International Journal of Approximate Reasoning 160 (2023) 108976

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

On conditional belief functions in directed graphical models in the Dempster-Shafer theory

Radim Jiroušek [a,b], Václav Kratochvíl [a,b], Prakash P. Shenoy [c,*]

[a] Faculty of Management, University of Economics, Jindrichuv Hradec, Czechia
[b] Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czechia
[c] School of Business, University of Kansas, 1654 Naismith Dr., Capitol Federal Hall, Lawrence, KS, 66045, USA

ARTICLE INFO

Article history:
Received 12 April 2023
Received in revised form 10 June 2023
Accepted 29 June 2023
Available online 4 July 2023

Keywords:
Dempster-Shafer theory of belief functions
Conditional belief functions
Smets' conditional embedding
Belief-function directed graphical models

ABSTRACT

The primary goal is to define conditional belief functions in the Dempster-Shafer theory. We do so similarly to probability theory's notion of conditional probability tables. Conditional belief functions are necessary for constructing directed graphical belief function models in the same sense as conditional probability tables are necessary for constructing Bayesian networks. We provide examples of conditional belief functions, including those obtained by Smets' conditional embedding. Besides defining conditional belief functions, we state and prove a few basic properties of conditionals. In the belief-function literature, conditionals are defined starting from a joint belief function. Conditionals are then defined using the removal operator, an inverse of Dempster's combination operator. When such conditionals are well-defined belief functions, we show that our definition is equivalent to these definitions.

1. Introduction

The main goal of this article is to review the concept of conditional belief functions in the Dempster-Shafer (D-S) theory of belief functions [10,20], provide a formal definition, state some basic properties, and provide some examples.

Several theories of belief functions use the representation of belief functions but differ in the combination operators and corresponding semantics. The D-S theory uses Dempster's combination rule [10]. Fagin and Halpern [11] propose an alternative combination rule interpreting belief functions as credal sets [13]. These two theories of belief functions are different. A comparison of these two theories is outside the scope of this paper. Here, we are concerned exclusively with the D-S theory.

One of the earliest to define conditional belief functions for the D-S theory is Smets [29]. Other contributions on conditional belief functions are (in chronological order) [21,23,8,25,1,32,2,6,7].

Shafer [21] is concerned about parametric models. There is a discrete parameter variable Θ and a data variable X. We have a prior basic probability assignment (BPA) m_Θ for Θ. We have a conditional model for the data, BPA m_{X_θ} for X in the context $\theta \in \Omega_\Theta$. Based on a dataset of n independent observations of X, the task is to compute the posterior belief function for Θ. The BPAs m_{X_θ} for X in the context $\theta \in \Theta$ are converted to a conditional BPA $m_{X|\theta}$ for (Θ, X) using Smets' conditional embedding. The marginal of $m_{X|\theta}$ for Θ is vacuous. The conditionals BPA $m_{X|\theta}$ are combined using Dempster's rule resulting

* Corresponding author.
E-mail address: pshenoy@ku.edu (P.P. Shenoy).

https://doi.org/10.1016/j.ijar.2023.108976
0888-613X/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

in the conditional $m_{X|\Theta}$. This assumes that the BPAs $m_{X|\theta}$ are distinct, which may be reasonable if the number of elements of Ω_Θ is small. Shafer also looks at the case where BPAs $m_{X|\theta}$ are not independent, and some known distributions describe the dependency.

Shafer [23] discusses conditionals abstractly in the framework of a commutative semigroup $(\Phi, \otimes)$, where Φ is a set of potentials, and $\otimes : \Phi \times \Phi \to \Phi$ is a binary combination operator that is commutative and associative. He discusses conditionals as potentials extending another potential's domain. He calls such conditionals '*continuers*.' Thus, ψ is a continuer of σ from r to $r \cup s$ if and only if $\sigma^{\downarrow r} \oplus \psi = \sigma^{\downarrow r \cup s}$. Here, ψ and σ are potentials, and $\sigma^{\downarrow r}$ denotes the marginal of σ for r, and r and s are disjoint subsets of variables. The paper focuses on the computation of marginals of a commutative semigroup with a marginalization operator, and some interesting properties of continuers are stated.

Cano et al. [8] define conditionals abstractly in the framework of valuation-based systems. Still, they do require that the marginal $m(s|r)^{\downarrow r}$ of conditional $m(s|r)$ is a vacuous valuation for r. The focus is on finding marginals by propagating conditional valuations in a directed acyclic graph.

Shenoy [25] describes conditional valuations using the removal operator, which is an inverse of the combination operator. For the D-S theory, the removal operator corresponds to the pointwise division of commonality functions followed by normalization. If σ is a BPA for subset t of variables, and r and s are disjoint subsets of t, then conditional belief function $\sigma(s|r)$ is defined as $\sigma^{\downarrow r \cup s} \ominus \sigma^{\downarrow r}$. A consequence of this definition is that the marginal of $\sigma(s|r)$ for r is vacuous for r. One disadvantage of this definition is that conditionals are defined starting from the joint. This does not help construct joint belief functions. Another disadvantage is that $\sigma(s|r)$ may result in a BPA with negative masses. Such BPAs are called pseudo-BPAs.[1]

Almond [1] defines conditional belief functions as those obtained from a joint BPA by Dempster's conditioning and marginalization. Suppose $m_{X,Y}$ is a BPA for $\{X, Y\}$. He defines the corresponding conditional BPA $m_{Y|x}$, where $x \in \Omega_X$ is as follows. Suppose $m_{X=x}$ is a deterministic BPA for X such that $m_{X=x}(\{x\}) = 1$. Let $m_{Y_x} = (m_{X,Y} \oplus m_{X=x})^{\downarrow Y}$ denote the BPA for Y in the context $X = x$. Then, BPA $m_{Y|x}$ for $\{X, Y\}$ is obtained by Smets' conditional embedding of m_{Y_x}. He then discusses the problem of going from conditionals to joints. He argues that there is not a unique joint associated with a group of conditionals, e.g., $\{m_{Y|x}\}_{x \in \Omega_X}$. Smets' conditional embedding is discussed whereby a BPA m_{Y_x} for Y in the context $X = x$ is embedded into a BPA $m_{Y|x}$ for $\{X, Y\}$ (details of Smets' conditional embedding are discussed in Section 3). Next, BPA $m_{Y|X}$ for $\{X, Y\}$ is constructed from conditional embeddings $m_{Y|x}$ for $x \in \Omega_X$ as follows:

$$m_{Y|X} = \bigoplus_{x \in \Omega_X} m_{Y|x}. \tag{1}$$

Eq. (1) implicitly assumes that the conditionally embedded BPAs $m_{Y|x}$ are distinct. Almond claims this assumption is unrealistic except where we start from conditional BPAs $m_{Y|x}$ that are Bayesian. Almond [1] also defines what he calls "effectively conditional belief functions" with some caveat. The caveat is that $m_{Y|X} \oplus m_X$ represents "our joint belief about X and Y" ([1, p. 96]. Our definition of conditionals is similar to this definition without such a caveat.

Xu and Smets [32] discuss conditionals m_{Y_a} for Y when proposition a is observed, where $\emptyset \neq \mathsf{a} \subseteq \Omega_X$. Let $m_{Y|\mathsf{a}}$ denote the BPA for $\{X, Y\}$ after conditional embedding of m_{Y_a}. [1] and [32] discuss Dempster's combination of all such conditionals ($\emptyset$ denotes the empty set):

$$\bigoplus_{\emptyset \neq \mathsf{a} \subseteq \Omega_X} m_{Y|\mathsf{a}}. \tag{2}$$

While it may be reasonable to assume that BPAs $m_{Y|x}$ for $x \in \Omega_X$ are distinct as in Eq. (1), assuming that all BPAs $m_{Y|\mathsf{a}}$ for $\emptyset \neq \mathsf{a} \subseteq \Omega_X$ are distinct may be unreasonable. The focus of [32] is on computing marginals.

Ben-Yaghlane and Mellouli [2] encodes conditional knowledge as an if-then rule encoded as a conditional belief function. The focus of the paper is on making inferences from directed graphical models.

Boukhris et al. [6,7] analyze belief-function graphical models where the knowledge is Bayesian, i.e., they start with a Bayesian network. When a node has several parent nodes, instead of modeling the conditional for the node (given its parents) as a single conditional, they model each arrow from a parent node to the child node as a separate conditional. This leads to a belief-function graphical model that is different from the Bayesian network model. In [26], each conditional distribution for the child node given a state of its parents is modeled as a Bayesian BPA, which is then converted to a conditional BPA using Smets' conditional embedding. If all such conditional BPAs are combined using Dempster's combinational rule, we get a non-Bayesian BPA representing the conditional probability table for a variable. It is shown in [26, Theorems 1 and 2] that the joint BPA for all variables is a Bayesian BPA representing the same joint probability distribution as in a Bayesian network.

When constructing a belief-function directed graphical model, we do not start with a joint BPA. Instead, we construct a joint BPA using priors and conditionals. In this context, the current definitions in the literature could be more helpful. What exactly is a conditional BPA? What are their properties? Where do conditionals come from? How do our conditionals

[1] This phenomenon has been observed, e.g., in [16,25,18].

compare with the existing definitions? What are some examples of conditionals? Answering these questions is the primary goal of this article.

