

Compositional Models in Valuation-Based Systems

Radim Jiroušek

Faculty of Management, University of Economics, and
Institute of Information Theory and Automation, Academy of Sciences
Jindřichův Hradec and Prague, Czech Republic
radim@utia.cas.cz

&

Prakash P. Shenoy

School of Business, University of Kansas
Lawrence, Kansas, USA
pshenoy@ku.edu

Abstract

Compositional models were initially described for discrete probability theory, and later extended for possibility theory and for belief functions in Dempster-Shafer (D-S) theory of evidence. Valuation-based system (VBS) is an unifying theoretical framework generalizing some of the well known and frequently used uncertainty calculi. This generalization enables us to not only highlight the most important theoretical properties necessary for efficient inference (analogous to Bayesian inference in the framework of Bayesian network), but also to design efficient computational procedures. Some of the specific calculi covered by VBS are probability theory, a version of possibility theory where combination is the product t -norm, Spohn's epistemic belief theory, and D-S belief function theory. In this paper, we describe compositional models in the general framework of VBS using the semantics of no-double counting, which is central to the VBS framework. Also, we show that conditioning can be expressed using the composition operator. We define a special case of compositional models called decomposable models, again in the VBS framework, and demonstrate that for the class of decomposable compositional models, conditioning can be done using local computation. As all results are obtained for the VBS framework, they hold in all calculi that fit in the VBS framework. For the D-S theory of belief functions, the compositional model defined here differs from the one studied by Jiroušek, Vejnarová, and Daniel. The latter model can also be described in the VBS framework, but with a combination operator that is different from Dempster's rule of combination. For the version of possibility theory in which combination is the product t -norm, the compositional model defined here reduces to the one studied by Vejnarová.

1. Introduction

The framework of valuation-based systems (VBS) was introduced in [28, 32, 34]. The main idea behind VBS is to capture the common features of various uncertainty calculi and other domains such as optimization, decision-making theories, database systems, and solving systems of equations. Briefly, knowledge about a set of variables is represented by a set of functions called valuations. Each valuation is associated with a subset of variables. There are two operators called combination and marginalization. Combination allows us to aggregate knowledge, and marginalization allows us to coarsen knowledge to a smaller set of variables. The combination of all valuations, called the joint valuation, represents the joint knowledge of all variables. Making inferences can be described as finding marginals of the joint valuation for variables of interest. The VBS framework can be used to describe various uncertainty theories such as probability theory, a version of possibility theory where combination is the product t -norm [43], Spohn's epistemic belief theory [37, 30], and Dempster-Shafer (D-S) belief function theory [26]. It can also be used to describe, e.g., propositional logic [29], solving systems of equations [25], optimization using dynamic programming [2, 31], Bayesian decision-making by maximizing expected utility [33], relational database theory [42], and other domains [24].

Besides the marginalization and combination operators in the VBS framework, we define an additional operator called removal. Removal is an inverse of combination, and is useful for defining conditionals in the VBS framework. Conditionals are useful in characterizing conditional independence relations. All of these operators are required to satisfy some basic properties described as axioms. These axioms enable us to make inferences using local computation, using architectures such as the Shenoy-Shafer architecture [35] that uses only the combination and marginalization operators, and the Lauritzen-Spiegelhalter architecture [20] that uses the combination, marginalization, and removal operators. The main focus of VBS is to enable local computation of marginals of the joint valuation.

The VBS framework has been expanded, and studied further in greater mathematical depth. Shafer [27] provides an axiomatic treatment of conditionals called continuers, which are defined without explicit reference to a removal operator. Lauritzen and Jensen [20] describe an alternative axiomatization of the removal operator. Kohlas [17] studies VBS using abstract algebra, and also studies a class of VBS (called information algebras) where the valuations are idempotent. Kohlas and Wilson [18] link VBS to the algebraic theory of semirings. Finally, Pouly and Kohlas [24] describe local computation in VBS in great detail, including different architectures, and normalization, and provide many examples of domains that fit in the VBS framework.

In a Bayesian network model, one usually starts with a specification of the joint probability distribution that is factorized into conditionals for each variable given a subset of variables. The joint probability distribution is then obtained as the combination of all the conditionals, i.e., a fundamental assumption of a Bayesian network model is that there is no double-counting of knowledge in combining all conditionals to form the joint distribution. In a compositional model, one starts from a different starting point. One starts with a set of marginal probability distributions, where each marginal distribution is for some subset of variables. We cannot combine the marginal distributions as this would lead to double counting of knowledge (for those variables that are in the intersections of subsets of variables for which we have marginals). This is why we use the composition operator because it allows us to aggregate knowledge in the marginal distributions without double counting of knowledge. We assume that each variable is included in some subset for which we have marginals. This goal can also be reached by the iterative proportional fitting procedure (IPFP) [5]. The IPFP solution is obtained by an iterative procedure of high computational complexity, where at each step I-projections of multidimensional probability distributions are computed. To substantially decrease the computational complexity of this process, Perez proposed an approximate solution [23] based on his idea of dependence structure simplifications. The approximation consists in the fact that not all marginals from the given set are taken into consideration.

Another popular method for representing complex models from sets of marginal distributions and a dependence structure is the method based on Sklar's copulas [36]. But while it is computationally difficult to apply copulas to problems of more than 10 variables, IPFP (especially when using its decomposable representation) can be applied to problems of several tens of variables. Perez's approximation and compositional models can be applied to problems with hundreds of variables.

The goal of this paper is to describe compositional models in the general framework of VBS. The composition operator, which is the central operator of compositional models, was first introduced in probability theory to compare Csiszár's I-projections [4] and Perez's dependence structure simplifications [23], and to make it easier to understand the differences between these two concepts. Soon after, the composition operator was used to introduce compositional models, as an alternative to Bayesian networks, in the framework of discrete probability theory [10, 11]. These models were later extended in [40] for possibility theory, and in [16] for belief functions in the D-S belief function theory.

In this paper, we use the VBS framework [34] to extend compositional models to all uncertainty calculi captured by the VBS framework, which includes calculi such as probability theory, a version of possibility theory with the product t -norm, Spohn's epistemic belief theory, and D-S belief function theory. We define a composition operator for valuations, and notice that conditional valuations can be described using the composition operator. Next, we define a class of compositional models called decomposable models, and for this class of models, we describe how conditioning can be done using local computation.

As the VBS framework includes the D-S theory of belief function, we have implicitly defined a compositional model for the D-S theory. We compare this compositional model with the one defined in [16] for belief functions. The two models are different. The compositional model described in [16] can be described in the VBS framework, but with a combination operator that is different from Dempster's rule of combination. Thus, the compositional model described in [16] is not for the D-S belief function theory that necessarily entails Dempster's rule of combination, but for an alternative belief function theory with the new rule of combination.

For the D-S belief function theory, if we remove a basic probability assignment (BPA) from another BPA, the resulting function may not be a BPA as the probability masses can be negative. This is true even if the BPA being removed is a marginal of the BPA it is being removed from. In this paper, we define a class of belief function models, called graphical belief, such that if we remove a BPA from another, the result is always a BPA.

We compare the VBS compositional model with the one described in [40] for possibility theory. The VBS framework captures only the version of possibility theory where the combination rule is the product t -norm. For this version of possibility theory, the two compositional models coincide. For the other versions of possibility theory, the combination rules (non-product t -norms) do not satisfy the axioms that the VBS operators are required to satisfy. Thus the applications of the local computation algorithms, such as the Shenoy-Shafer architecture [35] or the Lauritzen-Spiegelhalter architecture [20], are not theoretically supported by the results presented in this paper for the versions of possibility theory that do not use the product t -norm.

