjection from P to P_D. Then the argument in the last part of the proof of Theorem 3.22 works here simply by replacing x for $\nu_D(\varphi)$ and $f^{-1}(x)$ for $\nu(\varphi)$. $\square$

4. Belief functions on non-classical formalisms

The integration of belief functions and non-classical formalisms is intended to master two sources of ignorance. Non-classical formalisms usually take care of the incompleteness or inconsistency in the knowledge-base due to imperfect data while belief functions take care of the limitation of the information that the agent has *at his disposal*. In artificial intelligence especially in Knowledge Representation, non-classical formalisms play an important role in handling *imperfect* information in different forms. Most of these non-classical formalisms assume a mathematical setting of distributive lattices (quantum logic is probably one of the very few important exceptions [67]). Each of these formalisms was intended for reasoning about some specific form of information. For example, Kleene's three valued logic has been used to take into account statements that being undefined makes sense to be a third truth value, which is useful to model the situation in computer science when a computation does not return any result [23]; paraconsistent logics have been used to deal with contradictory knowledge bases [3,44]; relevance logic is used to deal with the famous logical omniscience problem in the foundation of knowledge representation, and used for reasoning in the presence of inconsistency in knowledge base systems [37,38,19].

Belief functions in the Dempster–Shafer theory are defined on Boolean algebras [52]. One essential difference of non-classical formalisms from the Boolean setting is their specific treatments of negation. Negation is closely related to the treatment of *bipolarity* in information [14], which means that there is an intrinsic positive and negative effect in dealing with information. In the classical Dempster–Shafer theory, negation is assumed to be *Boolean*, i.e., every element has a complement (for any element a, there is an element a' such that $a \vee a'$ is the top element and $a \wedge a'$ is the bottom), which is used to represent *complete* information. Any distributive lattice with a Boolean negation is a Boolean algebra and any Boolean algebra is represented as a power set with the usual set operations (Example 2.3). It is shown [18] that reasoning about belief functions in this case is the same as that for inner probability measures [19]. In Kleene's three valued logic, there are three truth values: true, false and undetermined. Logically, the treatment of negation considers some formulas to be neither true nor false (undetermined) but forbids any formula to be both true and false. In other words, positive and negative sides don't exhaust all possibilities. So this logic is used to represent incompleteness in information. There is an *implicit intuitionistic* negation in any finite distributive lattice. Since any finite distributive lattice $L = J(P)$ for some poset P is also a Heyting algebra, $\max\{I' \in J(P)\colon\ I' \cap I = \emptyset\}$ exists for any $I \in J(P)$, and is defined to be the negation of I (denoted by $\sim I$), which is in $J(P)$. The duality relation between belief function and plausibility function in Dempster–Shafer theory can be expressed in the following form: for any $I \in L$,

$$\begin{aligned}
1 - Bel(\sim I) &= 1 - Bel\bigl(\max\bigl\{I' \in J(P)\colon\ I' \cap I = \emptyset\bigr\}\bigr)\\
&= 1 - \sum_{\substack{I' \in J(P)\\ I' \subseteq \max\{I' \in J(P)\colon\ I' \cap I = \emptyset\}}} m\bigl(I'\bigr)\\
&= \sum_{\substack{I' \in J(P)\\ I' \cap I \neq \emptyset}} m\bigl(I'\bigr)
\end{aligned}$$

where *Bel* is a belief function on L and m is the corresponding mass assignment of *Bel*. So, the duality relation is essentially *intuitionistic*. One may also reason about belief functions in this case by replacing the classical proposition logic by intuitionistic propositional logic (see discussion in Section 6). A good example in this direction is bipolar belief pairs on vague propositions as an integrated model combining epistemic uncertainty and indeterminacy [36].

Kleene's three-valued logic finds a natural generalization in Belnap's four-valued logic [5], which can be naturally extended to distributive bilattices. A distributive bilattice is a distributive lattice with a second ordering which interacts with the original one in a certain way. In addition to their applications in logic programming [23], distributive bilattices are also used to represent the inconsistency in knowledge base systems. In these structures, another form of negation called de Morgan negation is employed. The most important aspect of de Morgan negations is their intrinsic ability to model inconsistency in knowledge base systems. In this paper, we will consider this type of negations and integrate belief functions with de Morgan lattices, which are distributive lattices with de Morgan negations, provide an axiomatization of reasoning about belief functions over this type of non-classical structures and discuss computational complexities of different problems in this setting. More importantly, this approach to reasoning about belief functions on de Morgan lattices also provides a framework to reason about belief functions on other non-classical structures.

A well-known slogan in algebraic methods for non-classical logics (Section 18 of [1]) tells us that algebra and logic are *dual* to each other. Given a non-classical logic $\mathbb{L}$, the Lindenbaum algebra $\mathbb{A}_{\mathbb{L}}$ of this logic is in the class of algebras which characterize the logic but also the possible-world-like structure derived from the Lindenbaum algebra $\mathbb{A}_{\mathbb{L}}$ through the Duality theorem for this class of algebras is the *canonical structure* for this logic. On the other hand, given any possible-world-like structure for the logic $\mathbb{L}$, the interpretations of the formulas is an algebra that characterizes $\mathbb{L}$. So, in the following section, before we reason about belief functions on specific classes of algebras, we elaborate on this kind of duality between logics and their corresponding algebraic structures (and possible-world-like semantics) and won't distinguish belief functions on algebras and for their logics.

In addition to this duality, algebra and logic are dual to each other in the sense that (order) ideals to algebra are the same as filters to logic. If we replace in Section 3 all posets $\langle P, \leqslant\rangle$ with its dual $\langle P, \leqslant^{\partial}\rangle$ (where $x \leqslant^{\partial} y$ iff $y \leqslant x$) and all $\leqslant$-order-ideals by $\leqslant^{\partial}$-filters, the dual forms of all propositions there remain valid. So, in order to apply the algebraic propositions in Section 3 to non-classical logics in this section, we have to keep this duality in mind.

4.1. Reasoning about bilattices

Bilattices are algebras with two separate lattice structures. Ginsberg [25] suggested using bilattices as the underlying framework for various AI inference systems including those based on default logics, truth maintenance systems, probabilistic logics, and others. These ideas were later pursued in the context of logic programming semantics [22,21]. Moreover, bilattices and their extensions have been used in the literature to model a variety of reasoning mechanisms about uncertainty in the presence of incomplete or contradictory information [33,3,54]. Also they have been employed to represent bipolar information [34]. In the following, we first present a well-known algebraic result about the representation of bilattices. In logic, we employ the Belnap's four-valued logic to *decouple* the interpretation of each formula φ into the set of states where φ is "true" and that of states where it is "false" [15]. In this way, the interpretations of all formulas form a distributive bilattices with two partial orderings: the *truth ordering* and the *knowledge ordering*.

One may refer to [40,23] for the technical details about the duality theorem of bilattices which is presented below. Recall that all lattices are assumed to be finite.

Definition 4.1. A *bilattice* is an algebra $\mathbf{B} = \langle B, \wedge_1, \vee_1, \bot_1, \top_1, \wedge_2, \vee_2, \bot_2, \top_2\rangle$ such that $\mathbf{B}_1 = \langle B, \wedge_1, \vee_1, \bot_1, \top_1\rangle$ and $\mathbf{B}_2 = \langle B, \wedge_2, \vee_2, \bot_2, \top_2\rangle$ are lattices. By a *negation* on $\mathbf{B}$ we mean a unary operation $\neg$ on B satisfying the conditions:

1. $\neg\neg x = x$;
2. $\neg(x \vee_1 y) = \neg x \wedge_1 \neg y$, $\neg(x \wedge_1 y) = \neg x \vee_1 \neg y$;
3. $\neg(x \vee_2 y) = \neg x \vee_2 \neg y$, $\neg(x \wedge_2 y) = \neg x \wedge_2 \neg y$.

$\mathbf{B}$ is called *distributive* if, for every $\Diamond, \Box \in \{\wedge_1, \vee_1, \wedge_2, \vee_2\}$ and all $x, y, z \in B$ $x\Diamond(y\Box z) = (x\Diamond y)\Box(x\Diamond z)$.

The lattice ordering corresponding to the lattice $\mathbf{B}_1$ will be denoted by $\leqslant_1$ and the lattice ordering corresponding to $\mathbf{B}_2$ by $\leqslant_2$; often the bilattice $\mathbf{B}$ is written in the form $\langle B, \leqslant_1, \leqslant_2\rangle$. Alternatively, $\leqslant_1$ and $\leqslant_2$ are often denoted by $\leqslant_t$ and $\leqslant_k$, respectively, reflecting the fact that they represent the "truth" and "knowledge" orderings, which will become clear in the following Definition 4.4 and in Remark 4.7.

Definition 4.2. Let $\mathbf{L} = \langle L, \wedge, \vee, \bot, \top\rangle$ and $\mathbf{L}' = \langle L', \wedge', \vee', \bot', \top'\rangle$ be lattices. Define $\mathcal{B}(\mathbf{L}, \mathbf{L}') = \langle L \times L', \sqcap_1, \sqcup_1, \bot_1, \top_1, \sqcap_2, \sqcup_2, \bot_2, \top_2\rangle$ as follows: for all $(x, x'), (y, y') \in L \times L'$,

- $(x, x') \sqcap_1 (y, y') = (x \wedge y, x' \vee' y')$, $(x, x') \sqcup_1 (y, y') = (x \vee y, x' \wedge' y')$;
- $(x, x') \sqcap_2 (y, y') = (x \wedge y, x' \wedge' y')$, $(x, x') \sqcup_2 (y, y') = (x \vee y, x' \vee' y')$;
- $\bot_1 = (\bot, \top')$, $\top_1 = (\top, \bot')$, $\bot_2 = (\bot, \bot')$, $\top_2 = (\top, \top')$.

