# On Conditional Belief Functions in the Dempster-Shafer Theory

Radim Jiroušek[1,2], Václav Kratochvíl[1,2], and Prakash P. Shenoy[3(✉)]

[1] Faculty of Management, University of Economics, Jindřichův Hradec, Czechia
[2] Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czechia
{radim,velorex}@utia.cas.cz
[3] School of Business, University of Kansas, Lawrence, KS 66045, USA
pshenoy@ku.edu
https://pshenoy.ku.edu

**Abstract.** The primary goal is to define conditional belief functions in the Dempster-Shafer theory. We do so similar to the notion of conditional probability tables in probability theory. Conditional belief functions are necessary for constructing directed graphical belief function models in the same sense as conditional probability tables for constructing Bayesian networks. Besides defining conditional belief functions, we state and prove a few basic properties of conditionals. We provide several examples of conditional belief functions, including those obtained by Smets' conditional embedding.

**Keywords:** Dempster-Shafer belief function theory · Conditional belief functions · Smets' conditional embedding

## 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 [4,13], 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 rules and corresponding semantics. The D-S theory uses Dempster's combination rule. [5] proposes an alternative combination rule interpreting belief functions as credal sets [7]. 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 [18]. Other contributions on conditional belief functions are (in chronological order) Shafer [14,15], Cano *et al.* [3], Shenoy [16], Almond [1], and Xu and Smets [19].

Shafer [14] is concerned about parametric models. There is a discrete parameter variable $\Theta$ and a data variable $X$. We have a prior basic probability assignment (BPA) $m_\Theta$ for $\Theta$. We have a conditional model for the data, BPA $m_{X|\theta}$


S. Le Hégarat-Mascle et al. (Eds.): BELIEF 2022, LNAI 13506, pp. 207–218, 2022.
https://doi.org/10.1007/978-3-031-17801-6_20

for $X$ given $\theta \in \Omega_\Theta$. Based on a dataset of $n$ independent observations of $X$, the task is to compute the posterior belief function for $\Theta$. The BPAs $m_{X|\theta}$ for $X$ given $\theta \in \Theta$ are converted to a conditional BPA $m_{\theta,X}$ for $(\Theta, X)$ using Smets' conditional embedding. The marginal of $m_{\theta,X}$ for $\Theta$ is vacuous. For all $\theta \in \Omega_\Theta$, the conditionals BPA $m_{\theta,X}$ are then combined using Dempster's rule resulting in the conditional $m_{X|\Theta}$. This assumes that the BPAs $m_{\theta,X}$ are distinct, which may be reasonable if the number of elements of $\Omega_\Theta$ is small. Shafer also looks at the case where BPAs $m_{\theta,X}$ are not independent, and some known distributions describe the dependency.

Shafer [15] discusses conditionals abstractly as potentials that extend the domain of a potential. He calls conditionals '*continuers*.' Thus, $\psi$ is a continuer of $\sigma$ from $a$ to $a \cup b$ if and only if $\sigma^{\downarrow a} \oplus \psi = \sigma^{\downarrow a \cup b}$. Here, $\sigma^{\downarrow a}$ denotes the marginal of $\sigma$ for $a$, $\oplus$ denotes Dempster's combination operator, and $a$ and $b$ are disjoint subsets of variables. The paper's focus is on the computation of marginals, but there are some interesting properties of continuers stated.

Cano *et al.* [3] define conditionals abstractly in the framework of valuation-based systems, but they do require that the marginal $m(b|a)^{\downarrow a}$ of conditional $m(b|a)$ is a vacuous valuation for $a$. The focus is on finding marginals by propagating conditional valuations in a directed acyclic graph.