This paper is a substantially extended version of [14]. The extension consists of inclusion of results from [15] (the results concerning decomposable models, and conditioning using local computation), and some results from an unpublished working paper [13], which makes this paper an up-to-date account of most of the results on compositional models in VBS.

An outline of the remainder of this article is as follows. In Section 2, we briefly sketch the basic definitions and results of the VBS framework. In Section 3, we describe a special case of valuations called conditionals, and define conditional independence in the VBS framework. Conditional independence is then characterized in terms of conditionals. Compositional models in the VBS framework are introduced in Section 4, and their special case called *decomposable* compositional models are described in Section 5. The latter topic is motivated by decomposable graphical models in probability theory that allows for computation of posterior marginals using local computation. In Section 6, we describe conditioning in decomposable compositional models using the compositional operator. Section 7 describes how the D-S belief function theory fits in the VBS framework. Thus, the compositional model described in Section 4 applies also to D-S belief function theory. However, this compositional model for the D-S belief function theory is different from the compositional model defined in [16] for belief functions. The comparison of these two compositional models for belief functions is performed in Section 8. We show that the compositional model described in [16] for belief functions fits in the VBS framework, but with a combination operator that is different from Dempster’s rule of combination. In Section 9, we describe how the possibility theory fits in the VBS framework. In Section 10, we discuss how the compositional model for possibility compares with the compositional model described in [40]. Finally, in Section 11, we summarize our contributions and conclude with a discussion of open issues that remain to be done.

2. Valuation-Based Systems

In this section, we briefly sketch the basic definitions and results of the VBS framework. Most of this material is taken from [34].

VBS consists of two parts—a static part that is concerned with representation of knowledge, and a dynamic part that is concerned with reasoning.

Variables and Valuations. The static part consists of objects called variables and valuations. Let Φ denote a set whose elements are called *variables*. Elements of Φ are denoted by upper-case Roman alphabets such as X, Y, Z , etc. Subsets of Φ are denoted by lower-case Roman alphabets such as r, s, t , etc.

Let Ψ denote a set whose elements are called *valuations*. Elements of Ψ are denoted by lower-case Greek alphabets such as ρ, σ, τ , etc. Each valuation is associated with a subset of variables, and represents some knowledge about the variables in the subset. Thus, we say that ρ is a valuation for r , where $r \subseteq \Phi$ is the subset associated with ρ .

We identify a subset of valuations $\Psi_n \subset \Psi$, whose elements are called *normal* valuations. Normal valuations are valuations that are coherent in some sense. E.g., in D-S belief function theory, normal valuations are basic probability assignment potentials whose values for all non-empty subsets add to one.

Valuation-Based Systems. A VBS is a couple, a finite set of variables Φ_S and a finite set of valuations Ψ_S , which are consistent in the sense that for each $X \in \Phi_S$ there exists (at least one) valuation $\rho \in \Psi_S$ for r such that $X \in r$ and that each valuation $\rho \in \Psi_S$ must be for variables $r \subseteq \Phi_S$. Such a VBS can be graphically depicted by graphs called valuation networks. A *valuation network* (VN) is a bi-partite graph with variables and valuations as nodes, and there

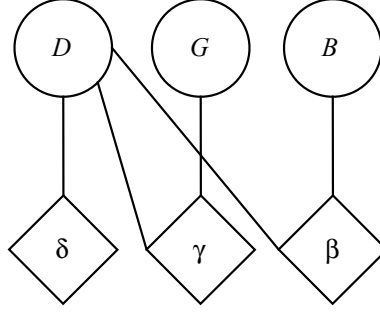


Figure 1: A valuation network

is an edge between each valuation and the variables in the subset associated with it. An example is shown in Figure 1. In this example, $\Phi_S = \{D, G, B\}$, $\Psi_S = \{\delta, \gamma, \beta\}$, where δ is a valuation for $\{D\}$, γ is a valuation for $\{D, G\}$, and β is a valuation for $\{D, B\}$.

The dynamic part of VBS consists of three operators—combination, marginalization, and removal—that are used to make inferences from the knowledge encoded in a VBS. We define these operators using axioms.

Marginalization. The simplest operator is marginalization $-X: \Psi \rightarrow \Psi$, which allows us to coarsen knowledge by marginalizing X out of the domain of a valuation. It must satisfy the following four axioms:

1. (*Domain*) If ρ is a valuation for r , and $X \in r$, then ρ^{-X} is a valuation for $r \setminus \{X\}$.
2. (*Normal*) ρ^{-X} is normal if and only if ρ is normal.
3. (*Order does not matter*) If ρ is a valuation for r , $X \in r$, and $Y \in r$, then $(\rho^{-X})^{-Y} = (\rho^{-Y})^{-X}$, which is denoted by $\rho^{-\{X,Y\}}$.
4. (*Local computation*) If ρ and σ are valuations for r and s , respectively, $X \in r$, and $X \notin s$, then $(\rho \oplus \sigma)^{-X} = (\rho^{-X}) \oplus \sigma$.

The domain axiom is self-explanatory. Marginalization preserves normal (and non-normal) property of valuations. The order does not matter axiom dictates that when we coarsen knowledge by marginalizing out several variables, the order in which the variables are marginalized does not matter in the final result. Occasionally, we let $\rho^{\downarrow r \setminus \{X,Y\}}$ denote $\rho^{-\{X,Y\}}$.

Combination. A substantial operator is the *combination* operator $\oplus: \Psi \times \Psi \rightarrow \Psi_n$, which represents aggregation of knowledge. It must satisfy the following three axioms:

1. (*Domain*) If ρ is a valuation for r , and σ is a valuation for s , then $\rho \oplus \sigma$ is a normal valuation for $r \cup s$.
2. (*Commutativity*) $\rho \oplus \sigma = \sigma \oplus \rho$.
3. (*Associativity*) $\rho \oplus (\sigma \oplus \tau) = (\rho \oplus \sigma) \oplus \tau$.

The domain axiom expresses the fact that if ρ represents some knowledge about variables in r , and σ represents some knowledge about variables in s , then $\rho \oplus \sigma$ represents the aggregated knowledge about variables in $r \cup s$. The commutativity and associativity axioms reflect the fact that the sequence in which knowledge is aggregated makes no difference in the aggregated result.

The set of all normal valuations with the combination operator \oplus forms a commutative semigroup. We let ι_0 denote the (unique) identity valuation of this semigroup. Thus, for any normal valuation ρ , $\rho \oplus \iota_0 = \rho$.

The set of all normal valuations for $s \subseteq \Phi$ with the combination operator \oplus also forms a commutative semigroup (which is different from the semigroup discussed in the previous paragraph). Let ι_s denote the (unique) identity for this semigroup. Thus, for any normal valuation σ for s , $\sigma \oplus \iota_s = \sigma$.

Notice that, in general, $\rho \oplus \rho \neq \rho$. Thus, it is important to ensure that we do not double count knowledge when double counting matters, i.e., it is okay to double count knowledge ρ that is idempotent, i.e., $\rho \oplus \rho = \rho$. In representing our knowledge as valuations in Ψ , we have to ensure that there is no double counting of non-idempotent knowledge. This can be ensured (for example when defining the operator of composition in Section 4) by the removal operator that is defined next.