$\mathcal{B}(\mathbf{L}, \mathbf{L}')$ is called the *product bilattice associated with* L and L'. If $L = L'$, we define

$$\sim (x, x') = (x', x),$$

and call $\mathcal{B}(L, L)$ the *square bilattice with negation associated with* L. Usually in this case we write $\mathcal{B}(L)$ for $\mathcal{B}(L, L)$.

Any abstract distributive bilattice with negation can be represented as a square bilattice with negation (Corollary 9 in [40]). Let $\mathbf{B} = \langle B, \leqslant_1, \leqslant_2\rangle$ be a distributive bilattice. An element $x \in B$ is called *positive* if, for every $y \in B$, $x \leqslant_1 y$ implies $x \leqslant_2 y$. It is called *negative* if, for every $y \in B$, $y \leqslant_1 x$ implies $x \leqslant_2 y$. Intuitively, an element x is positive (negative) if it should increase in the knowledge ordering whenever it increases (decreases) in the truth order. Denote by $POS(B)$ and $NEG(B)$ the set of positive and negative elements (they are called t-grounded and f-grounded in [25]), respectively, of B. An element $x \in B$ is called *positive* (*resp. negative*) $\leqslant_2$*-join-irreducible* if it is positive (resp. negative) and join-irreducible with respect to $\leqslant_2$-ordering. We denote by $\Im_2^+(B)$ (resp. $\Im_2^-(B)$) the set of non-bottom positive (resp. negative) $\leqslant_2$-join-irreducible elements of B. Note that both $\Im_2^+(B)$ and $\Im_2^-(B)$ are posets with respect to the ordering $\leqslant_2$. Moreover, $\Im_2(B)$ denotes the set of all $\leqslant_2$-join-irreducible elements of B. If a bilattice $\mathbf{B}$ is distributive and is represented as a square bilattice, then it is easy to recognize those positive and negative elements. Let L be a distributive lattice and let 0 be the least element. An element x of $\mathcal{B}(L)$ is positive iff $x = (y, 0)$ for some $y \in L$, and is negative iff $x = (0, y)$ for some $y \in L$ (Corollary 8 in [40]). The positive and negative elements here correspond to the two support relations $\models_T$ and $\models_F$ in the following four-valued models.

Let $P = \langle P, \preccurlyeq\rangle$ and $Q = \langle Q, \sqsubseteq\rangle$ be two disjoint partially ordered sets. Define the *lift* of P, denoted by $P_\bot$ by $P_\bot = \langle P \cup \{0\}, \leqslant\rangle$, where $0 \notin P$ and $x \leqslant y$ in $P_\bot$ iff $x = 0$ or $x \preccurlyeq y$ in P. Define the *disjoint union* $P \uplus Q = \langle P \cup Q, \leqslant\rangle$ to be the partially ordered set with $x \leqslant y$ iff either $x, y \in P$ and $x \preccurlyeq y$ or $x, y \in Q$ and $x \sqsubseteq y$.

Given two partially ordered sets P and Q, define the *separated sum* of P and Q, denoted $P \oplus_\bot Q$, to be the poset $P \oplus_\bot Q = (P \uplus Q)_\bot$. The following theorem is the bilattice counterpart of Birkhoff's representation theorem.

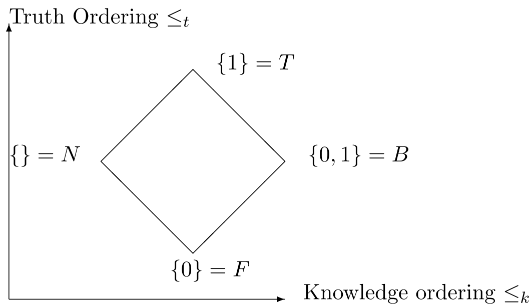

Fig. 4. Lattice *FOUR* interpreted by subsets of $\{0, 1\}$.

Theorem 4.3. *(See Corollary 23 in [40].) Let B be a distributive bilattice with negation. Then* $\Im_2(B) \cong \Im_2^+(B) \oplus_\perp \Im_2^-(B)$ *and* $\Im_2^+(B) \cong \Im_2^-(B)$*. Conversely, for any finite poset P, there is a finite distributive bilattice* $\mathbf{B}$ *such that* $\Im_2(B) \cong P \oplus_\perp P$*. So there is a one-to-one correspondence between finite distributive bilattices and finite posets.*

Intuitively, the theorem says that any distributive lattice with negation can be represented as a *concrete* square bilattice. Let $J_2(\Im_2(B))$ denote the set of $\leqslant_2$-order-ideals in $\Im_2(B)$. It is easy to check that it is the same as the Cartesian product of the set of order ideals in $\Im_2^+(B)$ with that in $\Im_2^-(B)$. Since $\Im_2^+(B) \cong \Im_2^-(B)$, $J_2(\Im_2(B)) \approx J(P) \times J(P)$ for some poset P which is isomorphic to $\Im_2^+(B)$. For any two elements $(I_1, I_2), (I'_1, I'_2) \in J_2(\Im_2(B))$,

- $(I_1, I_2) \leqslant_1 (I'_1, I'_2)$ if $I_1 \subseteq I_2$ and $I'_1 \supseteq I'_2$;
- $(I_1, I_2) \leqslant_2 (I'_1, I'_2)$ if $I_1 \subseteq I_2$ and $I'_1 \subseteq I'_2$.

The above representation theorem tells us that $\mathbf{B}$ is isomorphic to $\langle J_2(\Im_2(B)), \leqslant_1, \leqslant_2\rangle$. It also connects bilattice theory to the well-known four-valued logic [5].

FOUR, the structure that corresponds to Belnap's four-valued logic [5], is the minimal bilattice, exactly as the structure $\mathbf{2} = \{true, false\}$ or $\{0, 1\}$ that is based on the classical two valued logic is the minimal Boolean algebra. It plays an important role in bilattice-based multi-valued logics. The Hasse diagram of the lattice *FOUR* is illustrated in Fig. 4.

Following Dunn [17], we interpret *FOUR* in terms of the power set of $\{0, 1\}$. The meaning of the capital letters attached to the elements of *FOUR* in the above figure is obvious from this type of interpretation. For example, B informally means "both" and can be translated as both "true" (1) and "false" (0). The truth ordering $\leqslant_t$ can be formalized as follows: for any two elements $x, y \in \{T, F, N, B\}$,

$$x \leqslant_t y \quad \text{if both } 1 \in x \text{ implies } 1 \in y \text{ and } 0 \in y \text{ implies } 0 \in x.$$

The *lattice FOUR* is the tuple $\langle\{T, F, N, B\}, \vee, \wedge, \sim\rangle$ where $\wedge$ and $\vee$ are the lattice operations associated with the above truth ordering. Also we define $\sim$ as an order inverting that leaves N and B as fixed points, i.e., $\sim N = N$, $\sim B = B$, $\sim T = F$, $\sim F = T$. Then meaning of this negation $\sim$ will be clear from the following semantical meaning of formulas (See Remark 4.7).

It is interesting to note that there is a natural *knowledge* ordering implicit in the above lattice *FOUR*. The knowledge ordering $\leqslant_k$ is defined as follows: for any two elements $x, y \in \{T, F, N, B\}$,

$$x \leqslant_k y \quad \text{if both } 1 \in x \text{ implies } 1 \in y \text{ and } 0 \in x \text{ implies } 0 \in y.$$

So the lattices operations $\sqcap$ and $\sqcup$ associated with $\leqslant_k$ are simply the usual set operations $\cap$ and $\cup$. It is easy to check that $\langle\{T, F, N, B\}, \vee, \wedge, \sqcap, \sqcup, \sim, \rangle$ is a distributive bilattice. It is actually the smallest distributive bilattice. $\mathbf{2}$ is simply the sublattice $\langle\{T, F\}, \wedge, \vee\rangle$ of the reduct $\langle\{T, F, N, B\}, \wedge, \vee\rangle$.

The interpretations of standard propositional formulas in *FOUR* form exactly a distributive bilattice. Moreover, it was shown in [2] that all the natural bilattice-valued logics that we had introduced for various purposes can be characterized using only the four basic "epistemic truth values". The meaning of such epistemic truth values highly differs from the meaning of standard Boolean truth values since they are not intrinsic to propositions but are intended to reflect what an agent may have been informed about (regarding these propositions). Thus, interpreting a proposition φ as 0 (resp., 1) does not mean that φ is false (resp., true) but that the agent under consideration has some reasons to consider that φ is false (resp., true) or is told that φ is "false" (resp. "true"). The agent may have some reasons to consider that φ is false and other reasons to consider that φ is true, and the epistemic truth value B reflects this situation. Similarly, the agent may have no reasons to consider φ as true and no reasons to consider it as false; in this situation, φ is given the epistemic truth value N (or $\emptyset$). So B reflects a situation of inconsistency and N of ignorance. The epistemic nature in the interpretations into *FOUR* agrees well with Shafer's emphasis of the epistemic nature of the set of possibilities on the frame of discernment in his theory (page 36 in [52]) but under incomplete or inconsistent information.

In order to reason about belief functions for the four-valued logic, we expand Φ_0 in the last section to Φ by adding the connective negation $\sim$. In other words, a formula in Φ is formed by the following syntax:

$$\varphi := p \mid \sim\varphi \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2$$

where p is a propositional letter.

A *valuation* v into the lattice *FOUR* is a function from the set P of propositional letters into *FOUR*. It is easy to see that v can be extended to the set of formulas naturally as follows:

- $v(\sim\varphi) = \sim v(\varphi)$;
- $v(\varphi \wedge \psi) = v(\varphi) \wedge v(\psi)$;
- $v(\varphi \vee \psi) = v(\varphi) \vee v(\psi)$.