Shenoy [16] 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 pointwise division of commonality functions followed by normalization. If $\sigma$ is a BPA for subset $s$ of variables, and $a$ and $b$ are disjoint subsets of $s$, then conditional belief function $\sigma(b|a)$ is defined as $\sigma^{\downarrow a \cup b} \ominus \sigma^{\downarrow a}$. A consequence of this definition is that the marginal of $\sigma(b|a)$ for $a$ is vacuous for $a$. One disadvantage of this definition is that conditionals are defined starting from the joint. This is not helpful in constructing joint belief functions. We say $\sigma^{\downarrow a}$ is *included* in $\sigma^{\downarrow a \cup b}$ if $\sigma^{\downarrow a \cup b} = \sigma^{\downarrow a} \oplus \sigma(b|a)$. Another disadvantage is that if $\sigma^{\downarrow a}$ is not included in $\sigma^{\downarrow a \cup b}$, $\sigma(b|a)$ may result in a BPA with negative masses. Such BPAs are called quasi-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 $x \in \Omega_X$ as follows. Suppose $m_{X=x}$ is a deterministic BPA for $X$ such that $m_{X=x}(\{x\}) = 1$. Then $m_{Y|x}$ is defined as $(m_{X,Y} \oplus m_{X=x})^{\downarrow X}$. He then discusses the problem of going from conditionals to joint and argues that there isn't 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 conditional BPA $m_{Y|x}$ for $Y$ is embedded into a BPA $m_{x,Y}$ for $(X, Y)$ (details of Smets' conditional embedding are discussed in Sect. 3). Next, BPA $m_{Y|X}$ for $(X, Y)$ is constructed from conditional embeddings $m_{x,Y}$ for $x \in \Omega_X$ as follows:

$$m_{Y|X} = \oplus\{m_{x,Y} : x \in \Omega_X\}. \tag{1}$$

[1] This phenomenon has been observed, e.g., in [11,16], and [12]. An example is given in [10].

Equation (1) implicitly assumes that the conditionally embedded BPAs $m_{x,Y}$ are distinct. Almond claims this assumption is unrealistic except for the case where we start from conditional BPAs $m_{Y|x}$ that are Bayesian.

Xu and Smets [19] discuss conditionals $m_{Y|\mathsf{a}}$ for $Y$ when proposition $\mathsf{a}$ is observed, where $\emptyset \neq \mathsf{a} \in 2^{\Omega_X}$. Let $m_{\mathsf{a},Y}$ denote the BPA for $(X, Y)$ after conditional embedding of $m_{Y|\mathsf{a}}$. [1] and [19] discuss Dempster's combination of all such conditionals:

$$\oplus \{m_{\mathsf{a},Y} : \emptyset \neq \mathsf{a} \in 2^{\Omega_X}\}. \tag{2}$$

While it may be reasonable to assume that $m_{x,Y}$ for $x \in \Omega_X$ are distinct as in Eq. (1), assuming that all BPAs $m_{\mathsf{a},Y}$ for $\emptyset \neq \mathsf{a} \in 2^{\Omega_X}$ are distinct may be unreasonable. The focus of [19] is on computing marginals.

We do not start with a joint BPA when constructing a directed graphical belief function model. Instead, we construct a joint BPA using priors and conditionals. In this context, the current definitions in the literature are not helpful. What exactly is a conditional BPA? What are their properties? This is the primary goal of this article.

An outline of the remainder of the paper is as follows. In Sect. 2, we review the basics of D-S theory. In Sect. 3, we define conditional belief functions, and state some properties. Also, we describe where conditionals come from, including Smets' conditional embedding. We describe Almond's captain's problem [1], a directed graphical belief function model with several examples of conditionals. In Sect. 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 [4,13].

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.

Consider a set $s$ of variables. For each $X \in s$, let $\Omega_X$ denote its finite state space, and let $\Omega_s$ denote $\times_{X \in s} \Omega_X$. Let $2^{\Omega_s}$ denote the set of all subsets of $\Omega_s$. A *basic probability assignment* (BPA) $m$ for $s$ is a function $m : 2^{\Omega_s} \to [0, 1]$ such that

$$m(\emptyset) = 0, \text{ and } \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_s}} m(\mathsf{a}) = 1. \tag{3}$$

$m$ represents some knowledge about variables in $s$, and we say the *domain* of $m$ is $s$. $m(\mathsf{a})$ is the probability assigned to the proposition represented by the subset $\mathsf{a}$ of $\Omega_s$. Subsets $\mathsf{a}$ 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 $\Omega_s$, we say $m$ is *vacuous*.