An outline of the remainder of the paper is as follows. Section 2 reviews the basics of D-S theory, including conditional independence relations. In Section 3, we define conditional belief functions and state some properties. Section 3.2 describes where some conditionals come from, including Smets' conditional embedding. In Section 3.3, we compare our definition with the existing definitions in the belief-function literature. In Section 3.4, we describe an example called *Organizing a Conference*, a belief-function directed graphical model with several examples of conditionals. In Section 4, we conclude with a summary.

2. Basics of D-S theory of belief functions

This section sketches the basics of the D-S theory of belief functions [10,20].

2.1. Representations

Knowledge is represented by basic probability assignments, belief functions, plausibility functions, commonality functions, credal sets, etc. Here we focus only on basic probability assignments and commonality functions.

Notation Let $\mathcal{V}$ denote a finite set of variables. Elements of $\mathcal{V}$ are denoted by upper-case Roman letters, X, Y, Z, etc. Subsets of $\mathcal{V}$ are denoted by lower-case Roman alphabets r, s, t, etc. Each variable X is associated with a finite state space Ω_X that contains all possible values of X. For subset $r \subseteq \mathcal{V}$, let $\Omega_r = \times_{X\in r}\Omega_X$ denote the state space of r. Let 2^{Ω_r} denote the set of all subsets of Ω_r.

Projection of states means dropping some coordinates. If $(x, y) \in \Omega_{X,Y}$, then $(x, y)^{\downarrow X} = x$. The projection of a subset of states is achieved by projecting every state in the subset. Suppose $\mathsf{a} \subseteq \Omega_{X,Y}$. Then, $\mathsf{a}^{\downarrow X} = \{x \in \Omega_X : (x, y) \in \mathsf{a}\}$. If $\mathsf{a} \subseteq \Omega_X$, then $\mathsf{a}^{\uparrow\{X,Y\}} = \mathsf{a} \times \Omega_Y \subseteq \Omega_{X,Y}$ is a *vacuous extension* of a to $\{X, Y\}$.

Basic probability assignment A *basic probability assignment* (BPA) m for r is a function $m : 2^{\Omega_r} \to [0, 1]$ such that

$$m(\emptyset) = 0, \text{ and} \tag{3}$$

$$\sum_{\mathsf{a}\subseteq\Omega_r} m(\mathsf{a}) = 1. \tag{4}$$

m represents some knowledge about variables in r, and we say the *domain* of m is r. $m(\mathsf{a})$ is the probability assigned exactly to the subset a of Ω_r. Subsets a such that $m(\mathsf{a}) > 0$ are called *focal elements* of m. If m has only one focal element (with probability 1), we say m is *deterministic*. If the focal element of a deterministic BPA is Ω_r, we say m is *vacuous*. If all the focal elements of m are singleton subsets of Ω_r, we say m is *Bayesian*. A Bayesian BPA is, in essence, a probability mass function (PMF) of r.

Commonality functions The knowledge encoded in a BPA m for r can be represented as a corresponding commonality function. The *commonality function* (CF) Q_m corresponding to BPA m is such that for all $\mathsf{a} \subseteq \Omega_r$,

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b}\supseteq\mathsf{a}} m(\mathsf{b}). \tag{5}$$

$Q_m(\mathsf{a})$ represents the probability mass that could move to every state in the subset a. From Eqs. (4) and (5), it follows that $0 \le Q_m \le 1$. From Eqs. (3) and (5), it follows that $Q_m(\emptyset) = 1$. If m is a vacuous BPA for r, then $Q_m(\mathsf{a}) = 1$ for all $\mathsf{a} \subseteq \Omega_r$. CFs are non-increasing in the sense that if $\mathsf{a} \subseteq \mathsf{b}$, then $Q_m(\mathsf{a}) \ge Q_m(\mathsf{b})$.

The CF Q_m has the same information as the corresponding BPA m. Given a CF Q_m, we can recover the corresponding BPA m as follows [20]:

$$m(\mathsf{a}) = \sum_{\mathsf{b}\subseteq\Omega_r:\,\mathsf{b}\supseteq\mathsf{a}} (-1)^{|\mathsf{b}\setminus\mathsf{a}|} Q_m(\mathsf{b}). \tag{6}$$

Thus,[2] it follows that $Q : 2^{\Omega_r} \to [0, 1]$ is a *well-defined CF* for r iff

$$Q(\emptyset) = 1, \tag{7}$$

$$\sum_{\mathsf{b}\subseteq\Omega_r:\,\mathsf{b}\supseteq\mathsf{a}} (-1)^{|\mathsf{b}\setminus\mathsf{a}|} Q(\mathsf{b}) \ge 0, \quad \text{for all } \mathsf{a} \subseteq \Omega_r, \text{ and} \tag{8}$$

$$\sum_{\emptyset\neq\mathsf{a}\subseteq\Omega_r} (-1)^{|\mathsf{a}|+1} Q(\mathsf{a}) = 1. \tag{9}$$

[2] Q_m denotes the CF derived from BPA m. The question we are addressing is: when is Q a CF? (without reference to a BPA m).

The left-hand side (LHS) of Eq. (8) is $m_Q(\mathsf{a})$, the BPA corresponding to CF Q, and the LHS of Eq. (9) is $\sum_{\emptyset\neq\mathsf{a}\subseteq\Omega_r} m_Q(\mathsf{a})$. Eq. (9) can be regarded as a normalization condition for a CF. Thus, if we have a function $Q: 2^{\Omega_r} \to [0,1]$ that satisfies Eqs. (7) and (8), but not (9), then we can divide each of the values of the function for non-empty subsets in 2^{Ω_r} by $K = \sum_{\emptyset\neq\mathsf{a}\subseteq\Omega_r} (-1)^{|\mathsf{a}|+1} Q_m(\mathsf{a})$, and the resulting function will then qualify as a CF. If we have a function $Q: 2^{\Omega_r} \to \mathbb{R}$ ($\mathbb{R}$ is the set of real numbers) that satisfies Eqs. (7) and (9), but not Eq. (8), then such a function is called a *pseudo-CF*. The BPA corresponding to Q (using Eq. (6)) will have some negative masses, and such a BPA is called a *pseudo-BPA*.[3]

2.2. Inference operators

There are two basic inference operators in the D-S theory, marginalization, and combination. There is also a removal operator for computing conditionals from a joint belief function, which is an inverse of the combination operator [25]. The removal operator will be defined in Section 3.

Marginalization Suppose m is a BPA for r and suppose $s \subseteq r$. The marginalization operator transforms a BPA m for r to a BPA $m^{\downarrow s}$ for s by eliminating variables in $r \setminus s$.

$$m^{\downarrow s}(\mathsf{a}) = \sum_{\mathsf{b}\subseteq\Omega_r : \mathsf{b}^{\downarrow s}=\mathsf{a}} m(\mathsf{b}). \tag{10}$$

The marginalization operator satisfies the following property. Suppose m is a BPA for r and suppose X_1 and X_2 are two distinct variables in r. Then

$$(m^{\downarrow r\setminus\{X_1\}})^{\downarrow r\setminus\{X_1,X_2\}} = (m^{\downarrow r\setminus\{X_2\}})^{\downarrow r\setminus\{X_1,X_2\}}. \tag{11}$$

Thus, the order in which variables are eliminated does not matter.

Dempster's combination rule Suppose m_1 is a BPA for r_1 and m_2 is a BPA for r_2. Also, we assume that m_1 and m_2 are based on *distinct* pieces of knowledge.[4] We combine these two BPAs using Dempster's product-intersection rule [10] as follows. Let $m_1 \oplus m_2$ denote the BPA after the combination. Then $m_1 \oplus m_2$ is a BPA for $r_1 \cup r_2$ such that for all $\mathsf{a} \subseteq \Omega_{r_1\cup r_2}$

$$(m_1 \oplus m_2)(\mathsf{a}) = K^{-1} \sum_{\mathsf{a}_1\subseteq\Omega_{r_1}} \sum_{\mathsf{a}_2\subseteq\Omega_{r_2}} \{m_1(\mathsf{a}_1)\, m_2(\mathsf{a}_2) \mid \mathsf{a}_1^{\uparrow r_1\cup r_2} \cap \mathsf{a}_2^{\uparrow r_1\cup r_2} = \mathsf{a}\}, \tag{12}$$

where K is a normalization constant given by

$$K = 1 - \sum_{\mathsf{a}_1\subseteq\Omega_{r_1}} \sum_{\mathsf{a}_2\subseteq\Omega_{r_2}} \{m_1(\mathsf{a}_1)\, m_2(\mathsf{a}_2) \mid \mathsf{a}_1^{\uparrow r_1\cup r_2} \cap \mathsf{a}_2^{\uparrow r_1\cup r_2} = \emptyset\}. \tag{13}$$

We assume $K > 0$. If $K = 0$, then m_1 and m_2 are said to be in total conflict and cannot be combined. If $K = 1$, we say m_1 and m_2 are *non-conflicting*.

Dempster's combination rule can also be described using commonality functions. Consider two distinct BPAs m_1 for r_1 and m_2 for r_2, and let Q_1 and Q_2 denote the corresponding commonality functions. Then, as showed in [20], for all $\emptyset \neq \mathsf{a} \subseteq \Omega_{r_1\cup r_2}$,

$$(Q_1 \oplus Q_2)(\mathsf{a}) = K^{-1}\, Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}), \tag{14}$$

where K is a normalization constant defined as follows:

$$K = \sum_{\emptyset\neq\mathsf{a}\subseteq\Omega_{r_1\cup r_2}} (-1)^{|\mathsf{a}|+1}\, Q_1(\mathsf{a}^{\downarrow r_1})\, Q_2(\mathsf{a}^{\downarrow r_2}). \tag{15}$$

Thus, in terms of CFs, Dempster's rule is pointwise multiplication of CFs followed by normalization. As shown in [20], the normalization constant in Eq. (15) is the same as in Eq. (13).