Removal. The removal operator $\ominus: \Psi \times \Psi_n \rightarrow \Psi_n$ represents removing knowledge in the second valuation from the knowledge in the first valuation. It must satisfy the following three axioms:

1. (*Domain*): Suppose σ is a valuation for s and ρ is a normal valuation for r . Then $\sigma \ominus \rho$ is a normal valuation for $r \cup s$.
2. (*Identity*): For each normal valuation ρ for r , $\rho \oplus \rho \ominus \rho = \rho$. Thus, $\rho \ominus \rho$ acts as an identity for ρ , and we denote $\rho \ominus \rho$ by ι_ρ . Thus, $\rho \oplus \iota_\rho = \rho$.
3. (*Combination and Removal*): Suppose π and θ are valuations, and suppose ρ is a normal valuation. Then, $(\pi \oplus \theta) \ominus \rho = \pi \oplus (\theta \ominus \rho)$.

We call $\sigma \ominus \rho$ the valuation resulting after removing ρ from σ . The identity axiom defines the removal operator as an inverse of the combination operator. As we will see in Section 3, the removal operator is useful in defining conditionals, which are useful in turn in characterizing conditional independence. More importantly, the removal operator is necessary for defining the composition operator.

In [34], a number of properties of combination, marginalization, and removal operators are stated and proved. Here, we describe those that we will need in later sections.

Proposition 1. *Suppose σ and θ are valuations for s and t , respectively, and ρ is a normal valuation for r , and suppose $X \in s$ and $X \notin r$. Then,*

1. $(\sigma \oplus \theta) \ominus \rho = (\sigma \ominus \rho) \oplus \theta$.
2. $(\sigma \ominus \rho)^{-X} = \sigma^{-X} \ominus \rho$.

Domination. As defined in the identity property, $\rho \oplus \iota_\rho = \rho$. In general, if ρ' is a normal valuation for r that is distinct from ρ , then $\rho' \oplus \iota_\rho$ may not equal ρ' . However, there may exist a class of normal valuations for r such that if ρ' is in this class, then $\rho' \oplus \iota_\rho = \rho'$. Following the terminology in [10], we will call this class of normal valuations as valuations that are *dominated* by ρ . Thus, if ρ dominates ρ' , written as $\rho \gg \rho'$, then $\rho' \oplus \iota_\rho = \rho'$.

It may be useful to illustrate this concept with an example from probability theory. Here, a normal valuation ρ for r is a probability distribution for variables r . Combination is pointwise multiplication followed by normalization, and analogously, removal is a pointwise division followed by normalization. Therefore, $\iota_\rho = \rho \ominus \rho$ is a distribution

$$\iota_\rho(\mathbf{a}) = \begin{cases} K^{-1} & \text{if } \rho(\mathbf{a}) > 0, \\ 0 & \text{otherwise,} \end{cases} \quad \text{for all } \mathbf{a} \subseteq \Omega_r.$$

(Here, K is the number of states in Ω_r for which ρ is positive.) For a strictly positive distribution ρ , ι_ρ is a uniform distribution on Ω_r . In general, if ρ and ρ' are normal probability potentials for r such that $\rho(\mathbf{x}) = 0 \Rightarrow \rho'(\mathbf{x}) = 0$, then $\rho \gg \rho'$.

3. Conditionals and Conditional Independence in VBS

In this section, first we define conditionals. Next, we define conditional independence for variables, which does not use conditionals. However, conditionals can be used to characterize the definition of conditional independence. These characterizations of conditional independence correspond to the frequently used definitions of conditional independence in probability theory.

Suppose τ is a normal valuation for t , and suppose r and s are disjoint subsets of t . We call $\tau^{\downarrow(r \cup s)} \ominus \tau^{\downarrow r}$ the *conditional for s given r with respect to τ* . To simplify notation, we will let $\tau(s | r)$ denote $\tau^{\downarrow(r \cup s)} \ominus \tau^{\downarrow r}$. Also, if $r = \emptyset$, let $\tau(s)$ denote $\tau(s | \emptyset)$.

The following proposition is taken from [34].

Proposition 2. *Suppose τ is a normal valuation for t , and suppose r , s , and u are disjoint subsets of t . Then the following statements hold.*

1. $\tau(s) = \tau^{\downarrow s}$.
2. $\tau(r) \oplus \tau(s | r) = \tau(r \cup s)$.

3. $\tau(s \mid r) \oplus \tau(u \mid r \cup s) = \tau(s \cup u \mid r)$.
4. Suppose $X \in s$. Then, $\tau(s \mid r)^{-X} = \tau(s \setminus \{X\} \mid r)$.
5. $\tau(r) \oplus (\tau(s \mid r)^{-s}) = \tau(r)$.
6. $\tau(s \mid r)$ is a normal valuation for $r \cup s$.

In Section 4, we will present an alternative way how to express conditionals. For this, the following assertion will be useful.

Theorem 1. *Suppose τ is a normal valuation for t , and suppose $r \subseteq t$. Then*

$$\tau \oplus \iota_{\tau(r)} = \tau.$$

Proof: In the following equations, we use only the associativity and commutativity properties of combination, and Property 2 of Proposition 2.

$$\tau \oplus \iota_{\tau(r)} = (\tau(r) \oplus \tau(t \setminus r \mid r)) \oplus \iota_{\tau(r)} = (\tau(t \setminus r \mid r) \oplus \tau(r)) \oplus \iota_{\tau(r)} = \tau(t \setminus r \mid r) \oplus (\tau(r) \oplus \iota_{\tau(r)}) = \tau(t \setminus r \mid r) \oplus \tau(r) = \tau.$$

□

Conditional Independence for Variables. Suppose τ is a normal valuation for t , and suppose r , s , and v are disjoint subsets of t . We say r is *conditionally independent of s given v with respect to τ* , written as $r \perp_{\tau} s \mid v$, if $\tau \downarrow^{(r \cup s \cup v)}$ factorizes into valuations α for $r \cup v$ and β for $s \cup v$, i.e., if there exist valuations α for $r \cup v$ and β for $s \cup v$ such that $\tau \downarrow^{r \cup s \cup v} = \alpha \oplus \beta$.

Some observations. First, while τ has to be necessarily normal, valuations α and β do not have to be normal. Second, the definition of conditional independence does not involve the removal operator, only the combination and marginalization operators. However, we can characterize conditional independence in terms of conditionals, which are defined using the removal operator. This is done in Proposition 3 below. In [17], conditional independence is characterized using Shafer's [27] "continuers," which correspond to what we call conditionals, but which can be defined without explicit reference to a removal operator. Third, if $s = \emptyset$, then $r \perp_{\tau} \emptyset \mid v$ since we can let $\alpha = \tau \downarrow^{(r \cup v)}$ and $\beta = \iota_v$. This property is called *trivial independence* by Geiger and Pearl [8]. Fourth, if $v = \emptyset$, then we say r and s are *independent* with respect to τ , written as $r \perp_{\tau} s$, if $\tau \downarrow^{(r \cup s)} = \alpha \oplus \beta$, where α is a valuation for r and β is a valuation for s . Thus, independence is a special case of conditional independence.

The following result is proved in [34].