The knowledge encoded in a BPA $m$ can be represented by a corresponding commonality function. The *commonality function* (CF) $Q_m$ corresponding to BPA $m$ for $s$ is such that for all $\mathsf{a} \in 2^{\Omega_s}$,

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}). \tag{4}$$

$Q_m(\mathsf{a})$ represents the probability mass that could move to every state in $\mathsf{a}$. $Q_m$ has exactly the same information as $m$. Given a CF $Q$ for $s$, we can recover the corresponding BPA $m_Q$ for $s$ as follows [13]: For all $\mathsf{a} \in 2^{\Omega_s}$,

$$m_Q(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_s}:\mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q(\mathsf{b}). \tag{5}$$

Thus, $Q : 2^{\Omega_s} \to [0, 1]$ is a CF for $s$ if and only if

$$Q(\emptyset) = 1 \tag{6}$$

$$\sum_{\mathsf{b} \in 2^{\Omega_s}:\mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q(\mathsf{b}) \geq 0 \quad \text{for all } \emptyset \neq \mathsf{a} \in 2^{\Omega_s}, \text{ and} \tag{7}$$

$$\sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_s}} (-1)^{|\mathsf{a}|+1} Q(\mathsf{a}) = 1. \tag{8}$$

Equation (6) follows from Eq. (4), Eq. (7) corresponds to non-negativity of BPA values, and Eq. (8) corresponds to the second equation in Eq. (3).

There are two basic inference operators in the D-S theory, marginalization and combination.

Suppose $m$ is a BPA for a set of variables $r$ with state space $\Omega_r = \times_{X \in r} \Omega_X$ 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 $r \setminus s$. Projection of states means dropping some coordinates. If $(x, y) \in \Omega_{X,Y}$, then $(x, y)^{\downarrow X} = x$. Projection of subset of states is achieved by projecting every state in the subset. Suppose $a \in 2^{\Omega_{X,Y}}$. Then, $a^{\downarrow X} = \{x \in 2^{\Omega_X} : (x, y) \in \mathsf{a}\}$. Suppose $m$ is a BPA for $r$. Then, the marginal of $m$ for $s \subseteq r$, denoted by $m^{\downarrow s}$, is a BPA for $s$ such that for each $\mathsf{a} \in 2^{\Omega_s}$,

$$m^{\downarrow s}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_r}:\mathsf{b}^{\downarrow s} = \mathsf{a}} m(\mathsf{b}). \tag{9}$$

Dempster's combination rule is described using commonality functions. Consider two distinct BPAs $m_1$ for $r$ and $m_2$ for $s$, and let $Q_1$ and $Q_2$ denote the corresponding commonality functions. Then, as showed in [13], for all $\emptyset \neq \mathsf{a} \in 2^{\Omega_{r \cup s}}$

$$(Q_1 \oplus Q_2)(\mathsf{a}) = K^{-1} Q_1(\mathsf{a}^{\downarrow r})\, Q_2(\mathsf{a}^{\downarrow s}), \tag{10}$$

where $K$ is a normalization constant defined as follows:

$$K = \sum_{\emptyset \neq \mathsf{a} \in \Omega_{r \cup s}} (-1)^{|\mathsf{a}|+1} Q_1(\mathsf{a}^{\downarrow r})\, Q_2(\mathsf{a}^{\downarrow s}). \tag{11}$$

$(1 - K)$ can be regarded as a measure of *conflict* between $m_1$ and $m_2$. If $K = 1$, there is no conflict, and if $K = 0$, there is total conflict and Dempster's combination $Q_1 \oplus Q_2$ is undefined.

It is easy to show that Dempster's combination is commutative and associative: $m_1 \oplus m_2 = m_2 \oplus m_1$, and $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$.