It is easy to show that Dempster's combination is commutative and associative: $m_1 \oplus m_2 = m_2 \oplus m_1$, and $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$. Also, marginalization and Dempster's combination rule satisfies a vital property, called the local computation property [28] as follows.

[3] Pseudo-BPAs were first observed by [16] and also been noted in [18].

[4] See [27] for a discussion on distinct belief functions. A summary is included in Section 2.4.

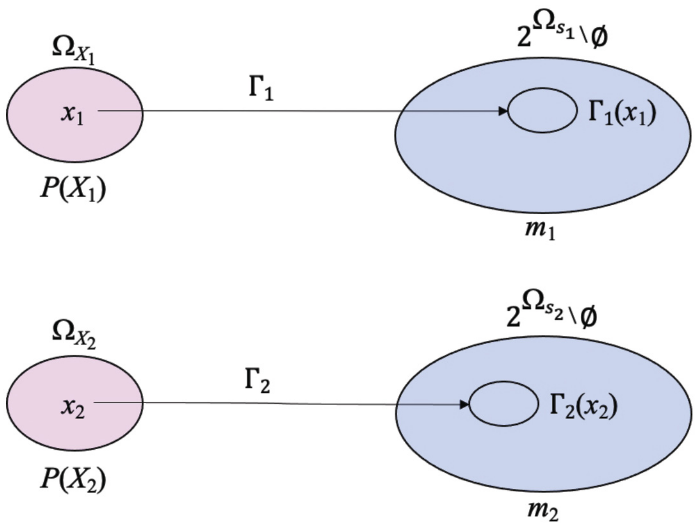

Fig. 1. Dempster's multi-valued semantics for BPAs.

Local computation property Suppose m_1 is a BPA for r_1 and m_2 is a BPA for r_2. Suppose $X \in r_1$ and $X \notin r_2$. Then,

$$(m_1 \oplus m_2)^{\downarrow(r_1 \cup r_2)\setminus\{X\}} = (m_1)^{\downarrow r_1 \setminus\{X\}} \oplus m_2 \tag{16}$$

This property is the basis of computing marginals of joint belief functions. [12] describes an implementation of a local computation algorithm in Matlab for computing marginals of joint belief function models.

2.3. Conditional independence

Shenoy [25] describes conditional independence relation in the framework of valuation-based systems using factorization semantics. Here, we describe it for the D-S theory of belief functions.

Definition 1 *(Conditional Independence).* Suppose $\mathcal{V}$ denotes the set of variables, and suppose r, s, and t are disjoint subsets of $\mathcal{V}$. Suppose m is a joint BPA for $\mathcal{V}$. We say r and s are conditionally independent (CI) given t with respect to BPA m if and only if $m^{\downarrow r\cup s\cup t} = m_{r\cup t} \oplus m_{s\cup t}$, where $m_{r\cup t}$ and $m_{s\cup t}$ are distinct BPAs for $r \cup t$ and $s \cup t$, respectively.

This definition generalizes the CI relation in probability theory [9]. There are other definitions of CI in the D-S literature, e.g., [17,31,3,4]. Definition 1 is closest to the definition in [17]. The definitions in [31,3,4] are based on the notion of non-interactivity, which are not useful in describing CI in belief-function graphical models.

The definition of CI in Definition 1 satisfies the graphoid properties of probabilistic CI [19].

2.4. Distinct belief functions

This material in this subsection is taken from [27]. Distinct belief functions are also called independent belief functions in the D-S literature.[5] Dempster's combination rule is only applicable to combining distinct BPAs. So, what are distinct BPAs? Dempster [10] provides a definition. Consider the multi-valued semantics of BPAs as shown in Fig. 1.

Suppose we have a probability mass function (PMF) $P(X_1)$ for X_1, a multivalued function $\Gamma_1 : \Omega_{X_1} \to 2^{s_1} \setminus \emptyset$ that defines the BPA m_1 for s_1. Similarly, suppose we have a probability mass function (PMF) $P(X_2)$ for X_2, a multivalued function $\Gamma_2 : \Omega_{X_2} \to 2^{s_2} \setminus \emptyset$ that defines the BPA m_2 for s_2. BPAs m_1 and m_2 are distinct if and only if X_1 and X_2 are independent random variables, i.e., $P(X_1, X_2) = P(X_1) \otimes P(X_2)$, where $\otimes$ is the probabilistic combination operator, point-wise multiplication followed by normalization.

Some comments about Dempster's definition.

1. In practice, not every belief function in a belief function model is associated with a multi-valued mapping. Thus, Dempster's definition cannot be used directly in practice.

[5] The terminology of 'distinct' belief functions is due to Smets [30]. As independence is usually associated with random variables, we prefer the terminology of distinct belief functions.

2. We say BPA m is idempotent if $m \oplus m = m$. Idempotent knowledge is knowledge encoded in a BPA m that is idempotent. For example, if m is deterministic, then m is idempotent. Thus, double-counting idempotent knowledge is not a problem; double-counting non-idempotent knowledge is.
3. If we assume independence of random variables X_1 and X_2 when they are not, and we combine m_1 and m_2, then we are double-counting common knowledge encoded in m_1 and m_2. If the common knowledge encoded in these two BPAs is non-idempotent, then we have a problem. Thus, the spirit of Dempster's definition is that two belief functions are distinct if, when combining them using Dempster's combination rule, we are not double-counting non-idempotent knowledge.
4. BPA m_X for X and conditional BPA $m_{Y|X}$ for Y given X are always distinct (regardless of the numeric values of these BPAs). Notice that $(m_X \oplus m_{Y|X})^{\downarrow X} = m_X$, and $(m_X \oplus m_{Y|X}) \ominus m_X = m_{Y|X}$.
5. BPAs m_X for X and m_Y for Y are distinct if and only if $X \perp\!\!\!\perp_{m_X \oplus m_Y} Y$.
6. BPA $m_{X,Y}$ for $\{X, Y\}$ and conditional BPA $m_{Z|Y}$ for Z given Y are distinct if and only if $X \perp\!\!\!\perp_{m_{X,Y} \oplus m_{Z|Y}} Z \mid Y$.
7. The discussion of distinct belief functions is valid more broadly to many uncertainty calculi, including probability theory.
8. Some references to the literature on distinct belief functions are as follows: [22],[6] and [30].

2.5. Belief-function directed graphical models

We start with some notation. A directed graph G is a pair $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{X_1, \ldots, X_n\}$ denotes the set of *nodes*, and $\mathcal{E}$ denotes the set of *directed edges* (X_i, X_j) between two distinct variables in $\mathcal{V}$. For any node X_i, let $Pa_G(X_i)$ denote $\{X_j \in \mathcal{V} : (X_j, X_i) \in \mathcal{E}\}$. A directed graph is said to be acyclic if and only if there exists a sequence of the nodes of the graph, say $(X_1, \ldots, X_n)$ such that if there is a directed edge $(X_i, X_j) \in \mathcal{E}$ then X_i precedes X_j in the sequence. Such a sequence is called a *topological* sequence.

Definition 2 *(BF directed graphical model).* Suppose $G = (\mathcal{V}, \mathcal{E})$ is a directed acyclic graph with n nodes in $\mathcal{V}$. A belief-function directed graphical model (BFDGM) is a pair $(G, \{m_1, \ldots, m_n\})$ such that BPA m_i is a conditional BPA for X_i given $Pa_G(X_i)$, for $i = 1, \ldots, n$. A fundamental assumption of a BFDGM is that $m_1, \ldots, m_n$ are all distinct, and the joint BPA m for $\mathcal{V}$ associated with the model is given by

$$m = \bigoplus_{i=1}^{n} m_i. \tag{17}$$

The assumption in Definition 2 that all conditionals are distinct allows the combination in Eq. (17). Given m, the joint BPA for $\mathcal{V}$, the definition of conditional independence in Definition 1 implies the following CI relations among the variables: For each variable X_i in a topological sequence $(X_1, \ldots, X_n)$, given $Pa_G(X_i)$, X_i is conditionally independent of the preceding variables in the sequence (excluding $Pa_G(X_i)$) with respect to the joint BPA m for $\mathcal{V}$. Thus, $m_1, \ldots m_n$ are distinct if and only if the CI assumptions of the model are valid [27].

3. Conditional belief functions

This section defines a conditional belief function similar to a conditional probability table in probability theory without starting from a joint distribution. In belief-function directed graphical models, we construct a joint using such conditional belief functions. We begin with the probabilistic case.