Following Belnap [5], we simply say "φ is at least true" if $1 \in v(\varphi)$; "φ is at least false" if $0 \in v(\varphi)$. It follows immediately that

- $1 \in v(\varphi)$ iff $0 \in v(\sim\varphi)$, $0 \in v(\varphi)$ iff $1 \in v(\sim\varphi)$;
- $1 \in v(\varphi \vee \psi)$ iff $1 \in v(\varphi)$ or $1 \in v(\psi)$, $0 \in v(\varphi \vee \psi)$ iff $0 \in v(\varphi)$ and $0 \in v(\psi)$;
- $1 \in v(\varphi \wedge \psi)$ iff $1 \in v(\varphi)$ and $1 \in v(\psi)$, $0 \in v(\varphi \wedge \psi)$ iff $0 \in v(\varphi)$ or $0 \in v(\psi)$.

In order to reason about belief functions on distributive bilattices, we simulate Kripke's semantics for intuitionistic logic [35] and choose to define belief structures in a more abstract form.

Definition 4.4. A *Belnap structure* is a tuple $S = \langle S, \leqslant, v\rangle$ where

- $\langle S, \leqslant\rangle$ is a poset where S is a non-empty set of states called situations [37] or set-ups [16], which are like possible worlds[3] except that they are not required to be either consistent or complete;
- for each $s \in S$, $v(s)$ is a valuation into *FOUR* on the set of propositional letters such that the following *persistency* condition is satisfied: for any $s_1, s_2 \in S$ and propositional letter p,

$$\text{if } s_1 \leqslant s_2, \quad \text{then } 1 \in v(s_1)(p) \text{ implies } 1 \in v(s_2)(p) \text{ and } 0 \in v(s_1)(p) \text{ implies } 0 \in v(s_2)(p).$$

It is a *Boolean structure* if $\leqslant$ is the identity relation and v is a valuation into **2**.

Two *support relations* between states and formulas [15,37] are defined inductively as follows:

- $S, s \models_T p$ if $1 \in v(s)(p)$,
  $S, s \models_F p$ if $0 \in v(s)(p)$;
- $S, s \models_T \varphi_1 \wedge \varphi_2$ if $S, s \models_T \varphi_1$ and $S, s \models_T \varphi_2$;
  $S, s \models_F \varphi_1 \wedge \varphi_2$ if $S, s \models_F \varphi_1$ or $S, s \models_F \varphi_2$;
- $S, s \models_T \varphi_1 \vee \varphi_2$ if $S, s \models_T \varphi_1$ or $S, s \models_T \varphi_2$;
  $S, s \models_F \varphi_1 \vee \varphi_2$ if $S, s \models_F \varphi_1$ and $S, s \models_F \varphi_2$;
- $S, s \models_T \sim\varphi$ if $S, s \models_F \varphi$;
  $S, s \models_F \sim\varphi$ if $S, s \models_T \varphi$.

Note that, for any $s \in S$ and any formula φ, $1 \in v(s)(\varphi)$ iff $S, s \models_T \varphi$ and $0 \in v(s)(\varphi)$ iff $S, s \models_F \varphi$. Actually the persistency condition for the valuation is satisfied by all formulas, as shown by the following lemma.

Lemma 4.5. *Let $\langle S, \leqslant, v\rangle$ be a Belnap structure. If $s_1, s_2 \in S$ and $s_1 \leqslant s_2$, then, for any formula $\varphi \in \Phi$,*

1. *$S, s_1 \models_T \varphi$ implies $S, s_2 \models_T \varphi$;*
2. *$S, s_1 \models_F \varphi$ implies $S, s_2 \models_F \varphi$.*

Proof. Let $\langle S, \leqslant, v\rangle$ be a Belnap structure and $s_1, s_2 \in S$ and $s_1 \leqslant s_2$. We prove by induction on the complexity of φ. Here we only prove the case that $\varphi = \sim\varphi'$. The proof of the other cases is straightforward.

We reason as follows:

$$S, s_1 \models_T \varphi \Rightarrow S, s_1 \models_F \varphi'$$

[3] Possible worlds are required to be both consistent and complete.

$$\begin{aligned}
&\Rightarrow S, s_2 \models_F \varphi' \quad \text{(Induction hypothesis)}\\
&\Rightarrow S, s_2 \models_T \sim\varphi'
\end{aligned}$$

and

$$\begin{aligned}
S, s_1 \models_F \varphi &\Rightarrow S, s_1 \models_T \varphi'\\
&\Rightarrow S, s_1 \models_T \varphi' \quad \text{(Induction hypothesis)}\\
&\Rightarrow S, s_1 \models_F \sim\varphi'. \qquad \square
\end{aligned}$$

Remark 4.6. The persistency condition in a Belnap structure is quite similar to that in the Kripke semantics for intuitionistic logic [35,64]. Following the interpretation of the non-standard propositional logic in [20], we analyze the above semantics in terms of knowledge-bases. Each state s can be seen as a pair of knowledge bases. The set of formulas that are at least true at s is the knowledge base of true facts and the set of formulas which are at least false at s constitutes the knowledge base for the false facts. This pair of knowledge evolves in the course of time. Both the knowledge base of true facts and that of false facts expand at every later stage. They are considered to be independent of each other and both take the *open-world assumption*, which also explains the incompleteness in the information in another way. So the essential difference of our interpretation from that in intuitionistic logic is that the persistency condition here considers not only the knowledge of true facts but also that of false facts.

Let $[[\varphi]]_T$ denote the set of all states where φ is "told true" $\{s \in S\colon\ S, s \models_T \varphi\}$ and $[[\varphi]]_F = \{s \in S\colon\ S, s \models_F \varphi\}$ for all formulas φ. It is easy to check that both of them are filters in $(S, \leqslant)$. Denote

$$B = \bigl\{\bigl([[\varphi]]_T, [[\varphi]]_F\bigr)\colon\ \varphi \in \Phi\bigr\}.$$

We can define two partial orders on B.

- $([[\varphi_1]]_T, [[\varphi_1]]_F) \leqslant_1 ([[\varphi_2]]_T, [[\varphi_2]]_F)$ if $[[\varphi_1]]_T \subseteq [[\varphi_2]]_T$ and $[[\varphi_1]]_F \supseteq [[\varphi_2]]_F$;
- $([[\varphi_1]]_T, [[\varphi_1]]_F) \leqslant_2 ([[\varphi_2]]_T, [[\varphi_2]]_F)$ if $[[\varphi_1]]_T \subseteq [[\varphi_2]]_T$ and $[[\varphi_1]]_F \subseteq [[\varphi_2]]_F$.

Remark 4.7. The four valued logic is employed to decouple the bipolar information in the semantics. And the interpretation of each formula φ is decomposed into two parts: the part for the epistemic truth and the other part for the epistemic falsity. For the partial ordering $\leqslant_1$, the agent has more reasons to consider φ_2 as true than φ_1 and more reasons to consider φ_1 as false than φ_2. In other words, φ_2 is *considered* at least as true as and at most as false as φ_1. The agent is more confident in considering that *overall* φ_2 is at least as true as φ_1. This is the reason why $\leqslant_1$ is also called the truth ordering. For the other ordering $\leqslant_2$, the agent has both more reasons to consider φ_2 as true than φ_1 and more reasons to consider φ_2 as false than φ_1. So the agent is more *informative* (in reasons) about φ_2 than about φ_1. This is the reason why $\leqslant_2$ is called the information or knowledge ordering.

It is easy to check that the associated structure $\mathcal{B} := \langle B, \leqslant_1, \leqslant_2\rangle$ is a distributive bilattice with the following negation:

$$\sim\bigl([[\varphi]]_T, [[\varphi]]_F\bigr) = \bigl([[\varphi]]_F, [[\varphi]]_T\bigr) = \bigl([[\sim\varphi]]_T, [[\sim\varphi]]_F\bigr).$$

Intuitively, for each formula φ, $[[\varphi]]_T$ and $[[\varphi]]_F$ are the *positive and negative sides* of the interpretation of this formula and they correspond to positive and negative elements in the bilattice $\mathcal{B}$, respectively.

Definition 4.8. A formula ψ is a *logical consequence* of a formula φ (φ *logically implies* ψ) with respect to the class $\mathcal{B}$ of Belnap structures (denoted $\varphi \models^{\mathcal{B}} \psi$) if, for any Belnap structure $S = \langle S, \leqslant, v\rangle$ and any $s \in S$, $S, s \models_T \varphi$ implies $S, s \models_T \psi$ and $S, s \models_F \psi$ implies $S, s \models_F \varphi$.

Now we investigate deductive system for this logical implication $\models^{\mathcal{B}}$ with respect to the class of Belnap structures. The following is the deductive system $\mathbf{R}_{fde}$ which is the well-known first-degree entailment fragment of the relevance logic **R** [1,16]. Without further notice, $\vdash$ denotes $\vdash$ and $\varphi \dashv\vdash \psi$ is short for both $\varphi \vdash \psi$ and $\psi \vdash \varphi$.