Proposition 3. *Suppose τ is a normal valuation for t , and suppose r , s , and v are disjoint subsets of t . The following statements are equivalent.*

1. $r \perp_{\tau} s \mid v$.
2. $\tau(r \cup s \cup v) = \tau(v) \oplus \tau(r \mid v) \oplus \tau(s \mid v)$.
3. $\tau(r \cup s \mid v) = \tau(r \mid v) \oplus \tau(s \mid v)$.
4. $\tau(r \cup s \cup v) \oplus \tau(v) = \tau(r \cup v) \oplus \tau(s \cup v)$.
5. $\tau(r \cup s \cup v) = \tau(r \mid v) \oplus \tau(s \cup v)$.
6. $\tau(r \mid s \cup v) = \tau(r \mid v) \oplus \iota_{\tau(s \cup v)}$.
7. $\tau(r \mid s \cup v) = \alpha \oplus \iota_{\tau(s \cup v)}$, where α is a valuation for $r \cup v$.

In [34], it is shown that the definition of conditional independence satisfies the so-called semi-graphoid axioms [22, 39], and also the graphoid axioms in the case where τ belongs to a sub-class of normal valuations called "positive normal."

4. Compositional Models in VBS

Suppose we have marginals for two overlapping subsets of variables, say for $\{D, G\}$ and $\{D, B\}$. How do we construct a joint distribution for $\{D, G, B\}$ that is consistent with the two marginals (assuming that it exists)? In [10], the operation of "composing" the two marginals to obtain a joint distribution is introduced. One way to view the composition

operator is in terms of no double counting. Notice that the two marginals are not distinct since the knowledge of $\{D\}$ is included in both marginals. So, the composition operator should aggregate the knowledge in the two marginals while adjusting for the double counting of knowledge of $\{D\}$.

In practice, it is extremely unlikely we would find marginals on non-disjoint subsets of variables with common marginals, i.e., for each pair of marginals, the marginals of the two marginals for the intersection of the subsets agree. In this case, there does not exist a joint that agrees with both marginals. So we relax the requirements so that the joint distribution that is constructed is required to agree only with the first marginal.

Composition. A general definition of composition is as follows. Suppose ρ and σ are normal valuations for r and s , respectively. The composition of ρ and σ , written as $\rho \triangleright \sigma$, is defined as follows:

$$\rho \triangleright \sigma = \rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s}. \quad (1)$$

It can be seen directly from the definition in Equation (1) that the composition operator is, in general, neither commutative nor associative. Its most important properties are summarized in the following assertion.

Theorem 2. *Suppose ρ and σ are normal valuations for r and s , respectively, and suppose that $\sigma^{\downarrow r \cap s} \gg \rho^{\downarrow r \cap s}$. Then the following statement hold.*

1. Domain: $\rho \triangleright \sigma$ is a normal valuation for $r \cup s$.
2. Composition preserves first marginal: $(\rho \triangleright \sigma)^{\downarrow r} = \rho$.
3. Non-commutativity: In general, $\rho \triangleright \sigma \neq \sigma \triangleright \rho$.
4. Commutativity under consistency: If ρ and σ have a common marginal for $r \cap s$, i.e., $\rho^{\downarrow r \cap s} = \sigma^{\downarrow r \cap s}$, then $\rho \triangleright \sigma = \sigma \triangleright \rho$.
5. Non-associativity: Suppose τ is a normal valuation for t , and suppose $\tau^{\downarrow (r \cup s) \cap t} \gg (\rho \triangleright \sigma)^{\downarrow (r \cup s) \cap t}$. Then, in general,

$$(\rho \triangleright \sigma) \triangleright \tau \neq \rho \triangleright (\sigma \triangleright \tau).$$

6. Associativity under a special condition: Suppose τ is a normal valuation for t , suppose $\tau^{\downarrow (r \cup s) \cap t} \gg (\rho \triangleright \sigma)^{\downarrow (r \cup s) \cap t}$, and suppose $s \supset (r \cap t)$. Then,

$$(\rho \triangleright \sigma) \triangleright \tau = \rho \triangleright (\sigma \triangleright \tau).$$

7. Stepwise composition: Suppose t is such that $(r \cap s) \subseteq t \subseteq s$. Then

$$(\rho \triangleright \sigma^{\downarrow t}) \triangleright \sigma = \rho \triangleright \sigma.$$

8. Exchangeability: Suppose τ is a normal valuation for t , suppose $\tau^{\downarrow r \cap t} \gg \rho^{\downarrow r \cap t}$, and suppose $r \supset (s \cap t)$. Then,

$$(\rho \triangleright \sigma) \triangleright \tau = (\rho \triangleright \tau) \triangleright \sigma.$$

Proof:

1. *Domain:* This follows directly from the definition of the removal operator.
2. *Composition preserves first marginal:* In the following computations we use local computations of marginalization, associativity of combination, and combination and removal from the definition of removal

$$\begin{aligned} (\rho \triangleright \sigma)^{\downarrow r} &= (\rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s})^{\downarrow r} = (\sigma \oplus \rho \ominus \sigma^{\downarrow r \cap s})^{\downarrow r} = (\sigma \oplus (\rho \ominus \sigma^{\downarrow r \cap s}))^{\downarrow r} \\ &= \sigma^{\downarrow r \cap s} \oplus (\rho \ominus \sigma^{\downarrow r \cap s}) = \rho \oplus \sigma^{\downarrow r \cap s} \ominus \sigma^{\downarrow r \cap s} = \rho \oplus \iota_{\sigma^{\downarrow r \cap s}} = \rho. \end{aligned}$$

Notice that the last equality holds true because of the assumption that $\sigma^{\downarrow r \cap s} \gg \rho^{\downarrow r \cap s}$.

3. *Non-commutativity:* Suppose ρ and σ are such that $\rho^{\downarrow r \cap s} \neq \sigma^{\downarrow r \cap s}$. Then, it follows from the composition preserves first marginal property that $(\rho \triangleright \sigma)^{\downarrow r \cap s} = \rho^{\downarrow r \cap s}$ and $(\sigma \triangleright \rho)^{\downarrow r \cap s} = \sigma^{\downarrow r \cap s}$. But since $\rho^{\downarrow r \cap s} \neq \sigma^{\downarrow r \cap s}$, the result follows.
4. *Commutativity under consistency:* For consistent valuations

$$\rho \triangleright \sigma = \rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s} = \rho \oplus \sigma \ominus \rho^{\downarrow r \cap s} = \sigma \oplus \rho \ominus \rho^{\downarrow r \cap s} = \sigma \triangleright \rho.$$

5. *Non-associativity*: Consider disjoint $t, u \subset r$ and a valuation ρ for which $\rho^{\downarrow r \cup u} \neq \rho^{\downarrow t} \oplus \rho^{\downarrow u}$. Then

$$(\rho^{\downarrow t} \triangleright \rho^{\downarrow u}) \triangleright \rho^{\downarrow r \cup u} = (\rho^{\downarrow t} \oplus \rho^{\downarrow u} \ominus \rho(\emptyset)) \triangleright \rho^{\downarrow r \cup u} = (\rho^{\downarrow t} \oplus \rho^{\downarrow u}) \triangleright \rho^{\downarrow r \cup u} = \rho^{\downarrow t} \oplus \rho^{\downarrow u},$$

(where the last equality holds because composition preserves first marginal), whereas

$$\rho^{\downarrow t} \triangleright (\rho^{\downarrow u} \triangleright \rho^{\downarrow r \cup u}) = \rho^{\downarrow t} \triangleright (\rho^{\downarrow r \cup u} \triangleright \rho^{\downarrow u}) = \rho^{\downarrow t} \triangleright \rho^{\downarrow r \cup u} = \rho^{\downarrow r \cup u} \triangleright \rho^{\downarrow t} = \rho^{\downarrow r \cup u}.$$