Suppose P_X denotes a PMF of X, and we wish to construct a joint PMF $P_{X,Y}$ of $\{X, Y\}$ such that P_X is the marginal of $P_{X,Y}$ for X (as is typically done in a probabilistic graphical model). One way to do this is to define a PMF of Y for each $x \in \Omega_X$ such that[7] $P_X(x) > 0$. Let $P_{Y|x} : \Omega_Y \to [0, 1]$ denote a PMF of Y when X is known to be x, i.e., for all $y \in \Omega_Y$, $P_{Y|x}(y) \geq 0$ and $\sum_{y \in \Omega_Y} P_{Y|x}(y) = 1$. We can embed all PMFs $P_{Y|x}$ of Y for each $x \in \Omega_X$ into a function $P_{Y|X} : \Omega_{X,Y} \to [0, 1]$ such that $P_{Y|X}(x, y) = P_{Y|x}(y)$. In the Bayesian network literature, the function $P_{Y|X}$ is called a *conditional probability table* (CPT). The joint PMF $P_{X,Y}$ of $\{X, Y\}$ can now be defined as $P_{X,Y}(x, y) = P_X(x) \cdot P_{Y|X}(x, y)$. Some observations:

1. Notice that if we marginalize the CPT $P_{Y|X}$ to X, we get a potential identically 1 for all values of $x \in \Omega_X$, which is a vacuous potential in probability theory. It is not a PMF but can be normalized to get an equally-likely PMF.
2. If we consider probabilistic combination operator $\otimes$ as pointwise multiplication followed by normalization, then we can write $P_{X,Y} = P_X \otimes P_{Y|X}$. The normalization constant is 1 for this combination.
3. It follows from the first observation that the marginal of $P_{X,Y}$ for X is P_X. So, the CPT $P_{Y|X}$ is used to extend P_X to $P_{X,Y}$ such that the marginal $(P_{X,Y})^{\downarrow X} = P_X$.

[6] This was published unchanged as [24].
[7] If $P_X(x) = 0$, then the conditional has no effect on the joint, and can be left undefined.

Consider the belief-function directed graphical model: $X \to Y \to Z$. In this model, we have BPA m_X for X, $m_{Y|X}$ for $\{X, Y\}$ that is a conditional for Y given X, and a BPA $m_{Z|Y}$ for $\{Y, Z\}$ that is a conditional for Z given Y. In this model, in general, $m_{Y|Z}$ is not distinct from $m_{X,Y} = m_X \oplus m_{Y|X}$. Assuming that Z and X are conditionally independent given Y, then $m_{Z|Y}$ is distinct from $m_{X,Y}$. We can combine the three conditionals using Dempster's rule to obtain the joint: $m_{X,Y,Z} = m_X \oplus m_{Y|X} \oplus m_{Z|Y}$. This is the motivation behind the definition of conditionals, which follows.

Definition 3 *(Conditional BPA).* Suppose r and s are disjoint subsets of variables and suppose $r' \subset r$. Suppose $m_{s|r'}$ is a BPA for $r' \cup s$. We say $m_{s|r'}$ is a conditional BPA for s given r' if and only if

1. $(m_{s|r'})^{\downarrow r'}$ is a vacuous BPA for r', and
2. If BPAs m_r for r and $m_{s|r'}$ for $r' \cup s$ are distinct, then $m_r \oplus m_{s|r'}$ is a BPA for $r \cup s$.

For conditional $m_{s|r'}$, s is called the *head* of the conditional, and r', its *tail*. Some comments:

1. The first condition in Definition 3 says that $m_{s|r'}$ tells us nothing about the tail r'.
2. The notion of distinct belief functions is discussed in [27]. The requirement that m_r and $m_{s|r'}$ are distinct is equivalent to: s is conditionally independent of $r \setminus r'$ given r' with respect to the joint BPA $m_r \oplus m_{s|r'}$ for $r \cup s$. This condition is essential. Without this condition, we cannot claim that $m_r \oplus m_{s|r'}$ is a BPA for $r \cup s$, as the two BPAs would not be distinct, and Dempster's combination would not be justified.
3. If $r = r' = \{X\}$ and $s = \{Y\}$, then we have a belief function analog of $P(Y|X)$ discussed in the second paragraph in this section. $P(X)$ and $P(Y|X)$ are always distinct. Similarly, BPA m_X and conditional BPA $m_{Y|X}$ for Y given X are always distinct. This is because no conditional independence assumptions exist in a *complete* directed graphical model. A complete directed graphical model is one where we have a unique topological sequence. In a complete directed graphical model, the preceding variables are its parents for each variable in the topological sequence (except the first). The model $X \to Y$ is an example of a complete model with the topological sequence (X, Y).
4. If $r' = r$, then m_r and $m_{s|r}$ are always distinct, and for this special case, condition 2 in Definition 3 is trivially satisfied. This is similar to the case in Comment 3 above. In graphical models, it is rarely the case that the tail of all conditionals includes all variables that precede it in some topological sequence (except for toy problems involving a small number of variables).
5. Notice that the marginal $(m_r \oplus m_{s|r'})^{\downarrow r} = m_r$. This follows from the local computation property of Dempster's rule and condition 1 of Definition 3. Thus, a conditional $m_{s|r'}$ allows us to extend a BPA m_r for r to a BPA $m_r \oplus m_{s|r'}$ for $r \cup s$ (assuming m_r and $m_{s|r'}$ are distinct).
6. Notice that m_r and $m_{s|r'}$ are non-conflicting, i.e., the normalization constant K in $m_r \oplus m_{s|r'}$ is 1 (Eq. (13)).
7. If s is a singleton subset, say $\{Y\}$, and $r' = Pa_G(Y)$, where $Pa_G(Y)$ denotes the parents of Y in some directed acyclic graph, then the conditional $m_{Y|pa(Y)}$ is a belief-function analog of a CPT for Y in Bayesian networks.

In a belief-function directed graphical model, we have a conditional associated with each variable X. The head of the associated condition is X, and the tail consists of the parents of X. For variables with no parents, we have priors associated with such variables. For convenience, priors can be considered conditionals with empty tails. For such BPAs, both conditions in the definition are trivially true–the sum of the probability masses in a BPA is 1 (Eq. (4)), which can be regarded as a BPA for the $\emptyset$, and every BPA m is distinct from the BPA for the $\emptyset$.

3.1. Properties of conditionals

The following theorem was stated in [25] where conditionals were defined using an inverse of the combination operator called removal. Here we prove the same results using the definition of conditionals above that include only combination and marginalization operators.

Theorem 1 *(Properties of conditionals [25]). Suppose r, s, and t are disjoint subsets of variables. Let m_r denote a BPA for r, $m_{s|r}$ denote a conditional BPA with head s and tail r, etc. Then,*

1. $m_r \oplus m_{s|r} \oplus m_{t|r\cup s} = m_{r\cup s\cup t}$.
2. $m_{s|r} \oplus m_{t|r\cup s} = m_{s\cup t|r}$.
3. *Suppose $s' \subseteq s$. Then, $(m_{s|r})^{\downarrow r\cup s'} = m_{s'|r}$.*
4. $(m_{s|r} \oplus m_{t|r\cup s})^{\downarrow r\cup t} = m_{t|r}$.
5. *Suppose $r' \subseteq r$, $s' \subseteq s$, and $m_{r\cup s}$ is distinct from $m_{t|r'\cup s'}$. Then, $m_{r\cup s} \oplus m_{t|r'\cup s'} = m_{r\cup s\cup t}$.*

Proof of Theorem 1.

1. $m_r \oplus m_{s|r} \oplus m_{t|r\cup s} = (m_r \oplus m_{s|r}) \oplus m_{t|r\cup s} = m_{r\cup s} \oplus m_{t|r\cup s} = m_{r\cup s\cup t}$. Condition 2 is always true, as we have a complete model.

2. Let ι_r denote the vacuous BPA for r. To show condition 1, notice that

$$\begin{aligned}(m_{s|r} \oplus m_{t|r\cup s})^{\downarrow r} &= ((m_{s|r} \oplus m_{t|r\cup s})^{\downarrow r\cup s})^{\downarrow r}\\ &= (m_{s|r} \oplus (m_{t|r\cup s})^{\downarrow r\cup s})^{\downarrow r}\\ &= (m_{s|r} \oplus \iota_{r\cup s})^{\downarrow r}\\ &= (m_{s|r})^{\downarrow r} = \iota_r.\end{aligned}$$

   To show condition 2, suppose m_r is a BPA for r. Then, it follows from Statement 1 that $m_r \oplus (m_{s|r} \oplus m_{t|r\cup s}) = m_{r\cup s\cup t}$.
3. To show condition 1, notice that $((m_{s|r})^{\downarrow r\cup s'})^{\downarrow r} = (m_{s|r})^{\downarrow r} = \iota_r$. To show condition 2, suppose m_r is a BPA for r. Then, $m_r \oplus (m_{s|r})^{\downarrow r\cup s'} = m_{r\cup s'}$.
4. To show condition 1, notice that

$$\begin{aligned}((m_{s|r} \oplus m_{t|r\cup s})^{\downarrow r\cup t})^{\downarrow r} &= ((m_{s|r} \oplus m_{t|r\cup s})^{\downarrow r\cup s})^{\downarrow r}\\ &= (m_{s|r} \oplus (m_{t|r\cup s})^{\downarrow r\cup s})^{\downarrow r}\\ &= ((m_{s|r} \oplus \iota_{r\cup s})^{\downarrow r}\\ &= (m_{s|r})^{\downarrow r} = \iota_r.\end{aligned}$$

   To show condition 2, suppose m_r is a BPA for r. Then,

$$\begin{aligned}m_r \oplus (m_{s|r} \oplus m_{t|r\cup s})^{\downarrow r\cup t} &= (m_r \oplus m_{s|r} \oplus m_{t|r\cup s})^{\downarrow r\cup t}\\ &= (m_{r\cup s\cup t})^{\downarrow r\cup t}\\ &= m_{r\cup t}.\end{aligned}$$

5. It follows from Definition 3 that $m_{r\cup s} \oplus m_{t|r'\cup s'} = m_{r\cup s\cup t}$. $\square$

3.2. Where do conditionals come from?

A conditional BPA $m_{r|s}$ describes the relationship between the variables in r and s. One source of conditionals is Smets' conditional embedding [29]. To describe conditional embedding, consider the case of two variables, X and Y. To describe the dependency between Y and X, suppose that in the context $X = x$, our belief in Y is described by a BPA m_{Y_x} for Y. The BPA m_{Y_x} for Y needs to be embedded into a BPA $m_{Y|x}$ for $\{X, Y\}$ such that

1. $m_{Y|x}$ is a conditional BPA for Y given $X = x$, i.e., $(m_{Y|x})^{\downarrow X}$ is vacuous BPA for X, and
2. suppose $m_{X=x}$ is a deterministic BPA for X such that $m_{X=x}(x) = 1$. Then $(m_{Y|x} \oplus m_{X=x})^{\downarrow Y} = m_{Y_x}$.