There is an important property satisfied by marginalization and Dempster's combination rule called the *local computation* property [17]. Suppose $m_1$ is a

BPA for $r$ and $m_2$ is a BPA for $s$ (subsets $r$ and $s$ may not be disjoint) and suppose $X \in r$ and $X \notin s$. Then,

$$(m_1 \oplus m_2)^{\downarrow (r \cup s) \setminus \{X\}} = (m_1)^{\downarrow r \setminus \{X\}} \oplus m_2 \tag{12}$$

This property is the basis of computing marginals of joint belief functions. [6] describes an implementation of a local computation algorithm for computing marginals of graphical belief function models.

Next, we define the removal operator, which is motivated by the following situation in probability theory. Suppose $P_{X,Y}$ is a joint probability mass function (PMF) for $(X, Y)$, and we need to compute the conditional probability table (CPT) $P_{Y|X}$. We know that $P_{X,Y} = P_X \otimes P_{Y|X}$, where $P_X = (P_{X,Y})^{\downarrow X}$ is the marginal PMF for $X$, and $\otimes$ is the probabilistic combination operator pointwise multiplication followed by normalization. This suggests that $P_{Y|X} = P_{X,Y} \oslash P_X$, where $\oslash$ is the inverse combination operator, pointwise division followed by normalization. If $P_X(x) = 0$, then $P_{X,Y}(x, y)$ must also be zero, and we can consider $0/0$ as undefined (using the symbol $0/0 = ?$) or define it as 1. Thus, if we regard combination $\otimes$ as aggregation of knowledge, then $\oslash$ can be regarded as removal of knowledge, and computing a CPT $P_{Y|X}$ is removing $P_X$ from $P_{X,Y}$.

As we saw in Eq. (10), Dempster's combination is pointwise multiplication of CFs followed by normalization. Thus, removal in the D-S theory can be defined as pointwise division of CFs followed by normalization. Formally, suppose $Q_{X,Y}$ is a joint CF for $(X, Y)$, and let $Q_X = (Q_{X,Y})^{\downarrow X}$ denote the marginal CF for $X$. Then, we define removal of $Q_X$ from $Q_{X,Y}$ as follows: For all $\emptyset \neq \mathsf{a} \in 2^{\Omega_{X,Y}}$,

$$(Q_{X,Y} \ominus Q_X)(\mathsf{a}) = K^{-1}\, Q_{X,Y}(\mathsf{a})/Q_X(\mathsf{a}^{\downarrow X}), \tag{13}$$

where $K$ is a normalization constant given by:

$$K = \sum_{\emptyset \neq \mathsf{a} \in 2^{\Omega_{X,Y}}} (-1)^{|\mathsf{a}|+1}\, Q_{X,Y}(\mathsf{a})/Q_X(\mathsf{a}^{\downarrow X}) \tag{14}$$

As in the probabilistic case, if $Q_X(\mathsf{a}^{\downarrow X}) = 0$, then $Q_{X,Y}(\mathsf{a})$ must also be 0, and we can define $0/0$ as 1.

Unlike probability theory, if we start with an arbitrary joint CF $Q_{X,Y}$, then $Q_{X,Y} \ominus Q_X$ may fail to be a CF because the corresponding BPA has negative masses adding to 1[2]. In the next section, we state a proposition that characterizes when removal results in a well-defined CF.

## 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. Our task is constructing a joint using conditional belief functions as in a graphical model. We begin with the probabilistic case.

[2] An example is given in [10].

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[3] $P_X(x) > 0$. 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 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 $X$, then we get a potential that is identically 1 for all values of $x \in \Omega_X$, which is the vacuous potential in probability theory.
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 $(P_{X,Y})^{\downarrow X} = P_X$.

A formal definition of a conditional belief function for $Y$ given $X$ in the D-S theory follows.