6. *Associativity under a special condition*: Note that in the considered case $t \cap (r \cup s) = t \cap s$ and $r \cap (t \cup s) = r \cap s$. By definition,

$$(\rho \triangleright \sigma) \triangleright \tau = (\rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s}) \oplus \tau \ominus \tau^{\downarrow r \cap (r \cup s)} = (\rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s}) \oplus \tau \ominus \tau^{\downarrow r \cap t} = \rho \oplus \sigma \oplus \tau \ominus \sigma^{\downarrow r \cap s} \ominus \tau^{\downarrow r \cap t},$$

and

$$\rho \triangleright (\sigma \triangleright \tau) = \rho \oplus (\sigma \triangleright \tau) \ominus (\sigma \triangleright \tau)^{\downarrow r \cap (r \cup s)} = \rho \oplus (\sigma \triangleright \tau) \ominus \sigma^{\downarrow r \cap s} = \rho \oplus (\sigma \oplus \tau \ominus \tau^{\downarrow s \cap t}) \ominus \sigma^{\downarrow r \cap s}$$

(where the second equality holds because composition preserves first marginal). Since both sides are equal, the result follows.

7. *Stepwise composition*:

$$\begin{aligned} (\rho \triangleright \sigma^{\downarrow t}) \triangleright \sigma &= (\rho \oplus \sigma^{\downarrow t} \ominus \sigma^{\downarrow r \cap t}) \triangleright \sigma = (\rho \oplus \sigma^{\downarrow t} \ominus \sigma^{\downarrow r \cap s}) \oplus \sigma \ominus \sigma^{\downarrow (r \cup t) \cap s} \\ &= \rho \oplus \sigma^{\downarrow t} \ominus \sigma^{\downarrow r \cap s} \oplus \sigma \ominus \sigma^{\downarrow t} = \rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s} = \rho \triangleright \sigma, \end{aligned}$$

where we used in addition to associativity of combination the fact that under the given assumption $r \cap s = r \cap t$ and $(r \cup t) \cap s = t$.

8. *Exchangeability*: Under the given assumption $(r \cup s) \cap t = r \cap t$ and $(r \cup t) \cap s = r \cap s$, and therefore

$$(\rho \triangleright \sigma) \triangleright \tau = \rho \oplus \sigma \ominus \sigma^{\downarrow (r \cap s)} \oplus \tau \ominus \tau^{\downarrow (r \cup s) \cap t} = \rho \oplus \tau \ominus \tau^{\downarrow (r \cap t)} \oplus \sigma \ominus \sigma^{\downarrow (r \cup t) \cap s} = (\rho \triangleright \tau) \triangleright \sigma.$$

□

Using Properties 2, 7, and 8 of the above Theorem we can prove an important assertion that enable us, under a special condition, to compute marginals of composed valuations.

Theorem 3. *Suppose ρ and σ are normal valuations for r and s , respectively, and suppose that $\sigma^{\downarrow r \cap s} \gg \rho^{\downarrow r \cap s}$. If t is such that $(r \cap s) \subseteq t \subseteq r \cup s$. Then*

$$(\rho \triangleright \sigma)^{\downarrow t} = \rho^{\downarrow r \cap t} \triangleright \sigma^{\downarrow s \cap t}.$$

Proof: First, let us compute $(\rho \triangleright \sigma)^{\downarrow r \cup t}$ using Properties 7, 2, and 8 of Theorem 2.

$$(\rho \triangleright \sigma)^{\downarrow r \cup t} = ((\rho \triangleright \sigma^{\downarrow s \cap t}) \triangleright \sigma)^{\downarrow r \cup t} = \rho \triangleright \sigma^{\downarrow s \cap t} = (\rho^{\downarrow r \cap s} \triangleright \rho) \triangleright \sigma^{\downarrow s \cap t} = (\rho^{\downarrow r \cap s} \triangleright \sigma^{\downarrow s \cap t}) \triangleright \rho.$$

Now, let us compute $(\rho \triangleright \sigma)^{\downarrow t} = ((\rho \triangleright \sigma)^{\downarrow r \cup t})^{\downarrow t}$ using again Properties 7, 2, and 8 of Theorem 2.

$$\begin{aligned} (\rho \triangleright \sigma)^{\downarrow t} &= ((\rho^{\downarrow r \cap s} \triangleright \sigma^{\downarrow s \cap t}) \triangleright \rho)^{\downarrow t} = (((\rho^{\downarrow r \cap s} \triangleright \sigma^{\downarrow s \cap t}) \triangleright \rho^{\downarrow r \cap t}) \triangleright \rho)^{\downarrow t} = (\rho^{\downarrow r \cap s} \triangleright \sigma^{\downarrow s \cap t}) \triangleright \rho^{\downarrow r \cap t} \\ &= (\rho^{\downarrow r \cap s} \triangleright \rho^{\downarrow r \cap t}) \triangleright \sigma^{\downarrow s \cap t} = \rho^{\downarrow r \cap t} \triangleright \sigma^{\downarrow s \cap t}. \end{aligned}$$

□

As we mentioned in Section 3, there is an alternative way of expressing conditionals using the composition operator. Though, as it can be seen from the proof, it is just a formal reformulation of the definition, a possibility to express conditionals as a composition of two valuations enables us to design the respective computational procedures taking advantage of the properties of this operator – especially the associativity under a special condition.

Theorem 4. *Suppose τ is a normal valuation for t , and suppose r and s are nonempty disjoint subsets of t such that $r \cup s = t$. Then*

$$\tau(s \mid r) = \iota_{\tau(r)} \triangleright \tau.$$

Proof:

$$\iota_{\tau(r)} \triangleright \tau = \iota_{\tau(r)} \oplus \tau \ominus \tau^{\downarrow r} = \tau \ominus \tau^{\downarrow r} = \tau(s \mid r).$$

□

5. Decomposable Compositional Models in VBS

In probability theory, inference with Bayesian networks is usually based on the idea of *local computation* of Lauritzen and Spiegelhalter [21]. This idea, which was in a restricted way for belief functions used already in [12], can be briefly expressed as follows. A Bayesian network is first transformed into a decomposable model (using well-known operations *moralization* and *triangulation* of a directed graph), and the required posterior marginal is then computed by a process exploiting the “tree” structure of decomposable models. Therefore, it is not surprising that we speak about decomposable compositional models in the VBS framework.

The tree structure of decomposable models is expressed as a running intersection property. We say that a sequence of sets s_1, s_2, \dots, s_n meets *running intersection property* (RIP) if for each $j = 2, 3, \dots, n$ there exists a $k < j$ such that

$$s_j \cap (s_1 \cup \dots \cup s_{j-1}) = s_j \cap s_k.$$

Decomposable compositional models are formed by multiple applications of the composition operator. Since it is not always associative (Property 5 of Theorem 2), we use the following convention. If we do not specify an order using brackets, the operators will always be performed from left to right, i.e., $\tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \tau^{\downarrow s_3} \triangleright \dots \triangleright \tau^{\downarrow s_n}$ denotes $(\dots((\tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2}) \triangleright \tau^{\downarrow s_3}) \triangleright \dots \triangleright \tau^{\downarrow s_n})$.

Decomposable VBS. Suppose τ is a normal valuation for t . We say τ is *decomposable* if there exists a sequence (s_1, s_2, \dots, s_n) of subsets of t such that it meets RIP and

$$\tau = \tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_n}.$$

In this case we also say that τ is decomposable with respect to the sequence (s_1, s_2, \dots, s_n) .