One way to do this is to take each focal element $\mathsf{b} \subseteq \Omega_Y$ of m_{Y_x} and convert it to the corresponding focal element

$$(\{x\} \times \mathsf{b}) \cup ((\Omega_X \setminus \{x\}) \times \Omega_Y) \subseteq \Omega_{X,Y} \tag{18}$$

of BPA $m_{Y|x}$ for $\{X, Y\}$ with the same mass. It is easy to confirm that this embedding method satisfies both conditions mentioned above. Suppose we have several distinct conditionals, e.g., $m_{Y|x_1}$, $m_{Y|x_2}$, etc., where x_1, and x_2 are distinct values of X. We combine the conditional embeddings by Dempster's combination rule to obtain $m_{Y|X}$. An example of conditional embedding follows.

Example 1 *(Conditional embedding).* Consider binary variables X and Y, with $\Omega_X = \{x, \bar{x}\}$ and $\Omega_Y = \{y, \bar{y}\}$. Suppose we have a BPA m_{Y_x} for Y in the context $X = x$ as follows:

$$m_{Y_x}(y) = 0.8,\ m_{Y_x}(\Omega_Y) = 0.2,$$

then its conditional embedding into the conditional BPA $m_{Y|x}$ for $\{X, Y\}$ is as follows:

$$m_{Y|x}(\{(x, y), (\bar{x}, y), (\bar{x}, \bar{y})\}) = 0.8,\ m_{Y|x}(\Omega_{X,Y}) = 0.2.$$

Similarly, if we have a BPA $m_{Y_{\bar{x}}}$ for Y in the context $X = \bar{x}$ as follows:

$$m_{Y_{\bar{x}}}(\bar{y}) = 0.3,\ m_{Y_{\bar{x}}}(\Omega_Y) = 0.7,$$

then its conditional embedding into the conditional BPA $m_{Y|\bar{x}}$ for $\{X, Y\}$ is as follows:

$$m_{Y|\bar{x}}(\{(x, y), (x, \bar{y}), (\bar{x}, \bar{y})\}) = 0.3,\ m_{Y|\bar{x}}(\Omega_{X,Y}) = 0.7.$$

Assuming we have these two BPAs and their corresponding embeddings, it is clear that the two BPA $m_{Y|x}$ and $m_{Y|\bar{x}}$ are distinct and can be combined with Dempster's rule of combination, resulting in the conditional BPA $m_{Y|X} = m_{Y|x} \oplus m_{Y|\bar{x}}$ for $\{X, Y\}$ as follows:

$$\begin{aligned}
m_{Y|X}(\{(x, y), (\bar{x}, \bar{y})\}) &= 0.24, \\
m_{Y|X}(\{(x, y), (\bar{x}, y), (\bar{x}, \bar{y})\}) &= 0.56, \\
m_{Y|X}(\{(x, y), (x, \bar{y}), (\bar{x}, \bar{y})\}) &= 0.06, \text{ and} \\
m_{Y|X}(\Omega_{X,Y}) &= 0.14.
\end{aligned}$$

$m_{Y|X}$ has the following properties.

1. $(m_{Y|X})^{\downarrow X}$ is the vacuous BPA for X.
2. First notice that any BPA for X is distinct from conditional BPA $m_{Y|X}$. If we combine $m_{Y|X}$ with deterministic BPA $m_{X=x}(\{x\}) = 1$ for X, and marginalize the combination to Y, then we get m_{Y_x}, i.e., $(m_{Y|X} \oplus m_{X=x})^{\downarrow Y} = m_{Y_x}$.
3. Similarly, $(m_{Y|X} \oplus m_{X=\bar{x}})^{\downarrow Y} = m_{Y_{\bar{x}}}$.
4. $m_{Y|X}$ is the belief function analog of CPT $P_{Y|X}$ in probability theory. $\square$

Smets' conditional embedding is one way to obtain conditionals. [31] argues that his conditional embedding method satisfies the principle of minimal commitment. Shenoy [26, Theorem 2, p. 15] shows that a Bayesian network (BN) can be modeled exactly by a corresponding belief function model if the conditional probability tables in a Bayesian network are modeled using Smets' conditional embedding. By exactly, we mean the joint BPA in a corresponding belief-function graphical model is a Bayesian BPA corresponding to a BN's joint probability mass function. Black and Laskey [5] propose other methods to get conditionals, but not much is known about these methods.

3.3. Comparison of our definition with existing definitions

As is discussed in the introduction, conditionals are defined in the belief function literature starting from a joint belief function. So, how does our definition in Definition 3 compare with the literature? First, we start with the definition of the removal operator, which is an inverse of Dempster's combination operator [25].

Removal operator Suppose r and s are disjoint sets of variables, and suppose we have a joint BPA m for $r \cup s$ obtained by combining a BPA m_r for r and a conditional BPA $m_{s|r}$ for s given r, i.e., $m = m_r \oplus m_{s|r}$. Notice that $m^{\downarrow r} = m_r$. Can we reconstruct the conditional from the joint? The answer is yes, using the removal operator defined as follows.

Definition 4 *(Removal operator).* Suppose r and s are disjoint sets of variables and suppose Q is a CF for $r \cup s$. Then the removal of $Q^{\downarrow r}$ from Q, denoted by $Q \ominus Q^{\downarrow r}$, is a CF for $r \cup s$ defined as follows:

$$(Q \ominus Q^{\downarrow r})(\mathsf{a}) = \begin{cases} K_q^{-1}\, Q(\mathsf{a})/Q^{\downarrow r}(\mathsf{a}^{\downarrow r}) & \text{if } Q^{\downarrow r}(\mathsf{a}^{\downarrow r}) > 0, \\ \text{undefined} & \text{if } Q^{\downarrow r}(\mathsf{a}^{\downarrow r}) = 0, \end{cases} \tag{19}$$

for all $\mathsf{a} \subseteq \Omega_{r\cup s}$, where K_q is a normalization constant given by

$$K_q = \sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_{r\cup s}} (-1)^{|\mathsf{a}|+1}\, Q(\mathsf{a})/Q^{\downarrow r}(\mathsf{a}^{\downarrow s}). \tag{20}$$

Some comments.

1. In probability theory, the conditional PMF $P(Y|x) = P(X, Y)(x, y)/P(X)(x)$ is well-defined only for $x \in \Omega_X$ such that $P(X)(x) > 0$. So, when we are defining the joint in terms of $P(X)$ and $P(Y|x)$, it doesn't matter how we define $P(Y|x)$ when $P(X)(x) = 0$ because $P(X, Y)(x, y) = P(X)(x) \cdot P(Y|x)(y)$, and $P(X, Y)(x, y) = 0$ when $P(X)(x) = 0$. But, when we compute the conditional from the joint, we encounter 0/0 when $P(Y|x)(y) = P(X, Y)(x, y)/P(X)(x)$ and $P(X)(x) = 0$. We have a similar situation in the D-S theory when we define removal in terms of CFs. If we encounter 0/0 in the definition of removal, what it really means is that the $(Q \ominus Q^{\downarrow r})(\mathsf{a})$ is undefined, which is what the definition of removal says in Eq. (19).
2. The removal operation $Q \ominus Q^{\downarrow r}$ corresponds to removing knowledge encoded in $Q^{\downarrow r}$ from the knowledge encoded in Q.
3. If we start with an arbitrary joint CF $Q_{X,Y}$ for $\{X, Y\}$, and remove its marginal CF $(Q_{X,Y})^{\downarrow X}$ for X, the removal operation may result in a pseudo-CF. This is demonstrated in Example 2, which follows these comments. [18] argue that pseudo-CF are useful in making inferences from a belief-function model. This is because $(Q \ominus Q^{\downarrow r}) \oplus Q^{\downarrow r} = Q$, which is a well-defined CF.

Table 1
The computation of $m_{X,Y} \ominus m_X$ in Example 2. Empty cell values are assumed to be 0. The last row, labeled K, denotes the sum in Eq. (4) for BPA values and the sum in Eq. (9) for CF values.