**Definition 1.** *Suppose $m_{Y|X}$ is a BPA for $(X, Y)$, where $X$ and $Y$ are distinct variables. We say $m_{Y|X}$ is a conditional BPA for $Y$ given $X$ if and only if*

1. *$(m_{Y|X})^{\downarrow X}$ is a vacuous BPA for $X$, and*
2. *for any BPA $m_X$ for $X$, $m_X$ and $m_{Y|X}$ are distinct. Thus, $m_X \oplus m_{Y|X}$ is a BPA for $(X, Y)$.*

The first condition says that $m_{Y|X}$ tells us nothing about $X$. We will refer to the BPA $m_X \oplus m_{Y|X}$ as the *joint* BPA for $(X, Y)$ and denote it by $m_{X,Y}$. It follows from the local computation property (Eq. (12)) that $(m_{X,Y})^{\downarrow X} = (m_X \oplus m_{Y|X})^{\downarrow X} = m_X \oplus (m_{Y|X})^{\downarrow X} = m_X$. Thus, the second condition says the conditional $m_{Y|X}$ allows us to *extend* any BPA $m_X$ for $X$ to a joint BPA $m_{X,Y}$ for $(X, Y)$ without changing its marginal for $X$. Notice that $m_X$ and $m_{Y|X}$ are non-conflicting, i.e., the normalization constant $K$ in $m_X \oplus m_{Y|X}$ is 1 (Eq. (11)).

Given a conditional BPA $m_{Y|X}$ for $Y$ given $X$, we will refer to $Y$ as the *head* of the conditional, and $X$ as the *tail*. A conditional describes the dependency between the head and tail variables. Although we have defined a conditional BPA with the head and tail being single variables, the definition generalizes when the head and tail are disjoint subsets of variables.

[3] If $P_X(x) = 0$, then the conditional has no effect on the joint, and 0/0 can be left undefined, or defined as 1.

**Definition 2.** *Suppose $r$ and $s$ are disjoint subsets of variables, and $m_{s|r}$ 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. *for any BPA $m_r$ for $r$, $m_r$ and $m_{s|r}$ are distinct. Thus, $m_r \oplus m_{s|r}$ is a BPA for $r \cup s$.*

In a directed graphical belief function model, we have a conditional associated with each variable $X$ in the model. 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, we can consider priors as conditionals with empty tails. For such BPAs, the first condition in the definition is trivially true as the sum of the probability masses in a BPA is 1.

*Properties of Conditionals.* The following lemma was stated in [16] 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.

**Lemma 1.** *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, the following statements are true.*

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. *$(m_{s|r} \oplus m_{t|r\cup s})^{\downarrow r\cup t} = m_{t|r}$.*

*Proof.* 1. $m_r$, $m_{s|r}$, and $m_{t|r\cup s}$ are all distinct by definition of conditionals. Thus,
$m_r \oplus m_{s|r} \oplus m_{t|r\cup s} = (m_r \oplus m_{s|r}) \oplus m_{t|r\cup s} = m_{r\cup s} \oplus m_{t|r\cup s} = m_{r\cup s\cup t}$.

2. Let $\iota_r$ denote the vacuous BPA for $r$. Using the local computation property,

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

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. First, notice that $((m_{s|r})^{\downarrow r\cup s'})^{\downarrow r} = (m_{s|r})^{\downarrow r} = \iota_r$. Suppose $m_r$ is a BPA for $r$. As $m_r$ and $m_{s|r}$ are distinct, $m_r$ and $m_{s'|r}$ are distinct. Thus, $m_r \oplus (m_{s|r})^{\downarrow r\cup s'} = m_{r\cup s'}$.

4. Using the local computation property,

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

Suppose $m_r$ is a BPA for $r$. As $m_r$, $m_{s|r}$, and $m_{t|r\cup s}$ are all distinct,

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

□

*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 [18]. To describe conditional embedding, consider the case of two variables $X$ and $Y$. To describe the dependency between $X$ and $Y$, suppose that when $X = x$, our belief in $Y$ is described by a BPA $m_{Y|x}$ for $Y$. Thus, $m_{Y|x} : 2^{\Omega_Y} \to [0, 1]$ such that $\sum_{\mathsf{a} \in 2^{\Omega_Y}} m_{Y|x}(\mathsf{a}) = 1$. The BPA $m_{Y|x}$ for $Y$ needs to be embedded into a BPA $m_{x,Y}$ for $(X, Y)$ such that

1. $m_{x,Y}$ is a conditional BPA for $(X, Y)$, i.e., $(m_{x,Y})^{\downarrow X}$ is vacuous BPA for $X$, and
2. when we add the belief that $X = x$ and marginalize the result to $Y$, we obtain $m_{Y|x}$.