Axioms:

$$\begin{array}{rll}
\varphi \vdash \varphi & & \text{Self-implication}\\
\varphi \wedge \psi \vdash \varphi, & \varphi \wedge \psi \vdash \psi & (\wedge\text{-elimination})\\
\varphi \vdash \varphi \vee \psi, & \psi \vdash \varphi \vee \psi & (\vee\text{-introduction})\\
\varphi \wedge (\psi \vee \gamma) \vdash (\varphi \vee \psi) \wedge (\psi \vee \gamma) & \varphi \dashv\vdash \sim\sim\varphi & \text{(Double Negation)}\\
\sim(\varphi \wedge \psi) \dashv\vdash \sim\varphi \vee \sim\psi & \sim(\varphi \vee \psi) \dashv\vdash \sim\varphi \wedge \sim\psi & \text{(de Morgan laws)}
\end{array}$$

Rules:

- From $\varphi \vdash \psi$ and $\psi \vdash \gamma$, infer $\varphi \vdash \gamma$ (Transitivity)
- From $\varphi \vdash \psi$ and $\varphi \vdash \gamma$, infer $\varphi \vdash \psi \wedge \gamma$ ($\wedge$-introduction)
- From $\varphi \vdash \gamma$ and $\psi \vdash \gamma$, infer $\varphi \vee \psi \vdash \gamma$ ($\vee$-elimination)
- From $\varphi \vdash \psi$, infer $\sim\psi \vdash \sim\varphi$ (Contraposition)

Actually the logical implication relation in the class of Belnap structures coincides with the above consequence relation $\vdash$.

Theorem 4.9. *(See Theorem 7 in [17].) For any two formulas φ and ψ in Φ,*

$$\varphi \vdash \psi \quad \textit{iff} \quad \varphi \models^{\mathcal{B}} \psi.$$

Definition 4.10. A *Dempster–Shafer structure* B (*DS*-structure for short) on a Belnap structure is a tuple $\langle S, \leqslant, v, Bel\rangle$ where

- $\mathrm{S} = \langle S, \leqslant, v\rangle$ is a Belnap structure;
- Bel is a belief function on $\langle F(S) \times F(S), \leqslant_t\rangle = \langle\{(I, I') : I \text{ and } I' \text{ are filters in } S\}, \leqslant_t\rangle$.

Note that, for any $(I_1, I_2), (I'_1, I'_2) \in F(S) \times F(S)$, if $(I_1, I_2) \leqslant_t (I'_1, I'_2)$, i.e., $I_1 \subseteq I'_1$ and $I_2 \supseteq I'_2$, then $Bel(I_1, I_2) \leqslant Bel(I'_1, I'_2)$. For the Belnap structure $\langle S, \leqslant, v\rangle$, $([[\varphi]]^{\mathrm{S}}_T, [[\varphi]]^{\mathrm{S}}_F)$ denotes the pair of the set of states where φ is satisfied, i.e., $[[\varphi]]^{\mathrm{S}}_T = \{s \in S\colon\ S, s \models_T \varphi\}$, which can be shown to be a filter in S, and that of states where φ is falsified, i.e., $[[\varphi]]^{\mathrm{S}}_F = \{s \in S\colon\ S, s \models_F \varphi\}$, which can be shown to be also a filter in S. So the belief in the formula φ is defined to be $Bel([[\varphi]]^{\mathrm{S}}_T, [[\varphi]]^{\mathrm{S}}_F)$. It is easy to see that the above definition agrees with Definition 3.20.

Definition 4.11. For any two formulas φ and ψ in Φ, φ *probabilistically entails* ψ with respect to the class of Belnap structures (denoted as $\varphi \models^{B}_{DS} \psi$) if, for any *DS*-structure $S = \langle S, \leqslant, Bel, v\rangle$ where $\langle S, \leqslant, v\rangle$ is a Belnap structure, $Bel([[\varphi]]^{\mathrm{S}}_T, [[\varphi]]^{\mathrm{S}}_F) \leqslant Bel([[\psi]]^{\mathrm{S}}_T, [[\psi]]^{\mathrm{S}}_F)$.

The following theorem tells us that the deductive system $\mathbf{R}_{fde}$ provides a sound and complete system for both logical implication but also probabilistic entailment.

Theorem 4.12. *For any formulas φ and ψ in Φ, $\varphi \vdash \psi$ if and only if $\varphi \models^{B}_{DS} \psi$.*

Proof. The left-to-right direction is straightforward. Now we show the other direction by contraposition. Assume that $\varphi \nvdash \psi$. According to Theorem 4.9, $\varphi \not\models^{B} \psi$. It follows that there is a Belnap model $B = \langle S, \leqslant, v\rangle$ such that either $[[\varphi]]^{B}_T \nsubseteq [[\psi]]^{B}_T$ or $[[\psi]]^{B}_F \nsubseteq [[\varphi]]^{B}_F$. In other words, $([[\varphi]]^{B}_T, [[\varphi]]^{B}_F) \nleq^1_{\Phi} ([[\psi]]^{B}_T, [[\varphi]]^{B}_F)$. It is easy to find a belief function Bel on B such that $Bel([[\varphi]]^{B}_T, [[\varphi]]^{B}_F) \nleq Bel([[\psi]]^{B}_T, [[\varphi]]^{B}_F)$. This implies that $\varphi \not\models^{B}_{DS} \psi$. $\square$

4.2. Reasoning about de Morgan lattices

De Morgan lattices are important for the study of the mathematical aspects of fuzzy logic [69]. The standard fuzzy algebra $F = \langle[0, 1], \max(x, y), \min(x, y), 0, 1, 1 - x\rangle$ is an example of a de Morgan lattice. Moreover, de Morgan lattices are the algebraic semantics for relevance logic [16]. In this part, we investigate belief functions on de Morgan lattices which covers those for fuzzy events [68,55,66]. It is interesting to note that the first degree entailments $\mathbf{R}_{fde}$ also provides a calculus for reasoning about de Morgan lattices. In the following, we will give a presentation of the semantics in terms of de Morgan lattices which is in parallel to that for bilattices. For simplicity, we will not repeat those motivations which are similar to those in last part about bilattices but simply present the main ideas. Note that the $\wedge_1$-$\vee_1$-$\neg$-reduct of a distributive bilattice $\mathbf{B} = \langle B, \wedge_1, \vee_1, \wedge'_1, \vee'_2, \neg\rangle$ is a de Morgan lattice.

We need the following duality theorem for finite de Morgan lattices which is based on [7,16,62,45]:

Theorem 4.13. *Any finite de Morgan lattice D can be represented as the lattice $J(P_D)$ of order ideals in the subposet P_D of join-irreducibles with an order-reversing involution g. And there is a one-to-one correspondence between de Morgan lattices and posets with order-reversing involutions.*

So each de Morgan lattice can be regarded as a poset with an order-reversing involution. Similarly, in order to reason about belief functions on de Morgan lattices, we choose to define underlying structures in a more abstract form.

Definition 4.14. A *Routley structure* is a tuple $S = \langle S, \leqslant, g, v\rangle$ where

- $\langle S, \leqslant, g\rangle$ is a poset with an order-reversing involution g;
- For each $s \in S$, $v(s)$ is a valuation into $\mathbf{2} = \{true, false\}$ on the set of propositional letters satisfying the following *persistency* condition: for any $s_1, s_2 \in S$ and propositional letter p,

$$\text{if } s_1 \leqslant s_2 \text{ and } v(s_1)(p) = true, \quad \text{then } v(s_2)(p) = true.$$

It is a Boolean structure if $\leqslant$ is the identity relation and g is the identity function on S.

From the persistency condition, we may immediately derive a "reverse" persistency condition as follows: for any $s_1, s_2 \in S$ and propositional letter p,

$$\text{if } s_1 \leqslant s_2 \text{ and } v\big(g(s_2)\big)(p) = true, \quad \text{then } v\big(g(s_1)\big)(p) = true.$$

A satisfaction relation between states and formulas can be defined exactly as in standard propositional logic except the following clause for negation

$$S, s \models \sim\varphi \quad \text{if } S, g(s) \not\models \varphi.$$

Actually the persistency condition for the valuation is satisfied by all formulas, as shown in the following proposition.

Lemma 4.15. *Let $\langle S, \leqslant, g, v\rangle$ be a Routley structure. If $s_1, s_2 \in S$ and $s_1 \leqslant s_2$, then, for any formula φ,*

1. $S, s_1 \models \varphi$ *implies* $S, s_2 \models \varphi$;
2. $S, g(s_2) \models \varphi$ *implies* $S, g(s_1) \models \varphi$.

Remark 4.16. The persistency condition in a Routley structure is quite similar to that in the Kripke semantics for intuitionistic logic [64]. Each pair $(s, g(s))$ of a state s and its adjunct can be seen as a pair of knowledge bases. s is the knowledge base *consisting of* true facts and $g(s)$ is the knowledge base *for* the false facts. Here we take the *closed-world assumption* [46] for $g(s)$ in the sense that, if a proposition is not implied in $g(s)$, then the negation of this proposition is implied at s. This pair of knowledge evolves in the course of time. The knowledge base of true facts expands at every later stage while the knowledge base for false facts decreases. So the essential difference from those in Belnap structures is that the knowledge of true facts and that for false facts are rather dual to each other than independent of each other as in Belnap structures.

If the second (reverse) persistency condition is replaced by a new *adjunct but independent* valuation v^* which is defined as $v^*(s) = v(g(s))$, then we may also define a Routley structure as a poset $\langle S, \leqslant g\rangle$ with an order-reversing involution g and two "independent" but adjunct-to-each-other valuations v and v^* satisfying the following two persistency conditions:

- if $s_1 \leqslant s_2$, $v(s_1)(p) = true$ implies $v(s_2)(p) = true$ for all propositional letters p;
- if $s_1 \leqslant s_2$, $v^*(s_2)(p) = true$ implies $v^*(s_1)(p) = true$ for all propositional letters p.

The new valuation v^* is for false facts. So the semantics with this new type of Routley structures is the same as the above except that for the negated formulas:

$$(S, v), s \models \sim\varphi \quad \text{if } \big(S, v^*\big), s \not\models \varphi.$$

The adjunct valuation is used to decouple the semantics to interpret negation. Since there is a straightforward interpretation between this semantics and the above one, they are equivalent. So each possibility in the epistemic frame for belief functions on Routley structures (Definition 4.19) consists of a *pair* of valuations. It would be interesting to compare this pair with the two support relations $\models_T$ and $\models_F$ in the valuation for the four-valued logic. Note that, although Routley semantics and Belnap semantics are different, the notion of satisfiability in the two semantics are equivalent (Proposition 9.1 in [20]).