$2^{\Omega_{X,Y}}$	$m_{X,Y}$	$m_X^{\uparrow\{X,Y\}}$	$Q_{m_{X,Y}}$	$Q_{m_X^{\uparrow\{X,Y\}}}$	$Q_{m_{X,Y}} \ominus Q_{m_X^{\uparrow\{X,Y\}}}$	$m_{X,Y} \ominus m_X$
$\emptyset$			1	1	1	
$\{(x,y)\}$	0.9		1	1	1	0.9
$\{(x,\bar{y})\}$			0.1	1	0.1	
$\{(\bar{x},y)\}$				0.1		
$\{(\bar{x},\bar{y})\}$			0.1	0.1	1	
$\{(x,y),(x,\bar{y})\}$		0.9	0.1	1	0.1	−0.9
$\{(x,y),(\bar{x},y)\}$				0.1		
$\{(x,y),(\bar{x},\bar{y})\}$			0.1	0.1	1	
$\{(x,\bar{y}),(\bar{x},y)\}$				0.1		
$\{(x,\bar{y}),(\bar{x},\bar{y})\}$			0.1	0.1	1	
$\{(\bar{x},y),(\bar{x},\bar{y})\}$				0.1		
$\{(x,y),(x,\bar{y}),(\bar{x},y)\}$				0.1		
$\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}$	0.1		0.1	0.1	1	1
$\{(x,y),(\bar{x},y),(\bar{x},\bar{y})\}$				0.1		
$\{(x,\bar{y}),(\bar{x},y),(\bar{x},\bar{y})\}$				0.1		
$\Omega_{X,Y}$		0.1		0.1		
K	1	1	1	1	1	1

4. If we start with a CF $Q = Q_r \oplus Q_{s|r}$ for $r \cup s$, then $Q^{\downarrow r} = Q_r$. In this case, $Q \ominus Q^{\downarrow r} = Q_{s|r}$ is a well-defined CF. Also, the normalization constant K_q in Eq. (20) is 1. Notice that in this case, $Q^{\downarrow r} = Q_r$ is explicitly included in Q. Theorem 2 (which follows Example 2) shows that $Q \ominus Q^{\downarrow r}$ is a well-defined CF if and only if $Q^{\downarrow r}$ is explicitly included in Q.
5. To compute removal more efficiently, we could use inverses defined in [25, pp. 212–213]. In terms of inverses, $m_{X,Y} \ominus m_X = m_{X,Y} \oplus m_X^{-1}$, where the pseudo-CF $Q_{m_X^{-1}}$ corresponding to pseudo-BPA m_X^{-1} is defined as follows:

$$Q_{m_X^{-1}}(\mathsf{a}) = 1/Q_{m_X}(\mathsf{a}) \tag{21}$$

for all $\mathsf{a} \subseteq \Omega_X$ assuming $Q_{m_X}(\mathsf{a}) > 0$.

Example 2. Consider variables X and Y with $\Omega_X = \{x, \bar{x}\}$, and $\Omega_Y = \{y, \bar{y}\}$. Consider joint BPA $m_{X,Y}$ for $\{X, Y\}$ as follows:

$$m_{X,Y}(\{(x,y)\}) = 0.9,\ m_{X,Y}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) = 0.1.$$

The marginal BPA $(m_{X,Y})^{\downarrow X} = m_X$ for X is as follows:

$$m_X(\{x\}) = 0.9,\ m_X(\{x,\bar{x}\}) = 0.1$$

The computation of $Q_{m_{X,Y}} \ominus Q_{m_X}$ is shown in Table 1. The last column in the table is the pseudo-BPA corresponding to $Q_{m_{X,Y}} \ominus Q_{m_X}$ computed using Eq. (6). This is because the m_X is not explicitly included in $m_{X,Y}$ (see Theorem 2 which follows). □

A comment concerning the computation of the removal operator. In practical applications, we represent knowledge using BPAs as the number of focal elements is limited. However, when we convert a BPA to a corresponding CF, the CF is usually non-zero for all subsets of the state space. For example, in Table 1, even though $m_X^{\uparrow\{X,Y\}}$ has only two focal elements, the corresponding CF in column five is non-zero for all subsets of $\Omega_{X,Y}$. This is true because $m_X(\Omega_X) = 0.1 > 0$. As mentioned in the comments about the removal operator, we can compute removal in terms of BPAs.

In probability theory, a joint distribution $P_{X,Y}$ can always be factored into marginal $P_X = (P_{X,Y})^{\downarrow X}$ and a conditional $P_{Y|X}$ such that $P_{X,Y} = P_X \otimes P_{Y|X}$. This is not always true for belief functions. The following theorem from [15] describes when a joint belief function can be factored into a marginal and a corresponding conditional.

Theorem 2 *([15]). Suppose $m_{X,Y}$ is a BPA for $\{X, Y\}$ with corresponding CF $Q_{m_{X,Y}}$. Let m_X denote the marginal of $m_{X,Y}$ for X, i.e., $m_X = (m_{X,Y})^{\downarrow X}$. Then, $Q_{m_{X,Y}} \ominus Q_{m_X}$ is a CF if and only if there exists a conditional BPA m for Y given X such that $m_{X,Y} = m_X \oplus m$.*

Here, we give a different proof than in [15].

Proof of Theorem 2. ($\Rightarrow$) Suppose $m_{X,Y} = m_X \oplus m$. Then, for each $\mathsf{a} \subseteq \Omega_{\{X,Y\}}$,

$$(Q_{m_{X,Y}} \ominus Q_{m_X})(\mathsf{a}) = Q_{m_{X,Y}}(\mathsf{a})/Q_{m_X}(\mathsf{a}^{\downarrow X})$$

$$= Q_{m_X}(\mathrm{a}^{\downarrow X})\, Q_m(\mathrm{a})/Q_{m_X}(\mathrm{a}^{\downarrow X})$$
$$= Q_m(\mathrm{a})$$

Thus, $Q_{m_{X,Y}} \ominus Q_{m_X}$ is a well-defined CF.

($\Leftarrow$) Suppose $Q_{m_{X,Y}} \ominus Q_{m_X}$ is a well-defined CF. Let Q_m denote $Q_{m_{X,Y}} \ominus Q_{m_X}$. Then, for each $\mathrm{a} \subseteq \Omega_{\{X,Y\}}$,

$$Q_m(\mathrm{a}) = Q_{m_{X,Y}}(\mathrm{a})/Q_{m_X}(\mathrm{a}^{\downarrow X})$$

So, $Q_{m_{X,Y}}(\mathrm{a}) = Q_{m_X}(\mathrm{a}^{\downarrow X})\, Q_m(\mathrm{a})$, i.e.,

$$Q_{m_{X,Y}} = Q_{m_X} \oplus Q_m.$$

Thus, Q_{m_X} is explicitly included in $Q_{m_{X,Y}}$. $\square$

It follows from Theorem 2 that if we construct a joint BPA function m for $r \cup s$ from a marginal BPA m_r for r and a conditional BPA $m_{s|r}$ for s given r, and we remove m_r from m, then the result is a conditional BPA for s given r. Thus, our definition of a conditional belief function is consistent with the definitions in the literature.

If the conditions in Theorem 2 are not met, then a joint belief function cannot be factored into a marginal and a conditional. In Example 2, we have already demonstrated that the removal operation does not result in a conditional. Example 3 (which follows) constructs conditionals starting from a joint using Dempster's conditioning (a definition of a conditional suggested in [1]). However, when the conditionals are combined with the marginal, it results in a joint BPA that differs from the one we started with.

Example 3 *(Constructing conditionals by conditioning).* Consider the BPA $m_{X,Y}$ for $\{X,Y\}$ as described in Example 2:

$$m_{X,Y}(\{(x,y)\}) = 0.9,\ m_{X,Y}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) = 0.1.$$

The marginal BPA m_X for X is as follows:

$$m_X(\{x\}) = 0.9,\ m_X(\{x,\bar{x}\}) = 0.1$$

The BPA m_{Y_x} for Y in the context where $X = x$ is $(m_{X,Y} \oplus m_{X=x})^{\downarrow Y}$, where $m_{X=x}$ is a deterministic BPA for X such that $m_{X=x}(\{x\}) = 1$. It is easy to show that $m_{Y_x}(\{y\}) = 0.9$, and $m_{Y_x}(\{y,\bar{y}\}) = 0.1$. After conditional embedding, conditional BPA $m_{Y|x}$ for $\{X,Y\}$ is as follows:

$$m_{Y|x}(\{(x,y),(\bar{x},y),(\bar{x},\bar{y})\}) = 0.9,\ m_{Y|x}(\Omega_{X,Y}) = 0.1.$$

The BPA $m_{Y_{\bar{x}}}$ for Y in the context where $X = \bar{x}$ is $(m_{X,Y} \oplus m_{X=\bar{x}})^{\downarrow Y}$, where $m_{X=\bar{x}}$ is a deterministic BPA for X such that $m_{X=\bar{x}}(\{\bar{x}\}) = 1$. It is easy to show that $m_{Y_{\bar{x}}}(\{\bar{y}\}) = 1$. After conditional embedding, conditional BPA $m_{Y|\bar{x}}$ for $\{X,Y\}$ is as follows:

$$m_{Y|\bar{x}}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) = 1.$$

If we combine the two conditionals, we get $m_{Y|X} = m_{Y|x} \oplus m_{Y|\bar{x}}$ as follows;

$$m_{Y|X}(\{(x,y),(\bar{x},\bar{y})\}) = 0.9$$
$$m_{Y|X}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) = 0.1$$

If we combine m_X for X and $m_{Y|X}$ for $\{X,Y\}$ by Dempster's rule, we get

$$(m_X \oplus m_{Y|X})(\{(x,y)\}) = 0.81,$$
$$(m_X \oplus m_{Y|X})(\{(x,y),(x,\bar{y})\}) = 0.09,$$
$$(m_X \oplus m_{Y|X})(\{(x,y),(\bar{x},\bar{y})\}) = 0.09,$$
$$(m_X \oplus m_{Y|X})(\{(x,y),(\bar{x},\bar{y}),(\bar{x},\bar{y})\}) = 0.01.$$

Notice that $m_X \oplus m_{Y|X} \neq m_{X,Y}$. $\square$

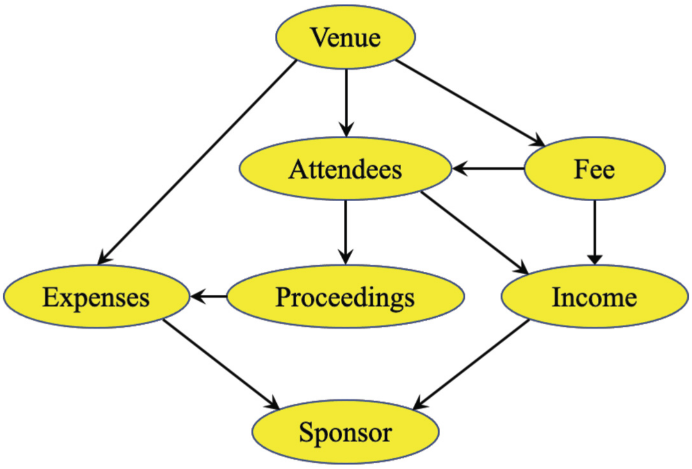

Fig. 2. The directed acyclic graph for the *Organizing a Conference* example.