One way to do this is to take each focal element $\mathsf{b} \in 2^{\Omega_Y}$ of $m_{Y|x}$, and convert it to the corresponding focal element

$$(\{x\} \times \mathsf{b}) \cup ((\Omega_X \setminus \{x\}) \times \Omega_Y) \in 2^{\Omega_{X,Y}} \tag{15}$$

of BPA $m_{x,Y}$ for $(X, Y)$ with the same mass. It is easy to confirm that this method of embedding satisfies both conditions mentioned above. If 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 $X$, then we do conditional embedding of each of these BPAs and then combine the 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 $Y$ given $X = x$ as follows:

$$m_{Y|x}(y) = 0.8,\ m_{Y|x}(\Omega_Y) = 0.2,$$

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

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

Similarly, if we have a BPA $m_{Y|\bar{x}}$ for $Y$ given $X = \bar{x}$ 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 $m_{\bar{x},Y}$ for $(X, Y)$ is as follows:

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

Assuming we have these two BPAs, and their corresponding embeddings, it is clear that the two BPA $m_{x,Y}$ and $m_{\bar{x},Y}$ are distinct, and can be combined with Dempster's rule of combination, resulting in the conditional BPA $m_{Y|X} = m_{x,Y} \oplus m_{\bar{x},Y}$ for $(X, Y)$. $m_{Y|X}$ has the following properties. First, $(m_{Y|X})^{\downarrow X} = \iota_X$, where $\iota_X$ denotes the vacuous BPA for $X$. Second, 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., $(m_{Y|X} \oplus m_{X=x})^{\downarrow Y} = m_{Y|x}$. Third, $(m_{Y|X} \oplus m_{X=\bar{x}})^{\downarrow Y} = m_{Y|\bar{x}}$. $m_{Y|X}$ is the belief function equivalent of CPT $P_{Y|X}$ in probability theory. □

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 true in the D-S theory. The following proposition describes when a joint belief function can be factored into a marginal and a conditional.

**Proposition 1.** *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 BPA $m$ for $\{X, Y\}$ such that $m_{X,Y} = m_X \oplus m$, and $m$ is a conditional for $Y$ given $X$.*

A proof of this proposition can be found in [8]. The proposition states that if we remove BPA $m_X$ from $m_{X,Y}$ such that $m_X$ is included in $m_{X,Y}$ in the sense that $m_{X,Y}$ is Dempster's combination of the marginal $m_X$ for $X$ and a conditional $m$ for $Y$ given $X$, then such removal always results in a well-defined CF.

Smets' conditional embedding is only one way to obtain conditionals. Black and Laskey [2] propose other methods to get conditionals. The following example from [1], called the captain's problem, has many examples of conditionals. The description of Almond's captain's problem is taken from [9].

*Example 2 (Captain's problem).* A ship's captain is concerned about how many days his ship may be delayed before arrival at a destination. The arrival delay is the sum of departure delay and sailing delay. Departure delay may be a result of maintenance (at most one day), loading delay (at most one day), or a forecast of bad weather (at most one day). Sailing delays may result from bad weather (at most one day) and whether repairs are needed at sea (at most one day). If maintenance is done before sailing, chances of repairs at sea are less likely. The forecast is 80% reliable. The captain knows the loading delay and whether maintenance is done before departure.