It is well-known that if a sequence (s_1, s_2, \dots, s_n) meets RIP, then we can find another sequence starting with, say, s_j that also meets RIP. More precisely, for each $j = 1, 2, \dots, n$ there exists (at least one) permutation $(s_{\ell_1}, s_{\ell_2}, \dots, s_{\ell_n})$ that meets RIP and such that $s_{\ell_j} = s_j$. Therefore, the following assertion is of great importance.

Theorem 5. *If τ is decomposable with respect to (s_1, s_2, \dots, s_n) , and $(s_{j_1}, s_{j_2}, \dots, s_{j_n})$ is a permutation of (s_1, s_2, \dots, s_n) such that it meets RIP, then τ is decomposable with respect to $(s_{j_1}, s_{j_2}, \dots, s_{j_n})$, i.e.,*

$$\tau = \tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_n} = \tau^{\downarrow s_{j_1}} \triangleright \tau^{\downarrow s_{j_2}} \triangleright \dots \triangleright \tau^{\downarrow s_{j_n}}.$$

Proof: The proof of this assertion is based on an important result concerning decomposable graphs, which follows from the results of Haberman ([9], Lemma 2.8) saying that the system of subsets (more exactly, *multiset*)

$$\{s_2 \cap s_1, s_3 \cap (s_1 \cup s_2), s_4 \cap (s_1 \cup s_2 \cup s_3), \dots, s_n \cap (s_1 \cup \dots \cup s_{n-1})\}$$

does not depend on the selected RIP ordering of the sequence (s_1, s_2, \dots, s_n) . Taking into account the running intersection property, we know that each element of this multiset is an intersection of two sets from the sequence (s_1, s_2, \dots, s_n) . Therefore, the above mentioned property can be expressed as follows: For any pair of distinct sets s_i, s_j from a system $\{s_1, s_2, \dots, s_n\}$, which can be ordered to meet RIP, the number of times the set $s_i \cap s_j$ appears in the sequence

$$s_{j_2} \cap s_{j_1}, s_{j_3} \cap (s_{j_1} \cup s_{j_2}), s_{j_4} \cap (s_{j_1} \cup s_{j_2} \cup s_{j_3}), \dots, s_{j_n} \cap (s_{j_1} \cup \dots \cup s_{j_{n-1}})$$

does not depend on the RIP ordering $(s_{j_1}, s_{j_2}, \dots, s_{j_n})$.

Suppose τ is decomposable with respect to (s_1, s_2, \dots, s_n) . Using the definition of composition, we have:

$$\tau = \tau^{\downarrow s_1} \oplus (\tau^{\downarrow s_2} \ominus \tau^{\downarrow s_2 \cap s_1}) \oplus (\tau^{\downarrow s_3} \ominus \tau^{\downarrow s_3 \cap (s_1 \cup s_2)}) \oplus \dots \oplus (\tau^{\downarrow s_n} \ominus \tau^{\downarrow s_n \cap (s_1 \cup \dots \cup s_{n-1})})$$

which can be reorganized independently of the RIP ordering (using the properties of combination and removal and Proposition 1) as follows:

$$\tau = (\tau^{\downarrow s_1} \oplus \tau^{\downarrow s_2} \oplus \dots \oplus \tau^{\downarrow s_n}) \ominus \tau^{\downarrow s_2 \cap s_1} \ominus \tau^{\downarrow s_3 \cap (s_1 \cup s_2)} \ominus \dots \ominus \tau^{\downarrow s_n \cap (s_1 \cup \dots \cup s_{n-1})}$$

□

6. Conditioning in Decomposable Compositional Models

In this section, we assume that τ is a normal valuation for t , and that it is decomposable with respect to (s_1, s_2, \dots, s_n) . Suppose we wish to compute the conditional $\tau(t \setminus \{X\} \mid \{X\})$.

First, we have to find an ordering of s_1, s_2, \dots, s_n such that it meets RIP, and such that the first set from this ordering contains X . We know from Theorem 5 that τ is decomposable also with respect to this new sequence. Therefore, without loss of generality we can assume that it is (s_1, s_2, \dots, s_n) , which means that we assume $X \in s_1$. Thus, using Theorem 4, we compute

$$\tau(t \setminus \{X\} \mid \{X\}) = \iota_{\tau(X)} \triangleright \tau = \iota_{\tau(X)} \triangleright (\tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_n}).$$

However, due to Property 6 (associativity under a special condition) of Theorem 2, we have

$$\iota_{\tau(X)} \triangleright ((\tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_{n-1}}) \triangleright \tau^{\downarrow s_n}) = (\iota_{\tau(X)} \triangleright (\tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_{n-1}})) \triangleright \tau^{\downarrow s_n}, \quad (2)$$

because s_1 , and thus even more $s_1 \cup \dots \cup s_{n-1}$, contains $\{X\} \cap s_n$. Notice that also the other assumption of associativity under a special condition is fulfilled because

$$\tau^{\downarrow (s_1 \cup \dots \cup s_{n-1}) \cap s_n} \gg (\iota_{\tau(X)} \triangleright (\tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_{n-1}}))^{\downarrow (s_1 \cup \dots \cup s_{n-1}) \cap s_n}.$$

Repeating the idea behind equality (2), we get

$$\iota_{\tau(X)} \triangleright ((\tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_{n-2}}) \triangleright \tau^{\downarrow s_{n-1}}) = (\iota_{\tau(X)} \triangleright (\tau^{\downarrow s_1} \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_{n-2}})) \triangleright \tau^{\downarrow s_{n-1}}.$$

Thus, eventually, after repeating this step $(n - 1)$ times we get

$$\tau(t \setminus \{X\} \mid \{X\}) = \iota_{\tau(X)} \triangleright \tau = (\iota_{\tau(X)} \triangleright \tau^{\downarrow s_1}) \triangleright \tau^{\downarrow s_2} \triangleright \dots \triangleright \tau^{\downarrow s_n},$$

from which we see that $\tau(t \setminus \{X\} \mid \{X\})$ is again a decomposable model with respect to (s_1, s_2, \dots, s_n) . Let $\hat{\tau}$ denote $\tau(t \setminus \{X\} \mid \{X\})$. We can compute the marginal valuations of $\hat{\tau}$ (that are necessary to represent this multidimensional valuation as a compositional model) as follows:

$$\begin{aligned} \hat{\tau}^{\downarrow s_1} &= \iota_{\tau(X)} \triangleright \tau^{\downarrow s_1} \\ \hat{\tau}^{\downarrow s_2} &= \hat{\tau}^{\downarrow s_2 \cap s_1} \triangleright \tau^{\downarrow s_2} \\ &\vdots \\ \hat{\tau}^{\downarrow s_n} &= \hat{\tau}^{\downarrow s_n \cap (s_1 \cup \dots \cup s_{n-1})} \triangleright \tau^{\downarrow s_n}. \end{aligned} \quad (3)$$

Notice that this computation is tractable because, thanks to RIP, at each step $\hat{\tau}^{\downarrow s_i \cap (s_1 \cup \dots \cup s_{i-1})}$ is easily computable since $s_i \cap (s_1 \cup \dots \cup s_{i-1})$ must be contained in some s_k for $k < i$. So, it is important to realize that we use decomposability of the valuation τ at two steps of the described process. First, the decomposability assumption enables us to reorder marginals $\tau^{\downarrow s_i}$ in the sequence defining the decomposable model so that s_1 contains the conditioning variable X , and we can then apply the 'associativity under a special condition' property (statement 6 of Theorem 2) of compositional models. Second, the decomposability made the computational process (3) tractable. From this, one can immediately see that conditioning may be for general (i.e., non-decomposable) compositional models a very hard problem.