3.4. Examples of conditionals

There are other ways of getting conditionals. We have discussed obtaining conditionals using Smets' conditional embedding of conditional knowledge in Section 3.2.

Another source of conditionals is deterministic knowledge. In Example 4 (which follows), we have Income (I) = Number of attendees (A) * Conference fee (F). This results in a deterministic conditional for Income given A and F as follows:

$$m_{F,A,I}(\{(200, 50, 10), (400, 50, 20), (200, 100, 20),$$
$$(400, 100, 40), (200, 150, 30), (400, 150, 60)) = 1.$$

Notice that we could use Smets' conditional embedding of each piece of knowledge (e.g., if $F = 200$, $A = 50$, then $I = 10$, etc.) and then combine all conditionals by Dempster's rule of combination. We get the same deterministic conditional. Of course, conditional embedding is unnecessary. We have to ensure we have all states of parent variables in the deterministic conditional.

Another source of conditionals is the reliability of two nodes, say X_1 with state space $\Omega_{X_1} = \{t_1, f_1\}$ and X_2 with state space $\Omega_{X_2} = \{t_2, f_2\}$. Suppose these two nodes are linked via a communication link with reliability, say 90%. This can be modeled as a BPA m_{12} for $\{X_1, X_2\}$ as follows:

$$m_{12}(\{(t_1, t_2), (f_1, f_2)\}) = 0.90$$
$$m_{12}(\Omega_{\{X_1,X_2\}}) = 0.10$$

Notice that m_{12} can be considered as a conditional for X_1 given X_2, or as a conditional for X_2 given X_1 as $m_{12}^{\downarrow X_1}$ is vacuous for X_1 and $m_{12}^{\downarrow X_2}$ is vacuous for X_2.

The list above is incomplete. We may get other examples of conditionals depending on the domain of interest.

The following example, the *Organizing a Conference*, has several examples of conditionals.

Example 4 *(Organizing a Conference).* Members of a local organizing committee for a conference would like to know whether the registration fees paid by the participants will cover the necessary expenses or whether they must find a sponsor to fund the shortfall between income and expenses. The answer depends on the selected venue, the amount of the conference fee, and other factors. Table 2 lists all variables in a simplified model.

The only disadvantage of choosing a luxury hotel is its price; it charges 40 thousand euros regardless of the number of participants, while the mountain hut charges only 10 thousand euros. However, the mountain hut venue limits the number of participants to one hundred at maximum. The organizers consider this a minor drawback because they believe some potential participants will be discouraged from participating if the location is in the mountains. The organizers will fix a higher conference fee if the venue is a luxury hotel than if it is a mountain hut. The model assumes that the venue and the conference fee will influence the number of participants. If the number of attendees is at least one hundred, they will consider including a printed version of the conference proceedings, which would cost 20 thousand euros.

Fig. 2 shows the directed acyclic graph associated with this problem. The details of the conditional BPAs are as follows.

Venue. The organizers voted for the location. Two voted for the hotel, four for the mountain hut, and four abstained. Thus, BPA m_V for V is as follows:

Table 2
The variables, names, state spaces, and meaning of the states.

Variable	Name	State space Ω	Meaning	Conditionals
V	Venue	$\{h, m\}$	luxury hotel, mountain hut	m_V
F	Fee	$\{200, 400\}$	in euros	$m_{F\|V}$
A	Attendees	$\{50, 100, 150\}$	# attendees	$m_{A\|V,F}$
I	Income	$\{10, 20, 30, 40, 60\}$	in 1000 euros	$m_{F,A,I}$
E	Expenses	$\{10, 30, 40, 60\}$	in 1000 euros	$m_{V,P,E}$
P	Proceedings	$\{p, e\}$	paper, electronic	$m_{P\|A}$
S	Sponsor	$\{y, n\}$	yes, no	$m_{E,I,S}$

$$m_V(\{h\}) = 0.2, \;\; m_V(\{m\}) = 0.4, \;\; m_V(\{h,m\}) = 0.4.$$

Fee. The organizers agreed that if the event is organized in the mountain hut, the registration fee is described by BPA m_{F_m} for F as follows:

$$m_{F_m}(\{200\}) = 0.5, m_{F_m}(\{200, 400\}) = 0.5$$

And if it is organized in a luxury hotel, then BPA m_{F_h} for F as follows:

$$m_{F_h}(\{400\}) = 0.6, m_{F_h}(\{200, 400\}) = 0.4.$$

Using Smets' conditional embedding, we get conditionals $m_{F|m}$ and $m_{F|h}$ as follows:

$$m_{F|m}(\{(m,200), (h,200), (h,400)\}) = 0.5, m_{F|m}(\Omega_{V,F}) = 0.5,$$
$$m_{F|h}(\{(h,400), (m,200), (m,400)\}) = 0.6, \;\; m_{F|h}(\Omega_{V,F}) = 0.4.$$

Therefore, conditional BPA $m_{F|V} = m_{F|m} \oplus m_{F|h}$ is as follows:

$$\begin{aligned} m_{F|V}(\{(m,200), (h,400)\}) &= 0.30, \\ m_{F|V}(\{(m,200), (h,200), (h,400)\}) &= 0.20, \\ m_{F|V}(\{(h,400), (m,200), (m,400)\}) &= 0.30, \\ m_{F|V}(\Omega_{V,F}) &= 0.20. \end{aligned}$$

Attendees. To specify the conditional $m_{A|V,F}$ using Smets' embedding, the organizers had to estimate one-dimensional BPAs for variable A in all four situations described by the combinations of values of variables V and F. Let they be

$$\begin{aligned} & m_{A_{(h,200)}}(\{150\}) = 1, \\ & m_{A_{(m,200)}}(\{50\}) = 0.2, \;\; m_{A_{(m,200)}}(\{100\}) = 0.8, \\ & m_{A_{(h,400)}}(\{100, 150\}) = 0.9, \;\; m_{A_{(h,400)}}(\Omega_A) = 0.1, \\ & m_{A_{(m,400)}}(\{50, 100\}) = 1. \end{aligned}$$

After conditional embedding and Dempster's combination, we get the conditional BPA $m_{A|F,V}$. Details are omitted.

Income. Income is the product of the number of attendees and the conference fee. This is modeled by a deterministic conditional BPA $m_{F,A,I}$ as follows;

$$\begin{aligned} & m_{F,A,I}(\{(200,50,10), (400,50,20), (200,100,20), \\ & (400,100,40), (200,150,30), (400,150,60)) = 1. \end{aligned}$$

As the marginal BPA $m_{F,A,I}^{\{F,A\}}$ is the vacuous BPA for $\{F, A\}$, it is a conditional BPA for I given $\{F, A\}$.

Proceedings. To set up conditional $m_{P|A}$ using Smets' conditional embedding, we start with the condition that the proceedings are printed only if the number of participants is at least one hundred. Therefore $m_{P_{50}}(\{e\}) = 1$. For one hundred participants, the organizers could not achieve an agreement, and therefore they assigned $m_{P_{100}}(\{e, p\}) = 1$. Finally, for 150 participants, they assigned $m_{P_{150}}(\{p\}) = 0.8$ and $m_{P_{150}}(\{e, p\}) = 0.2$. From this, we get

$$\begin{aligned} m_{P|50}(\{(50,e), (100,e), (150,e), (100,p), (150,p)\}) &= 1, \\ m_{P|100}(\Omega_{A,P}) &= 1, \\ m_{P|150}(\{(150,p), (50,e), (100,e), (50,p), (100,p)\}) &= 0.8, \\ m_{P|150}(\Omega_{A,P}) &= 0.2, \end{aligned}$$

which, using Dempster's combination $m_{P|A} = m_{P|50} \oplus m_{P|100} \oplus m_{P|150}$, yields

$$m_{P|A}(\{(50,e),(100,e),(100,p),(150,p)\}) = 0.8,$$
$$m_{P|A}(\{(50,e),(100,e),(150,e),(100,p),(150,p)\}) = 0.2.$$

Expenses. Variable E indicates how much the organizers need to cover invoices issued by the hotel (or hut) and the publishing house if the proceedings are printed. Thus, BPA $m_{V,P,E}$ for $\{V,P,E\}$ is a deterministic BPA as follows:

$$m_{V,P,E}(\{(h,p,60),(m,p,30),(h,e,40),(m,e,10)\}) = 1.$$

Sponsor. $S = y$ if expenses exceed income. This situation is specified by subset c of $\Omega_{E,I}$

$$\mathsf{c} = \{(30,10),(30,20),(40,10),(40,20),$$
$$(40,30),(60,10),(60,20),(60,30),(60,40)\}$$

The deterministic BPA $m_{E,I,S}$ for $\{E,I,S\}$ models the logical relation $E > I \Rightarrow S = y$.

$$m_{E,I,S}((\mathsf{c} \times \{y\}) \cup ((\Omega_{E,I} \setminus \mathsf{c}) \times \{n\})) = 1.$$

This belief-function graphical model, as described above, has seven variables with a joint state space of 960 states. The joint belief function m represented by the graphical model is Dempster's combination of all conditionals. If we compute the marginal of the joint for S, $m^{\downarrow S}$ using local computation [28], we get the following BPA for S:

$$m^{\downarrow S}(\{n\}) = 0.08, m^{\downarrow S}(\{y,n\}) = 0.92.$$

Thus, if we interpret belief and plausibility (corresponding to $m^{\downarrow S}$) of S as lower and upper bounds of probabilities of S, we have $0 \le P(S=y) \le 0.92$, and $0.08 \le P(S=n) \le 1$.