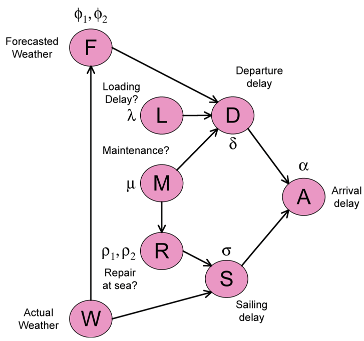


**Fig. 1.** The directed acyclic graph for the captain's problem. The Greek alphabets adjacent to a variable denote the prior or conditional or evidence associated with the variable.

**Table 1.** The variables, their state spaces, and associated conditionals in the captain's problem.

| Variable | Name | State space, $\Omega$ | Associated conditional |
|---|---|---|---|
| $W$ | Actual weather | $\{g_w, b_w\}$ | vacuous for $W$ |
| $F$ | Forecasted weather | $\{g_f, b_f\}$ | $\phi_1$ for $F\|W$ |
| $L$ | Loading delay? | $\{t_l, f_l\}$ | $\lambda$ for $L$ |
| $M$ | Maintenance done? | $\{t_m, f_m\}$ | $\mu$ for $M$ |
| $R$ | Repair at sea needed? | $\{t_r, f_r\}$ | $\rho_1, \rho_2$ for $R$ given $M = t_m, t_f$, resp. |
| $D$ | Departure delay (in days) | $\{0, 1, 2, 3\}$ | $\delta$ for $D\|\{F, L, M\}$ |
| $S$ | Sailing delay (in days) | $\{0, 1, 2, 3\}$ | $\sigma$ for $S\|\{W, R\}$ |
| $A$ | Arrival delay (in days) | $\{0, 1, 2, 3, 4, 5, 6\}$ | $\alpha$ for $A\|\{D, S\}$ |

Table 1 describes the variables, their state spaces, and associated conditionals, and Fig. 1 shows the directed acyclic graph associated with this problem. The details of some of the conditional BPAs are as follows.

1. Weather forecast is 80% accurate. $\phi_1$ is a conditional BPA for $F$ given $W$.

$$\phi_1(\{(g_w, g_f), (b_w, b_f)\}) = 0.8,\ \phi_1(\Omega_{W,F}) = 0.2.$$

2. Bad weather and repair at sea each adds a day to sailing delay. This proposition is true 90% of the time. $\sigma$ is a conditional for $S$ given $(W, R)$.

$$\sigma(\{(g_w, f_r, 0), (b_w, f_r, 1), (g_w, t_r, 1), (b_w, t_r, 2)\}) = 0.9,\ \sigma(\Omega_{W,R,S}) = 0.1.$$

3. Departure delay may be a result of maintenance (at most 1 day), loading delay (at most 1 day), or a forecast of bad weather (at most 1 day). $\delta$ is a deterministic conditional BPA for $D$ given $\{F, L, M\}$.

$$\delta(\{(g_f, f_l, f_m, 0), (b_f, f_l, f_m, 1), (g_f, t_l, f_m, 1), (g_f, f_l, t_m, 1),\\ (b_f, t_l, f_m, 2), (b_f, f_l, t_m, 2), (g_f, t_l, t_m, 2), (b_f, t_l, t_m, 3)\}) = 1.$$

4. The arrival delay is the sum of departure delay and sailing delay. $\alpha$ is a deterministic conditional BPA for $A$ given $\{D, S\}$.

$$\alpha(\{(0,0,0), (0,1,1), (0,2,2), (0,3,3), (1,0,1), (1,1,2), (1,2,3), (1,3,4),$$

$$(2,0,2), (2,1,3), (2,2,4), (2,3,5), (3,0,3), (3,1,4), (3,2,5), (3,3,6)\}) = 1.$$

## 4 Summary and 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 [16] 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 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. Another issue is that if $m_X$ is not already included in $m_{X,Y}$, the removal operation may result in a BPA with negative masses. We have stated some properties of conditionals given in [16] and these properties remain valid using our definition. Smets' conditional embedding [18] is one way to obtain conditionals. There are other ways to obtain conditionals, and some examples of conditionals are described using Almond's captain's problem [1].

**Acknowledgments.** This study was supported by the Czech Science Foundation Grant No. 19-06569S to the first two authors and by the Ronald G. Harper Professorship at the University of Kansas to the third author.