The connection (duality) between Routley structures and de Morgan lattices can be analyzed as follows. Define an equivalence relation $\asymp$ on Φ,

$$\varphi_1 \asymp \varphi_2 \quad \text{iff} \quad \varphi_1 \vdash \varphi_2 \text{ and } \varphi_2 \vdash \varphi_1.$$

Let $\Phi/{\asymp}$ denote the set of $\asymp$-equivalence classes $[\varphi]_\asymp$. Now we define the operations on this set as follows:

- $[\varphi_1]_\asymp \wedge_\asymp [\varphi_2]_\asymp = [\varphi_1 \wedge \varphi_2]_\asymp$;
- $[\varphi_1]_\asymp \vee_\asymp [\varphi_2]_\asymp = [\varphi_1 \vee \varphi_2]_\asymp$;
- $\sim_\asymp [\varphi]_\asymp = [\sim\varphi]_\asymp$.

It is easy to check that these operations are well-defined. Such a defined algebra $\langle \Phi/_{\asymp}, \wedge_{\asymp}, \vee_{\asymp}, \sim_{\asymp}\rangle$ is the Lindenbaum algebra on $\Phi/_{\asymp}$ and is actually a de Morgan lattice.

With respect to Φ, we define the *canonical* Routley structure as follows.

- S_Φ is the set of all prime filters in $\Phi/\asymp$;
- $g_\Phi : S_\Phi \to S_\Phi$ is defined: for each $F \in S_\Phi$, $g_\Phi(F) = \Phi/_{\asymp} \setminus \sim_{\asymp} F$ where $\sim_{\asymp} F = \{\sim_{\asymp} [\varphi]_{\asymp} : [\varphi]_{\asymp} \in F\}$, which is a prime filer;
- $v_\Phi(F)(p) = \mathit{true}$ if $[p]_{\asymp} \in F$ for any $F \in S_\Phi$.

It is easy to check that $\langle S_\Phi, \leqslant_\Phi, g_\Phi\rangle$ is a poset with the order-reversing involution g_Φ where $\leqslant_\Phi$ is the subset relation. If Φ is finite, according to the dual form of the representation theorem for finite de Morgan lattices, the Lindenbaum algebra $\langle \Phi/_{\asymp}, \wedge_{\asymp}, \vee_{\asymp}, \sim_{\asymp}\rangle$ is isomorphic to the concrete lattice of filters in the poset $\langle S_\Phi, \leqslant_\Phi, g_\Phi\rangle$ which underlies the canonical Routley structure.

Definition 4.17. A formula φ *logically implies* a formula ψ with respect to the class of Routley structures (denoted as $\varphi \models^R \psi$) if, for any Routley structure $S = \langle S, g, v\rangle$, $S, s \models \varphi$ implies $S, s \models \psi$.

The difference of this definition from that for Belnap structures (Definition 4.8) is that we don't need to consider the "negative side". Actually the logical implication relation in the class of Routley structures coincides with the above consequence relation $\vdash_{\mathbf{R}_{fde}}$.

Theorem 4.18. *(See Theorem 7 in [17].) For any two formulas φ and ψ in Φ,*

$$\varphi \vdash \psi \quad \textit{iff} \quad \varphi \models^R \psi.$$

Definition 4.19. A *Dempster–Shafer structure* B (*DS*-structure for short) on a de Morgan lattice is a tuple $\langle S, \leqslant, g, v, \mathit{Bel}\rangle$ where

- $\mathrm{S} = \langle S, \leqslant, g, v\rangle$ is a Routley structure;
- Bel is a belief function on $F(S) = \{I \subseteq S\colon\ I \text{ is a filter in } S\}$.

For the Routley structure $\langle S, \leqslant, g, v\rangle$, $[[\varphi]]_{\mathrm{S}}$ denotes the set of states where φ is satisfied, i.e., $[[\varphi]]_{\mathrm{S}} = \{s \in S\colon\ S, s \models \varphi\}$, which can be shown to be a filter in S. So the belief in the formula φ is defined to be $\mathit{Bel}([[\varphi]]_S)$. Compared with belief functions on bilattices, belief functions on de Morgan lattices are much simpler in the following sense. In a Belnap structure B, $[[\varphi]]^B_T \subseteq [[\psi]]^B_T$ does not necessarily imply that $[[\varphi]]^B_F \supseteq [[\psi]]^B_F$; in other words, the fact the agent have more reasons to consider ψ as true than φ does not imply that the agent have more reasons to consider φ false than ψ. However, this is not the case in a Routley structure. So, in the definition of belief functions for de Morgan lattices, we don't need to consider the second parameter of the set of facts that falsify the proposition. Our definition here agrees with the definition of *DS*-structures in Definition 3.20.

Definition 4.20. For any two formulas φ and ψ in Φ, φ *probabilistically entails* ψ with respect to the class of de Morgan lattices (denoted as $\varphi \models^R_{DS} \psi$) if, for any *DS*-structure $S = \langle S, \leqslant, g, v, \mathit{Bel}\rangle$, $\mathit{Bel}([[\varphi]]_S) \leqslant \mathit{Bel}([[\psi]]_S)$.

The following theorem tells us that the deductive system $\mathbf{R}_{fde}$ provides a sound and complete system for both logical implication but also probabilistic entailment with respect to both the class of Belnap structures and that of Routley structures.

Theorem 4.21. *For any formulas φ and ψ in Φ, $\varphi \vdash \psi$ if and only if $\varphi \models^R_{DS} \psi$.*

Proof. The proof is similar to that for Theorem 4.12. □

The following theorem tells us that the logical implication problem with respect to the class of Routley-structures (Belnap structures) is as hard as logical implication in standard propositional logic.

Theorem 4.22. *The complexity of deciding logical implication with respect to the class of Routley structures* (*Belnap structures*) *is co-NP-complete.*

Proof. The proof can be easily adapted from that of the similar problem in Section 8 in [20] (also [63,37]). □

5. Reasoning about belief functions for first-degree entailments

In this section, we provide a sound and complete deductive system for reasoning about belief functions for first degree entailments and show that the satisfiability problem of a belief formula with respect to the corresponding class of Dempster–Shafer structures is *NP*-complete.

In this part, we adapt the deductive machineries in [19,18] to provide a sound and complete axiomatization for reasoning about belief functions over Routley structures (Belnap structures).

Definition 5.1. For the above given set Φ of formulas, a *term* is an expression of the form $a_1 bel(\varphi_1) + a_2 bel(\varphi_2) + \cdots + a_k bel(\varphi_k)$, where $a_1, a_2, \ldots, a_k$ are integers, *bel* is the belief function symbol and $\varphi_1, \varphi_2, \ldots, \varphi_k$ are formulas in Φ. *A basic belief formula* is one of the form $t \geqslant b$, where t is a term and b is an integer. A *belief formula* is a Boolean combination of basic belief formulas. We can always allow rational numbers in our formulas as abbreviations for the formula that would be obtained by clearing the dominator. And other derived relations $=, \leqslant, <$ and $>$ can be defined as usual.

Definition 5.2. Given a *DS*-structure $B = \langle S, \leqslant, g, v, Bel\rangle$ on a Routley structure $S := \langle S, \leqslant, g, v\rangle$ and a basic belief formula $f := a_1 bel(\varphi_1) + a_2 bel(\varphi_2) + \cdots + a_k bel(\varphi_k) \geqslant b$, B *satisfies* f (denoted as $B \models f$) if

$$a_1 Bel\big([[\varphi_1]]_S\big) + a_2 Bel\big([[\varphi_2]]_S\big) + \cdots + a_k Bel\big([[\varphi_k]]_S\big) \geqslant b.$$

We then extend the above $\models$ in the obvious way to all belief formulas. Let $\mathcal{B}$ be a class of Dempster–Shafer structures. A belief formula f' is *satisfiable with respect to* $\mathcal{B}$ if it is satisfied in some $B \in \mathcal{B}$. It is *valid with respect to* $\mathcal{B}$ if $B \models f$ for all $B \in \mathcal{B}$.

Definition 5.3. Given a basic belief formula $f := a_1 bel(\varphi_1) + a_2 bel(\varphi_2) + \cdots + a_k bel(\varphi_k) \geqslant b$ and a probability structure $M = \langle S, L, L', v, \mu\rangle$ on a Routley structure $S := \langle S, \leqslant, g, v\rangle$ where $L = J(S)$, L' is a sublattice of L and μ is a probability function on L', M *satisfies* f (denoted as $M \models f$) if

$$a_1 \mu_*\big([[\varphi_1]]_S\big) + a_2 \mu_*\big([[\varphi_2]]_S\big) + \cdots + a_k \mu_*\big([[\varphi_k]]_S\big) \geqslant b.$$

We then extend the above $\models$ in the obvious way to all belief formulas. Let $\mathcal{M}$ be a class of probability structures. A belief formula f' is *satisfiable with respect to* $\mathcal{M}$ if it is satisfied in some $M \in \mathcal{M}$. It is *valid with respect to* $\mathcal{M}$ if $M \models f$ for all $M \in \mathcal{M}$.

The axiomatization $\mathcal{B}_{fde}$ of reasoning about belief functions for first degree entailments consists of three parts: the first-degree entailments, reasoning about linear inequalities and reasoning about belief functions.