Why do we get such wide bounds on $P(S)$? We have a non-Bayesian prior for V and non-Bayesian conditional knowledge for F, A, and P. If we had a Bayesian prior for V, and Bayesian conditional knowledge for F, A, and P, we would get a Bayesian marginal for S with point estimates for $P(S=y)$ and for $P(S=n)$ [26]. □

4. Summary & conclusions

We have explicitly defined conditionals in the D-S theory using only the marginalization and Dempster's combination operators. The main goal of the definition is to enable the construction of directed graphical belief function models.

Conditional belief functions are also defined in [25] using an inverse of Dempster's combination operator called *removal*. Since Dempster's combination is pointwise multiplication of commonality functions followed by normalization, removal consists of the division of commonality functions followed by normalization. Thus, $m_{Y|X} = m_{X,Y} \ominus m_X$. One issue with this definition is that a conditional BPA is defined starting from a joint BPA, which is not useful in constructing a joint BPA from conditionals as in a belief-function directed graphical model. Another issue is that if m_X is not already included in $m_{X,Y}$, the division operation will result in a BPA with negative masses.

We have stated some properties of conditionals in [25], and these properties remain valid using our definition. Smets' conditional embedding [29] is one way to obtain conditionals. There are other ways to obtain conditionals; some are described and illustrated in Example 4.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Radim Jirousek and Vaclav Kratochvil report financial support was provided by Czech Science Foundation Grant No. 21-07494S. Prakash Pundalik Shenoy reports financial support was provided by Ronald G. Harper Professorship, Univ. of Kansas School of Business.

Data availability

No data was used for the research described in the article.

Acknowledgements

This study was supported by the Czech Science Foundation Grant No. 21-07494S to the first two authors, and by the Ronald G. Harper Professorship at the University of Kansas to the third author. A 12-pp. version of this paper appeared as [14]. This paper has benefitted from constructive comments from two anonymous reviewers of the *International Journal of Approximate Reasoning*.

References

[1] R.G. Almond, Graphical Belief Modeling, Chapman & Hall, London, UK, 1995.
[2] B. Ben-Yaghlane, K. Mellouli, Inference in directed evidential networks based on the transferable belief model, Int. J. Approx. Reason. 48 (2008) 399–418.
[3] B. Ben-Yaghlane, P. Smets, K. Mellouli, Belief function independence: I. The marginal case, Int. J. Approx. Reason. 29 (2002) 47–70.
[4] B. Ben-Yaghlane, P. Smets, K. Mellouli, Belief function independence: II. The conditional case, Int. J. Approx. Reason. 31 (2002) 31–75.
[5] P.K. Black, K.B. Laskey, Hierarchical evidence and belief functions, in: R.D. Shachter, T.S. Levitt, L.N. Kanal, J.F. Lemmer (Eds.), Uncertainty in Artificial Intelligence 4, in: Machine Intelligence and Pattern Recognition, vol. 9, North-Holland, Amsterdam, Netherlands, 1990, pp. 207–215.
[6] I. Boukhris, S. Benferhat, Z. Elouedi, Representing belief function knowledge with graphical models, in: H. Xiong, W. Lee (Eds.), Proceedings of the International Conference on Knowledge Science, Engineering and Management (KSEM 2011), Springer-Verlag, Berlin, Heidelberg, 2011, pp. 233–245.
[7] I. Boukhris, Z. Eloudi, S. Benferhat, Analyzing belief function networks with conditional beliefs, in: Proceedings of the 11th International Conference on Intelligent Systems Design and Applications, IEEE Explore, 2011, pp. 959–964.
[8] J. Cano, M. Delgado, S. Moral, An axiomatic framework for propagating uncertainty in directed acyclic networks, Int. J. Approx. Reason. 8 (1993) 253–280.
[9] P. Dawid, Conditional independence in statistical theory, J. R. Stat. Soc., Ser. B 41 (1979) 1–15.
[10] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (1967) 325–339.
[11] R. Fagin, J.Y. Halpern, A new approach to updating beliefs, in: P. Bonissone, M. Henrion, L. Kanal, J. Lemmer (Eds.), Uncertainty in Artificial Intelligence, vol. 6, North-Holland, 1991, pp. 347–374.
[12] P. Giang, S. Shenoy, The belief function machine: an environment for reasoning with belief functions in Matlab, Working Paper, University of Kansas School of Business, Lawrence, KS, 2003, https://pshenoy.ku.edu/Papers/BFM072503.zip.
[13] J.Y. Halpern, R. Fagin, Two views of belief: belief as generalized probability and belief as evidence, Artif. Intell. 54 (1992) 275–317.
[14] R. Jiroušek, V. Kratochvíl, P.P. Shenoy, Conditional belief functions in the Dempster-Shafer theory, in: S.L. Hégarat-Mascle, I. Bloch, E. Aldea (Eds.), Belief Functions: Theory and Applications, 7th International Conference, BELIEF 2022, in: Lecture Notes in Artificial Intelligence, vol. 13506, Springer Nature, Switzerland, 2022, pp. 207–218.
[15] R. Jiroušek, V. Kratochvíl, P.P. Shenoy, Entropy for evaluation of Dempster-Shafer belief function models, Int. J. Approx. Reason. 151 (2022) 164–181.
[16] C.T.A. Kong, Multivariate belief functions and graphical models, Ph.D. thesis, Department of Statistics, Harvard University, Cambridge, Massachusetts, 1986.
[17] C.T.A. Kong, A belief function generalization of Gibbs ensembles, Technical Report 239, University of Chicago, Department of Statistics, Chicago, IL, 1988.
[18] S.L. Lauritzen, F.V. Jensen, Local computation with valuations from a commutative semigroup, Ann. Math. Artif. Intell. 21 (1997) 51–69.
[19] J. Pearl, A. Paz, Graphoids: graph-based logic for reasoning about relevance relations, in: B.D. Boulay, D. Hogg, L. Steele (Eds.), Advances in Artificial Intelligence II, North-Holland, Amsterdam, 1987, pp. 357–363.
[20] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
[21] G. Shafer, Belief functions and parametric models, J. R. Stat. Soc., Ser. B 44 (1982) 322–352.
[22] G. Shafer, The problem of dependent evidence, Working Paper 164, University of Kansas School of Business, Lawrence, KS, 1984.
[23] G. Shafer, An axiomatic study of computation in hypertrees, Working Paper 232, University of Kansas School of Business, Lawrence, KS, 1991, http://glennshafer.com/assets/downloads/hypertrees_91WP232.pdf.
[24] G. Shafer, The problem of dependent evidence, Int. J. Approx. Reason. 79 (2016) 41–44.
[25] P.P. Shenoy, Conditional independence in valuation-based systems, Int. J. Approx. Reason. 10 (1994) 203–234.
[26] P.P. Shenoy, Making inferences in incomplete Bayesian networks: A Dempster-Shafer belief-function approach, Int. J. Approx. Reason. 160 (2023), in press, https://pshenoy.ku.edu/Papers/IJAR23a.pdf.
[27] P.P. Shenoy, On distinct belief functions in the Dempster-Shafer theory, in: E. Miranda, I. Montes, E. Quaeghebeur, B. Vantaggi (Eds.), Proceedings of ISIPTA-23, Proc. Mach. Learn. Res. 215 (2023), in press, https://pshenoy.ku.edu/Papers/ISIPTA23a.pdf.
[28] P.P. Shenoy, G. Shafer, Axioms for probability and belief-function propagation, in: R.D. Shachter, T. Levitt, J.F. Lemmer, L.N. Kanal (Eds.), Uncertainty in Artificial Intelligence 4, in: Machine Intelligence and Pattern Recognition Series, vol. 9, North-Holland, Amsterdam, Netherlands, 1990, pp. 169–198.
[29] P. Smets, Un modele mathematico-statistique simulant le processus du diagnostic medical, Ph.D. thesis, Free University of Brussels, 1978.
[30] P. Smets, The concept of distinct evidence, in: B. Bouchon-Meunier (Ed.), Proceedings of the 4th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'92), 1992, pp. 789–794.
[31] P. Smets, Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem, Int. J. Approx. Reason. 9 (1993) 1–35.
[32] H. Xu, P. Smets, Reasoning in evidential networks with conditional belief functions, Int. J. Approx. Reason. 14 (1996) 155–185.