## References

1. Almond, R.G.: Graphical Belief Modeling. Chapman & Hall, London, UK (1995)
2. Black, P.K., Laskey, K.B.: Hierarchical evidence and belief functions. In: Shachter, R.D., Levitt, T.S., Kanal, L.N., Lemmer, J.F. (eds.) Uncertainty in Artificial Intelligence 4, Machine Intelligence and Pattern Recognition, vol. 9, pp. 207–215. North-Holland, Amsterdam, Netherlands (1990)
3. Cano, J., Delgado, M., Moral, S.: An axiomatic framework for propagating uncertainty in directed acyclic networks. Int. J. Approximate Reasoning **8**(4), 253–280 (1993)
4. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**(2), 325–339 (1967)
5. Fagin, R., Halpern, J.Y.: A new approach to updating beliefs. In: Bonissone, P., Henrion, M., Kanal, L., Lemmer, J. (eds.) Uncertainty in Artificial Intelligence 6, pp. 347–374, North-Holland (1991)
6. Giang, P., Shenoy, S.: The belief function machine: an environment for reasoning with belief functions in Matlab. Working paper, University of Kansas School of Business, Lawrence, KS 66045 (2003). https://pshenoy.ku.edu/Papers/BFM072503.zip
7. Halpern, J.Y., Fagin, R.: Two views of belief: belief as generalized probability and belief as evidence. Artif. Intell. **54**(3), 275–317 (1992)
8. Jiroušek, R., Kratochvíl, V., Shenoy, P.P.: Entropy for evaluation of Dempster-Shafer belief function models. Working Paper 342, University of Kansas School of Business, Lawrence, KS 66045 (2022). https://pshenoy.ku.edu/Papers/WP342.pdf
9. Jiroušek, R., Kratochvíl, V., Shenoy, P.P.: Computing the decomposable entropy of graphical belief function models. In: Studený, M., Ay, N., Coletti, G., Kleiter, G.D., Shenoy, P.P. (eds.) Proceedings of the 12th Workshop on Uncertain Processing (WUPES 2022), pp. 111–122. MatfyzPress, Prague, Czechia (2022). https://pshenoy.ku.edu/Papers/WUPES22a.pdf
10. Jiroušek, R., Shenoy, P.P.: Compositional models in valuation-based systems. Int. J. Approximate Reasoning **55**(1), 277–293 (2014)

11. Kong, A.: Multivariate belief functions and graphical models. Ph.D. thesis, Department of Statistics, Harvard University, Cambridge, Massachusetts (1986)
12. Lauritzen, S.L., Jensen, F.V.: Local computation with valuations from a commutative semigroup. Ann. Math. Artif. Intell. **21**(1), 51–69 (1997)
13. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
14. Shafer, G.: Belief functions and parametric models. J. Roy. Stat. Soc. B **44**(3), 322–352 (1982)
15. Shafer, G.: An axiomatic study of computation in hypertrees. Working Paper 232, University of Kansas School of Business, Lawrence, KS 66045 (1991). http://glennshafer.com/assets/downloads/hypertrees_91WP232.pdf
16. Shenoy, P.P.: Conditional independence in valuation-based systems. Int. J. Approximate Reasoning **10**(3), 203–234 (1994)
17. Shenoy, P.P., Shafer, G.: Axioms for probability and belief-function propagation. In: Shachter, R.D., Levitt, T., Lemmer, J.F., Kanal, L.N. (eds.) Uncertainty in Artificial Intelligence 4, Machine Intelligence and Pattern Recognition Series, vol. 9, pp. 169–198. North-Holland, Amsterdam, Netherlands (1990)
18. Smets, P.: Un modele mathematico-statistique simulant le processus du diagnostic medical. Ph.D. thesis, Free University of Brussels (1978)
19. Xu, H., Smets, P.: Reasoning in evidential networks with conditional belief functions. Int. J. Approximate Reasoning **14**(2–3), 155–185 (1996)