7. VBS for D-S Belief Function Theory

In D-S belief function theory, we can use basic probability assignments or belief functions or plausibility functions or commonality functions to represent knowledge. Since the combination and removal operations are easily defined using the commonality functions, and marginalization using basic probability assignments, we will represent knowledge using either commonality functions or basic probability assignments. We start with definitions of basic probability assignments and commonality functions.

Basic Probability Assignment. A basic probability assignment (BPA) μ for s is a function $\mu : 2^{\Omega_s} \rightarrow \mathbb{R}$ such that $\mu(\mathbf{a}) \geq 0$ for all $\mathbf{a} \subseteq \Omega_s$, and

$$\sum \{\mu(\mathbf{a}) \mid \mathbf{a} \subseteq \Omega_s\} = 1.$$

Commonality and Pseudo-Commonality Functions. A function $\theta : 2^{\Omega_s} \rightarrow \mathbb{R}$ is a *commonality function* for s if and only if there exists a BPA μ for s such that

$$\theta(\mathbf{a}) = \sum \{\mu(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\} \quad \text{for all } \mathbf{a} \subseteq \Omega_s. \quad (4)$$

It is evident from Equation (4) that $0 \leq \theta(\mathbf{a}) \leq 1$, and that $\theta(\mathbf{a}) \geq \theta(\mathbf{b})$ whenever $\mathbf{a} \subseteq \mathbf{b}$.

The following two propositions from [26] will help us understand the mathematical properties of commonality functions.

Proposition 4. Suppose μ and θ are real-valued functions defined on 2^{Ω_s} . Then Equation (4) holds if and only if

$$\mu(\mathbf{a}) = \sum \{(-1)^{|\mathbf{c} \setminus \mathbf{a}|} \theta(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\} \quad \text{for all } \mathbf{a} \subseteq \Omega_s. \quad (5)$$

Proposition 5. Suppose μ and θ are real-valued functions defined on 2^{Ω_s} , and suppose Equation (5) holds for all $\mathbf{a} \subseteq \Omega_s$. Then

$$\sum \{\mu(\mathbf{a}) \mid \mathbf{a} \subseteq \Omega_s\} = \sum \{(-1)^{|\mathbf{a}|+1} \theta(\mathbf{a}) \mid \mathbf{a} \subseteq \Omega_s\}.$$

From Proposition 4, we see that a BPA is completely determined from a commonality function. From Propositions 4 and 5 we see that a function $\theta : 2^{\Omega_s} \rightarrow \mathbb{R}$ is a commonality function if and only if the following two conditions are satisfied:

$$\mu(\mathbf{a}) = \sum \{(-1)^{|\mathbf{c} \setminus \mathbf{a}|} \theta(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\} \geq 0 \quad \text{for every } \mathbf{a} \subseteq \Omega_s, \quad (6)$$

and

$$\sum \{\mu(\mathbf{a}) \mid \mathbf{a} \subseteq \Omega_s\} = \sum \{(-1)^{|\mathbf{a}|+1} \theta(\mathbf{a}) \mid \mathbf{a} \subseteq \Omega_s\} = 1. \quad (7)$$

If θ satisfies Equation (7) but not Equation (6), then we say θ is a *pseudo-commonality function*.

Commonality Valuations. In D-S belief function theory, a commonality valuation (or *c*-valuation, in short) σ for s is a function $\sigma : 2^{\Omega_s} \rightarrow \mathbb{R}^+$. We say σ is *normal* if $\sum \{(-1)^{|\mathbf{a}|+1} \sigma(\mathbf{a}) \mid \mathbf{a} \subseteq \Omega_s\} = 1$. Thus, normal *c*-valuations correspond to pseudo-commonality or commonality functions. We say σ is *proper* if $\sum \{(-1)^{|\mathbf{c} \setminus \mathbf{a}|} \theta(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\} \geq 0$ for all $\mathbf{a} \subseteq \Omega_s$. Thus, proper normal *c*-valuations correspond to commonality functions.

BPA Valuations. A BPA valuation (or *b*-valuation, in short) σ for s is a function $\sigma : 2^{\Omega_s} \rightarrow \mathbb{R}$. We say σ is *normal* if $\sum \{\sigma(\mathbf{a}) \mid \mathbf{a} \subseteq \Omega_s\} = 1$, and we say σ is *proper* if $\sigma(\mathbf{a}) \geq 0$ for all $\mathbf{a} \subseteq \Omega_s$. Proper normal *b*-valuations represent BPA functions. Normal *b*-valuations that are not proper are called *pseudo-BPA*.

If we are given a *c*-valuation, we can convert it into a corresponding *b*-valuation by using Equation (5), and if we are given a *b*-valuation, we can convert it to a corresponding *c*-valuation by using Equation (4). A *c*-valuation is normal if and only if the corresponding *b*-valuation is normal, and a *c*-valuation is proper if and only if the corresponding *b*-valuation is proper.

In D-S belief function theory, Dempster's rule of combination is pointwise multiplication of commonality functions followed by normalization (assuming it is possible) [26]. Before we can define combination formally, we need to define projection of subsets of states.

Projection of Subsets of States. Suppose r and s are sets of variables, $r \subseteq s$, and $\mathbf{a} \subseteq \Omega_s$. The projection of \mathbf{a} to r , denoted by $\mathbf{a}^{\downarrow r}$, is given by $\mathbf{a}^{\downarrow r} = \{\mathbf{x}^{\downarrow r} \mid \mathbf{x} \in \mathbf{a}\}$. Notice that $\mathbf{a}^{\downarrow r} \subseteq \Omega_r$.

Combination. Suppose ρ and σ are *c*-valuations for r and s , respectively. Let K denote $\sum \{(-1)^{|\mathbf{a}^{\downarrow r}|+1} \rho(\mathbf{a}^{\downarrow r}) \sigma(\mathbf{a}^{\downarrow s}) \mid \mathbf{a} \subseteq \Omega_{r \cup s}\}$. The combination $\rho \oplus \sigma$ is a *c*-valuation for $r \cup s$ given by

$$(\rho \oplus \sigma)(\mathbf{a}) = \begin{cases} K^{-1} \rho(\mathbf{a}^{\downarrow r}) \sigma(\mathbf{a}^{\downarrow s}) & \text{if } K \neq 0 \\ 0 & \text{if } K = 0 \end{cases} \quad \text{for all } \mathbf{a} \subseteq \Omega_{r \cup s}. \quad (8)$$

If $K \neq 0$, then K is the normalization constant that ensures that $\rho \oplus \sigma$ is a normal *c*-valuation. It is shown by Shafer [26] that if ρ and σ are commonality functions (proper normal *c*-valuations), and $K \neq 0$, then $\rho \oplus \sigma$ is a commonality

function. It is shown in ([34]) that the definition of combination in Equation (8) satisfies the three properties of combination.

Marginalization. It is simpler to define marginalization in terms of b -valuations. Suppose σ is a b -valuation for s , and suppose $X \in s$. The marginal σ^{-X} is a b -valuation for $s \setminus \{X\}$ given by

$$\sigma^{-X}(\mathbf{a}) = \sum \{\sigma(\mathbf{b}) \mid \mathbf{b} \subseteq \Omega_s \text{ s.t. } \mathbf{b} \downarrow_{s \setminus \{X\}} = \mathbf{a}\} \quad \text{for all } \mathbf{a} \subseteq \Omega_{s \setminus \{X\}}. \quad (9)$$

It is shown in ([34]) that the definition of marginalization in Equation (9) satisfies the four properties of marginalization.