1. First-degree entailments
   - The complete system $\vdash_{\mathbf{R}_{fde}}$ of first degree entailment is provided in last section.
2. Reasoning about linear inequalities
   (a) $a_1 bel(\varphi_1) + a_2 bel(\varphi_2) + \cdots + a_k bel(\varphi_k) \geqslant b$ iff $a_1 bel(\varphi_1) + a_2 bel(\varphi_2) + \cdots + a_k bel(\varphi_k) + 0 bel(\varphi_{k+1}) \geqslant b$;
   (b) $a_1 bel(\varphi_1) + a_2 bel(\varphi_2) + \cdots + a_k bel(\varphi_k) \geqslant b$ iff $a_{j_1} bel(\varphi_1) + a_{j_2} bel(\varphi_2) + \cdots + a_{j_k} bel(\varphi_k) \geqslant b$ where $j_1, j_2, \ldots, j_k$ is a permutation of $1, 2, \ldots, k$;
   (c) $a_1 bel(\varphi_1) + a_2 bel(\varphi_2) + \cdots + a_k bel(\varphi_k) \geqslant b$ iff $ca_1 bel(\varphi_1) + ca_2 bel(\varphi_2) + \cdots + ca_k bel(\varphi_k) \geqslant cb$ where $c > 0$;
   (d) $(a_1 + a'_1) bel(\varphi_1) + (a_2 + a'_2) bel(\varphi_2) + \cdots + (a_k + a'_k) bel(\varphi_k) \geqslant b + b'$ if $a_1 bel(\varphi_1) + a_2 bel(\varphi_2) + \cdots + a_k bel(\varphi_k) \geqslant b$ and $a'_1 bel(\varphi_1) + a'_2 bel(\varphi_2) + \cdots + a'_k bel(\varphi_k) \geqslant b'$;
   (e) Either $t \geqslant b$ or $t \leqslant b$ where t is a term;
   (f) $t \geqslant b$ implies $t > b'$ where t is a term and $b' < b$.

   Let AX_{Iq} denote this deductive system reasoning about linear inequalities, which is shown to be complete [19].
3. Reasoning about belief functions
   (a) $bel(\varphi) \geqslant 0$ for all formulas $\varphi \in \Phi$;
   (b) $bel(\top) = 1$;
   (c) $bel(\bot) = 0$;
   (d) $bel(\varphi_1 \vee \varphi_2 \vee \cdots \vee \varphi_n) \geqslant \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} bel(\wedge_{i \in I} \varphi_i)$;
   (e) $bel(\varphi) \leqslant bel(\psi)$ if $\varphi \vdash_{\mathbf{R}_{fde}} \psi$.

Note that the principle 3. (e) is the connection of reasoning about belief functions to first degree entailments.

Theorem 5.4. *$\mathcal{B}_{fde}$ is a sound and complete axiomatization of belief formulas with respect to the class of DS-structures.*

Proof. The soundness is trivial except the principle 3(e), which follows from Theorem 4.21 (and Theorem 4.12). Here we mainly focus on the completeness proof. The completeness proof is similar to that of the system AX with respect to the class of general probability structures in [19] except that we need to replace propositional reasoning with reasoning about first degree entailments.

Fix a belief formula f. Without loss of generality, we may assume that f is a conjunction of basic belief formulas and negated basic belief formulas.[4] Let $\mathcal{P}_f = \{p_1, \ldots, p_k\}$ be the set of all the propositional letters occurring in f and $\Phi_{\mathcal{P}_f}$ be the set of formulas in Φ with propositional letters occurring in $\mathcal{P}_f$. Recall that two formulas are $\asymp$-equivalent iff they are deductively equivalent in $\mathbf{R}_{fde}$. As in the duality between de Morgan lattices and Routley structures, we define the Lindenbaum algebra $\langle \Phi_{\mathcal{P}_f}/\asymp, \wedge_\asymp, \vee_\asymp, \sim_\asymp \rangle$ as follows:

- $\Phi_{\mathcal{P}_f}/\asymp$ is the set of all $\asymp$-equivalence classes $[\varphi]_\asymp$ for all $\varphi \in \Phi_{\mathcal{P}_f}$, which can be easily shown to be finite;
- for any two formulas φ, ψ whose propositional letters are in $\mathcal{P}_f$
  1. $[\varphi]_\asymp \wedge_\asymp [\psi]_\asymp = [\varphi \wedge \psi]_\asymp$;
  2. $[\varphi]_\asymp \vee_\asymp [\psi]_\asymp = [\varphi \vee \psi]_\asymp$;
  3. $\sim_\asymp [\varphi]_\asymp = [\sim \varphi]_\asymp$.

It is easy to check that $\langle \Phi_{\mathcal{P}_f}/\asymp, \wedge_\asymp, \vee_\asymp, \sim_\asymp \rangle$ is a de Morgan lattice and it is isomorphic to the concrete lattice of the poset underlying the dual canonical Routley structure which is defined as follows:

- $S_{\mathcal{P}_f}$ is the set of all prime filters in $\Phi_{\mathcal{P}_f}/\asymp$;
- $g_{\mathcal{P}_f} : S_{\mathcal{P}_f} \to S_{\mathcal{P}_f}$ is defined: for each $F \in S_{\mathcal{P}_f}$, $g(F) = S_{\mathcal{P}_f} \setminus \sim_\asymp F$ where $\sim_\asymp F = \{\sim_\asymp [\varphi]_\asymp : [\varphi]_\asymp \in F\}$, which is a prime filer;
- $v_{\Phi_{\mathcal{P}_f}}(F)(p) = true$ if $[p]_\asymp \in F$ for any $F \in S_{\Phi_{\mathcal{P}_f}}$.

In other words, $\Phi_{\mathcal{P}_f}/\asymp$ is isomorphic to $J(S_{\Phi_{\mathcal{P}_f}})$. Let h denote this isomorphic function. In the following, we won't distinguish between a formula and its equivalence class. Enumerate all formulas $\varphi \in \Phi_{\mathcal{P}_f}$:

$$\varphi_0, \varphi_1, \ldots, \varphi_m \quad \text{where } \varphi_0 \equiv \bot \text{ and } \varphi_m \equiv \top.$$

Let $x_0, x_1, \ldots, x_m$ be the variables that correspond to these formulas. Now we turn f into a group of linear inequalities. For each conjunct of f, say, $a_1 bel(\varphi_1) + \cdots + a_k bel(\varphi_k) \geqslant b$, we replace each $bel(\varphi_i)$ by the variable, say, x_i that corresponds to φ_i and get

$$a_1 x_1 + \cdots + a_k x_k \geqslant b.$$

By collecting all these inequalities, we obtain a group $\hat{f}$ of linear inequalities that captures the linear relations of beliefs on different formulas in f. For each formula $\varphi_i \in \Phi_{\mathcal{P}_f}$, we have a new inequality:

$$\sum_{[h(\varphi_j), h(\varphi_i)] \text{ is a Boolean algebra}} (-1)^{|h(\varphi_i) \setminus h(\varphi_j)|} x_j \geqslant 0, \tag{10}$$

where $h(\varphi_j), h(\varphi_i)$ are order ideals in $S_{\mathcal{P}_f}$ and $[h(\varphi_j), h(\varphi_i)]$ is an interval in $J(S_{\mathcal{P}_f})$. Intuitively, it says that the mass assignment at each formula φ_i is nonnegative. The principle guiding the construction is Theorem 3.2. Let $\widehat{\Phi_{\mathcal{P}_f}}$ denote the set of all linear inequalities of this kind.

Claim 6. *If f is satisfiable, then $\widehat{\Phi_{\mathcal{P}_f}}$ has a solution.*

Assume that f is satisfied in a *DS*-structure $S = \langle S, \leqslant, g, v, Bel \rangle$. Let m be the corresponding mass assignment. Since each term on the left side of the inequality 10 is actually the mass assignment $m([[\varphi_i]]_S)$, it is nonnegative according to Theorem 3.3. So $x_i = Bel([[\varphi_i]]_S)(0 \leqslant i \leqslant m)$ is a solution of $\widehat{\Phi_{\mathcal{P}_f}}$ and hence we have proved the claim.

Claim 7. *f is satisfiable if and only if the following group of inequalities has a solution*:

$$\widetilde{f} := \hat{f} \cup \widehat{\Phi_{\mathcal{P}_f}} \cup \{x_0 = 0, x_m = 1\}.$$

From the above claim, the left-to-right direction follows immediately. Assume that $\widetilde{f}$ has a solution. Now we define $Bel(\varphi_i) = x_i (0 \leqslant i \leqslant m)$. From the constraints in $\widetilde{f}$ and Theorem 3.2, it follows that Bel is a belief function on the canonical Routley structure. Moreover, the constraints in $\hat{f}$ guarantees that f is satisfied in this canonical structure.

---

[4] If f is not a conjunction of this form, it is equivalent to a disjunction of conjunctions of this form. Then we consider any satisfiable conjunct if f is satisfiable.

Now we are back to show the theorem. Assume that f is unsatisfiable. It follows from the above claim that $\widetilde{f}$ does not have a solution. Since AX_{Iq} is a complete axiomatization for reasoning about linear inequalities, $\widetilde{f}$ is inconsistent in AX_{Iq}. Since $\widehat{\Phi_{\mathcal{P}_f}} \cup \{x_0 = 0, x_m = 1\}$ correspond to Axioms 3(b)(c)(d) in $\mathcal{B}_{fde}$, $\hat{f}$ is not consistent in $\mathcal{B}_{fde}$. This is equivalent to say that f is not consistent in $\mathcal{B}_{fde}$. So we have finished the whole proof of the theorem. $\square$

Let $|f|$ denote the length of the belief formula f which is the number of symbols required to write f, where we count the length of each coefficient as 1 and $\|f\|$ be the length of the longest coefficient appearing in f, when written in binary.

Lemma 5.5. *Let f be a belief formula that is satisfied in some DS-structure on a Routley structure (a Belnap structure). Then f is satisfied in a probability structure on a Boolean algebra with at most $|f|^2$ states, with a basis of size at most $|f|$, where the probability assigned to each member of the basis is a rational number with size $O(|f|\|f\| + |f|\log(|f|))$.*

Proof. From the results in Theorem 3.22 (particular from Corollary 3.23), we know that each *DS*-structure on a distributive lattice is equivalent to a probability structure on a Boolean algebra. Actually the whole argument there works to show that each *DS* structure $B := \langle S, \leqslant, g, v, Bel\rangle$ on a de Morgan lattice $J(S)$ (or on a Routley structure $\langle S, \leqslant, g, v\rangle$) is equivalent to a probability structure $M := \langle S', L, L', v', \mu'\rangle$ where $L = 2^{S'}$, L' is a Boolean algebra and μ' is a probability function on L' in the sense that, for any formula $\varphi \in \Phi$,

$$Bel\big([[\varphi]]_B\big) = \mu'_*\big([[\varphi]]_M\big).$$

Note that, compared to Theorems 3.22 (and Corollary 3.23), our equivalence here is stronger because the language Φ here includes the negation $\sim$ while Φ_0 in Theorems 3.22 (and Corollary 3.23) does not. Then the lemma follows immediately from Theorem 3.10 in [19]. $\square$

Theorem 5.6. *The time complexity of deciding whether a belief formula is satisfiable with respect to the class of DS-structures is NP-complete.*

Proof. Since every probability structure is also a *DS*-structure, the theorem follows from Theorem 3.11 in [19]. The interested reader may refer to [19] for the details. $\square$

We may define logical implication for belief formulas as usual. The above theorem tells us that the complexity of the logical implication problem with respect to the class of *DS*-structures on Routley structures is the same as that with respect to the class of Routley-structures (Theorem 4.22) and hence is not affected by the expansion of the propositional language with belief functions.

6. Discussion and conclusion

The main contribution of this paper is the extension of Dempster–Shafer theory of belief functions on Boolean algebras to the setting of distributive lattices and show that many intuitively appealing properties in the theory are transposed to this more general case. We use this more general theory to provide a framework for reasoning about belief functions in a deductive approach on many non-classical formalisms in artificial intelligence which assume a setting of distributive lattices. As an illustration of this deductive approach, we apply this general theory to a non-classical formalism in the foundations of Knowledge Representation the first-degree-entailment of relevance logic **R**.

This kind of integration of uncertainty measures such as belief functions and logics for knowledge representation is an important approach in reasoning under uncertain and imperfect knowledge in artificial intelligence [43]. Logics including many non-classical logics play a central role in the task of knowledge representation in artificial intelligence [42]. And each of these logics was intended for some particular focus. On the other hand, uncertainty measures are usually to deal with uncertainty in information [31]. However, non-classical logics are not expressive enough to express uncertainty in a *gradual* way and uncertainty measures such as belief functions are not enough to handling imperfect information. This is another motivation to combine uncertainty measures with non-classical logics in addition to that mentioned at the beginning of Section 4. Bernard and Lang applied possibility theory to non-classical logics especially paraconsistent logics and showed how to reason under uncertain and inconsistent information [6]. Saffiotti proposed a formal framework to integrate logics for knowledge including first order logic and belief functions [49,48,50]. But none of these papers has touched any issue about the mathematical foundation and computational complexity behind the theory of belief functions, just as we have done in this paper. Here, we provide a sound and complete deductive system for reasoning about belief functions for a simple epistemic logic the first-degree-entailment fragment of relevance logic **R** through different duality theorems between algebraic semantics and logic. This axiomatization can be used to show how to deduce one belief of some events from beliefs of others. Moreover, we have given the complexity result of the satisfiability problem of belief formulas in this kind of non-classical settings. The deductive approach for belief functions for first degree entailments can be applied to other non-classical formalisms $\mathbb{L}$ that assume a setting of distributive lattices. The axiomatization $\mathcal{B}_{\mathbb{L}}$ for reasoning about belief functions on $\mathbb{L}$ is simply obtained from the axiomatization $\mathcal{B}_{fde}$ by replacing Part 1 of first degree entailments by $\mathbb{L}$ and the implication $\varphi \vdash_{R_{fde}} \psi$ in Principle 3 (e) in Part 3 by $\varphi \vdash_{\mathbb{L}} \psi$. In particular, our deductive approach also covers the formalism developed by Bernard and Lang in [6].

In the literature, there are many similar kinds of integration of belief functions with other non-classical formalisms in Knowledge Representation. In the theory of evidential reasoning on fuzzy events [68,55,56,66], a belief functions is not defined in terms of mass assignments as in our paper but is defined to be a set function which is both monotonic and ∞-monotonic. Moreover, it is defined *indirectly* either as an expectation of some membership function with respect to some other given belief function on the referential space [55,56] or as a probability function through some generalized compatibility relation [66,68]. In the theory of bipolar belief pairs on vague propositions [36], a belief function is defined *indirectly* as a probability function on some supervaluation pair and hence not in terms of mass assignments. Usually it is difficult to do evidential reasoning with this kind of definitions *without* reference to mass assignments because Dempster's rule of combination is defined in terms of them. In contrast, in the theory developed in this paper, we define belief functions *in* terms of mass assignment and moreover, characterize them without reference to mass assignments. So, our theory of belief functions on distributive lattices provides both a more transparent framework and a uniform foundation to do evidential reasoning in the above settings.

This paper is a theoretical framework for our ongoing project to apply belief functions to multi-valued reasoning in AI [25], decision-making with bipolar information [14] and adding belief annotations to databases [24] with incomplete and/or inconsistent knowledge where structures of interest are usually assumed to be no more Boolean but distributive. Our immediate task is to investigate information fusion in belief functions for distributive *bilattice reasoning* [25].

The perspective of this paper mainly views belief functions as generalized probability functions. Now we are also working on another different perspective regarding belief functions for evidence and are doing evidential reasoning for non-classical formalisms. A further project, which is probably more of theoretical interest, is to develop a theory of belief functions for quantum structures (not necessarily distributive) which would relate quantum computation and AI [67].

Acknowledgements

This research is partly supported by Technische Universität Dortmund (Informatik 10), by the Key Project for basic research from the Ministry of Science and Technology of China (Grant Number: 2012CB316205), by NSF of China (Grant Number: 60905036), and by German Academic Exchange Service (DAAD Codenumber: A/10/00803).

I am more than grateful to Professor Ernst-Erich Doberkat for his support of my research these years. He carefully read the whole paper and helped me correct many mistakes and errors in *all* earlier drafts of this paper. Moreover, he gave many invaluable suggestions and advice to improve the presentation. Especially he brought Lemmas 2.5 and 2.6, which have greatly streamlined our proofs in Sections 2 and 3. I would like to thank Ingo Battenfeld for carefully reading the last draft of this paper and providing a nice counterexample about Definition 3.10, and for helpful discussions about lattices in other related areas in theoretical computer science. Also I want to thank Professor Mike Dunn for helpful discussions about duality theorem of de Morgan lattices and Professor Mingsheng Ying about belief functions on fuzzy events.

I am grateful to the referees for pointing out some mistakes in the early draft of the paper and for their helpful comments to improve the presentation.

Appendix A. Some properties of distributive lattices

A lattice L is called a *modular lattice* if its elements satisfy the following condition:

$$\text{if } x \leqslant z, \quad \text{then } x \vee (y \wedge z) \leqslant (x \vee y) \wedge z. \tag{11}$$

With the following proposition, we know that a belief function can be defined on any lattice while probability functions can live only on distributive lattices.

Proposition A.1. *(See Lemma 3 in [26].) Let L be a lattice. Then*

1. *L is a modular if and only if it admits a strictly monotone* 2*-valuation.*
2. *L is distributive if and only if it is modular and every strictly monotone* 2*-valuation on L is a* 3*-valuation.*
3. *L is distributive if and only if it admits a strictly monotone* 3*-valuation.*
4. *L is distributive if and only if it is modular and every strictly monotone* 2*-valuation on L is an* ∞*-valuation.*

Given a poset P, $J(P)$ denotes the lattice of order ideals of P. The lattice operations $\vee$ and $\wedge$ on order ideals are just ordinary union and intersection (as subsets of P). The interested reader may refer to Section 3.4 in [61] (and Chapter IX in [9]) for the detailed proofs of the following propositions.

Theorem A.2. *Let L be a finite distributive lattice. Then there is a unique (up to isomorphism) finite poset P for which* $L \cong J(P)$.

This theorem is so important in combinatorial aspects of distributive lattices that it is called by Stanley the fundamental theorem for finite distributive lattices [61].

Lemma A.3. *Let $I \leqslant I'$ in the distributive lattice $J(P)$. The interval $[I, I']$ is isomorphic to $J(I' \setminus I)$, where $I' \setminus I$ is regarded as an induced subposet of P. In particular, $[I, I']$ is a distributive lattice.*

Theorem A.4. *If P is an n-element poset, then $J(P)$ is graded of rank n. Moreover, the rank $\rho(I)$ of $I \in J(P)$ is just the cardinality of I.*

References

[1] B. Anderson, N. Belnap, Entailment: the Logic of Relevance and Necessity, Princeton University Press, New Jersey, 1975.
[2] O. Arieli, A. Avron, The value of the four values, Artif. Intell. 102 (1) (1998) 97–141.
[3] O. Arieli, A. Avron, A. Zamansky, What is an ideal logic for reasoning with inconsistency?, in: IJCAI, 2011, pp. 706–711.
[4] J.-P. Barthelemy, Monotone functions on finite lattices: An ordinal approach to capacities, belief and necessity functions, in: J. Fodor, B. Baets, P. Perny (Eds.), Preferences and Decisions Under Incomplete Knowledge, 2000, pp. 195–208.
[5] N. Belnap, A useful four-valued logic, in: J.M. Dunn, G. Epstein (Eds.), Modern Uses of Multi-Valued Logic, Reidel, Dordrecht, 1977, pp. 8–37.
[6] P. Besnard, J. Lang, Possibility and necessity functions over non-classical logics, in: David Poole, Ramon López de Mántaras (Eds.), UAI, Morgan Kaufmann, 1994, pp. 69–76.
[7] A. Bialynicki-Birula, H. Rasiowa, On the representation of quasi-boolean algebras, Bull. Acad. Pol. Sci. 5 (1957) 259–261.
[8] P. Billingsley, Measure and Probability, third edition, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., 1995.
[9] G. Birkhoff, Lattice Theory, 3rd edition, American Mathematical Society, 1967.
[10] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[11] A.P. Dempster, A generalization of bayesian inference, J. Roy. Statist. Soc. 73 (363) (1968) 205–247.
[12] T. Denoeux, M. Masson, Evidential reasoning in large partially ordered sets – application to multi-label classification, ensemble clustering and preference aggregation, Ann. OR 195 (1) (2012) 135–161.
[13] T. Denoeux, Z. Younes, F. Abdallah, Representing uncertainty on set-valued variables using belief functions, Artif. Intell. 174 (7–8) (2010) 479–499.
[14] D. Dubois, H. Prade, An introduction to bipolar representations of information and preference, Int. J. Intell. Syst. 23 (8) (2008) 866–877.
[15] M. Dunn, Intuitive semantics for first degree entailments and coupled trees, Philos. Stud. 29 (1976) 149–168.
[16] M. Dunn, Relevance logic and entailment, in: D. Gabbay, F. Guenthner (Eds.), Handbook of Philosophical Logics, D. Reidel Publishing Company, Dordrecht, 1986, pp. 117–224.
[17] M. Dunn, Partiality and its dual, Stud. Log. 66 (1) (2000) 5–40.
[18] R. Fagin, J. Halpern, Uncertainty, belief, and probability, Comput. Intell. 7 (1991) 160–173.
[19] R. Fagin, J. Halpern, N. Megiddo, A logic for reasoning about probabilities, Inform. and Comput. 87 (1990) 78–128.
[20] R. Fagin, J. Halpern, M. Vardi, A nonstandard approach to the logical omniscience problem, Artif. Intell. 79 (2) (1995) 203–240.
[21] M. Fitting, Bilattices in logic programming, in: ISMVL, 1990, pp. 238–246.
[22] M. Fitting, Bilattices and the semantics of logic programming, J. Log. Program. 11 (1&2) (1991) 91–116.
[23] M. Fitting, Kleene's three valued logics and their children, Fund. Inform. 20 (1/2/3) (1994) 113–131.
[24] W. Gatterbauer, M. Balazinska, N. Khoussainova, D. Suciu, Believe it or not: Adding belief annotations to databases, PVLDB 2 (1) (2009) 1–12.
[25] M. Ginsberg, Multivalued logics: a uniform approach to reasoning in artificial intelligence, Comput. Intell. 4 (1988) 265–316.
[26] M. Grabisch, Belief functions on lattices, Int. J. Intell. Syst. 24 (1) (2009) 76–95.
[27] M. Grabisch, The lattice of embedded subsets, Discrete Appl. Math. 158 (5) (2010) 479–488.
[28] M. Grabisch, C. Labreuche, The symmetric and asymmetric Choquet integrals on finite spaces for decision making, Statist. Papers 43 (2002) 37–52.
[29] M. Grabisch, C. Labreuche, Bi-capacities – I: definition, Möbius transform and interaction, Fuzzy Sets Syst. 151 (2) (2005) 211–236.
[30] P. Halmos, Measure Theory, Graduate Text in Mathematics, Van Nostrand, New York, 1950.
[31] J. Halpern, Reasoning About Uncertainty, MIT Press, 2005.
[32] J. Halpern, R. Fagin, Two views of belief: Belief as generalized probability and belief as evidence, Artif. Intell. 54 (2) (1992) 275–317.
[33] M. Kifer, E.L. Lozinskii, A logic for reasoning with inconsistency, J. Automat. Reason. 9 (2) (1992) 179–215.
[34] S. Konieczny, P. Marquis, P. Besnard, Bipolarity in bilattice logics, Int. J. Intell. Syst. 23 (10) (2008) 1046–1061.
[35] S. Kripke, Semantical analysis of intuitionistic logic, in: J.N. Crossley, M.A.E. Dummett (Eds.), Formal Systems and Recursive Functions, in: Studies in Logic and Foundations of Mathematics, Reidel, Dordrecht, 1965, pp. 92–130.
[36] J. Lawry, Y. Tang, On truth-gaps, bipolar belief and the assertability of vague propositions, Artif. Intell. 191–192 (2012) 20–41.
[37] H. Levesque, A logic of implicit and explicit belief, in: AAAI, 1984, pp. 198–202.
[38] J. Lin, A semantics for reasoning consistently in the presence of inconsistency, Artif. Intell. 86 (1) (1996) 75–95.
[39] M. Masson, T. Denoeux, Ensemble clustering in the belief functions framework, Int. J. Approx. Reason. 52 (1) (2011) 92–109.
[40] B. Mobasher, D. Pigozzi, G. Slutzki, G. Voutsadakis, A duality theory for bilattices, Algebra Univers. 43 (2000) 109–125.
[41] A. Nerode, Some stone spaces and recursion theory, Duke Math. J. 26 (1959) 397–406.
[42] N. Nilsson, Logic and artificial intelligence, Artif. Intell. 47 (1–3) (1991) 31–56.
[43] S. Parsons, Current approaches to handling imperfect information in data and knowledge bases, IEEE Trans. Knowl. Data Eng. 8 (3) (1996) 353–372.
[44] G. Priest, The logic of paradox, J. Philos. Logic 9 (1979) 415–435.
[45] H. Priestley, Representation of distributive lattices by means of ordered stone spaces, Bull. Lond. Math. Soc. 2 (1970) 186–190.
[46] R. Reiter, On closed world database, in: H. Gallaire, J. Minker (Eds.), Logic and Database, Plenum Press, New York, 1978, pp. 55–76.
[47] G.C. Rota, On the foundations of combinatorial theory i: Theory of Möbius functions, Z. Wahrscheinlichkeitstheor. Verw. Geb. 2 (1964) 340–368.
[48] A. Saffiotti, A hybrid framework for representing uncertain knowledge, in: Howard E. Shrobe, Thomas G. Dietterich, William R. Swartout (Eds.), AAAI, AAAI Press/The MIT Press, 1990, pp. 653–658.
[49] A. Saffiotti, Using Dempster–Shafer theory in knowledge representation, in: Piero P. Bonissone, Max Henrion, Laveen N. Kanal, John F. Lemmer (Eds.), UAI, Elsevier, 1990, pp. 417–434.
[50] A. Saffiotti, A belief-function logic, in: William R. Swartout (Ed.), AAAI, AAAI Press/The MIT Press, 1992, pp. 642–647.
[51] L. Savage, The Foundations of Statistics, Dover Publications, Inc., 1972.
[52] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[53] G. Shafer, Allocation of probability, Ann. Probab. 7 (5) (1979) 827–839.
[54] K. Sim, Bilattices and reasoning in artificial intelligence: Concepts and foundations, Artif. Intell. Rev. 15 (3) (2001) 219–240.
[55] P. Smets, The degree of belief in a fuzzy event, Inform. Sci. 25 (1) (1981) 1–19.

[56] P. Smets, Medical diagnosis: fuzzy sets and degree of belief, Fuzzy Sets Syst. 5 (1981) 259–266.
[57] P. Smets, The canonical decomposition of a weighted belief, in: IJCAI, Morgan Kaufmann, 1995, pp. 1896–1901.
[58] P. Smets, Decision making in the tbm: the necessity of the pignistic transformation, Internat. J. Approx. Reason. 38 (2) (2005) 133–147.
[59] P. Smets, R. Kennes, The transferable belief model, Artif. Intell. 66 (2) (1994) 191–234.
[60] M.F. Smiley, A note on measure functions in lattice, Bull. Amer. Math. Soc. 46 (4) (1940) 239–241.
[61] R. Stanley, Enumerative Combinatorics, Cambridge Studies in Advanced Mathematics, vol. 1, Cambridge University Press, 1997.
[62] A. Urquhart, Distributive lattices with a dual homomorphic operation, Stud. Log. 38 (2) (1979) 201–209.
[63] A. Urquhart, The complexity of decision problems in relevance logic, in: J. Dunn, A. Gupta (Eds.), Truth or Consequences: Essays in Honor of Nuel Belnap, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990, pp. 61–76.
[64] D. van Dalen, Logic and Structure, fourth edition, University Text in Mathematics, Springer, 2004.
[65] H. Xu, Y. Hsia, P. Smets, Transferable belief model for decision making in the valuation-based systems, IEEE Trans. Syst. Man Cybern., Part A 26 (6) (1996) 698–707.
[66] J. Yen, Generalizing the Dempster–Shafer theory to fuzzy sets, IEEE Trans. Syst. Man Cybern. 20 (3) (May/June 1990) 559–570.
[67] M. Ying, Quantum computation, quantum theory and ai, Artif. Intell. 174 (2) (2010) 162–176.
[68] L. Zadeh, Fuzzy sets and information granularity, in: Advances in Fuzzy Set Theory and Applications, 1979, pp. 3–18.
[69] L. Zadeh, Fuzzy logic, IEEE Computer 21 (4) (1988) 83–93.
[70] M. Zhang, Z. Zhou, Ml-knn: A lazy learning approach to multi-label learning, Pattern Recogn. 40 (7) (2007) 2038–2048.