Removal. Removal is division of c -valuations followed by normalization (assuming it is possible). Suppose σ is a c -valuation for s , and suppose ρ is a normal c -valuation for r . Let K denote $\sum \{(-1)^{|\mathbf{a}|+1} \left(\frac{\sigma(\mathbf{a} \downarrow_s)}{\rho(\mathbf{a} \downarrow_r)}\right) \mid \mathbf{a} \subseteq \Omega_{s \cup r} \text{ s.t. } \rho(\mathbf{a} \downarrow_r) \neq 0\}$. The valuation $\sigma \ominus \rho$ for $s \cup r$ is given by

$$(\sigma \ominus \rho)(\mathbf{a}) = \begin{cases} K^{-1} \left(\frac{\sigma(\mathbf{a} \downarrow_s)}{\rho(\mathbf{a} \downarrow_r)}\right) & \text{if } K \neq 0, \rho(\mathbf{a} \downarrow_r) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } \mathbf{a} \subseteq \Omega_{s \cup r}. \quad (10)$$

If $K \neq 0$, K is a normalization constant that ensures $\sigma \ominus \rho$ is a normal c -valuation. It is shown in ([34]) that the definition of removal in Equation (10) satisfies the three properties of removal.

Notice that if σ and ρ are commonality functions, it is possible that $\sigma \ominus \rho$ is a pseudo-commonality function. This may be true even if $r \subseteq s$ and ρ is a marginal of σ , i.e., $\sigma \ominus \sigma \downarrow_r$ may be a pseudo-commonality function. We demonstrate this by an example due to Studený [1].

Example 1. Suppose X and Y are variables with state spaces $\Omega_X = \{x, \bar{x}\}$ and $\Omega_Y = \{y, \bar{y}\}$. Consider a BPA μ for $\{X, Y\}$ as follows:

$$\begin{aligned} \mu(\{(x, \bar{y}), (\bar{x}, y)\}) &= 0.5, \\ \mu(\{(x, \bar{y}), (\bar{x}, \bar{y})\}) &= 0.5. \end{aligned}$$

The marginal BPA μ^{-X} for $\{Y\}$ is as follows:

$$\begin{aligned} \mu^{-X}(\{\bar{y}\}) &= 0.5, \\ \mu^{-X}(\{y, \bar{y}\}) &= 0.5, \end{aligned}$$

then, $\mu \ominus \mu^{-X}$ is as follows:

$$\begin{aligned} (\mu \ominus \mu^{-X})(\{(x, \bar{y})\}) &= -0.5, \\ (\mu \ominus \mu^{-X})(\{(x, \bar{y}), (\bar{x}, y)\}) &= 1, \\ (\mu \ominus \mu^{-X})(\{(x, \bar{y}), (\bar{x}, \bar{y})\}) &= 0.5. \end{aligned}$$

Thus, $\mu \ominus \mu^{-X}$ is a pseudo-BPA for $\{X, Y\}$.

Domination. Suppose ρ is a normal c -valuation for r . It is clear from the definition of removal in Equation (10) that $\rho \ominus \rho = \iota_\rho$ is a normal c -valuation for r whose values are as follows:

$$\iota_\rho(\mathbf{a}) = (\rho \ominus \rho)(\mathbf{a}) = \begin{cases} K^{-1} & \text{if } \rho(\mathbf{a}) > 0 \\ 0 & \text{if } \rho(\mathbf{a}) = 0 \end{cases} \quad \text{for all } \mathbf{a} \in \Omega_r,$$

where $K = \sum \{(-1)^{|\mathbf{a}|+1} \mid \mathbf{a} \subseteq \Omega_r \text{ s.t. } \rho(\mathbf{a}) \neq 0\}$. Now consider a normal c -valuation ρ' for r such that $\rho(\mathbf{a}) = 0 \Rightarrow \rho'(\mathbf{a}) = 0$. Then it is clear that $\rho' \oplus \iota_\rho = \rho'$. Thus, $\rho \gg \rho'$ if $\rho(\mathbf{a}) = 0 \Rightarrow \rho'(\mathbf{a}) = 0$ for all $\mathbf{a} \in \Omega_r$.

Graphical Belief Model. In Example 1, we saw that using the removal operator with b -valuations can result in a pseudo-BPA. This does not happen if the valuation being removed is already included in the valuation that it is

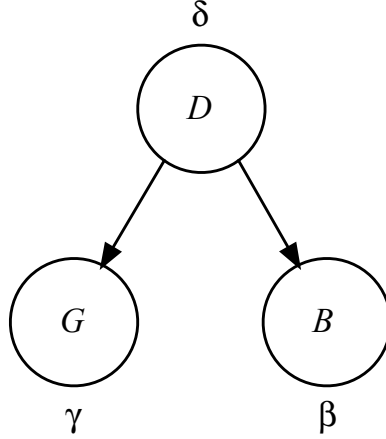


Figure 2: A graphical belief model for $\Phi = \{D, G, B\}$

removed from, i.e., $(\sigma \oplus \rho) \ominus \sigma$ is always a b -valuation assuming σ and ρ are b -valuations. Now, we describe a class of D-S belief function models for which removal of some valuations always results in BPA (or commonality) functions.

Suppose $G = (\mathbb{V}, \mathbb{E})$ is a directed acyclic graph, where \mathbb{V} is the set of the considered nodes (variables), and $\mathbb{E} \subset \mathbb{V} \times \mathbb{V}$ is a set of directed edges. For each $X \in \mathbb{V}$, let $Pa(X)$ denote the set of *parents* of variable X , i.e., $Pa(X) = \{Y \in \mathbb{V} \mid (Y, X) \in \mathbb{E}\}$. Let Ψ_G denote a set of valuations. We say Ψ_G is a *graphical belief model* corresponding to G if

$$\Psi_G = \{\alpha_X \mid X \in \mathbb{V}, \alpha_X \text{ is a proper valuation for } \{X\} \cup Pa(X) \text{ such that } \alpha_X^{-X} = \iota_{Pa(X)}\}.$$

Example 2. An example of a graphical belief model is as follows (see Figure 2).

$$\begin{aligned} \mathbb{V} &= \{D, G, B\}, \\ \mathbb{E} &= \{(D, G), (D, B)\}, \\ G &= (\Phi, \mathbb{E}), \\ \Psi_G &= \{\delta, \gamma, \beta\}, \end{aligned}$$

where δ is a BPA for $\{D\}$, γ is a BPA for $\{D, G\}$ such that $\gamma^{-G} = \iota_D$, and β is a BPA for $\{D, B\}$ such that $\beta^{-B} = \iota_D$.

Suppose we have graphical belief model Ψ_G , where $G = (\mathbb{V}, \mathbb{E})$ is a directed acyclic graph, and all valuations in Ψ_G are BPA (or commonality functions). Let $\tau = \bigoplus_{X \in \mathbb{V}} \alpha_X$ denote the joint b -valuation. Since G is acyclic, there exists a sequence of variables in \mathbb{V} such that for each edge $(X, Y) \in \mathbb{E}$, X precedes Y in the sequence. We will call such a sequence *compatible* with G (or simply G -compatible). We may have more than one G -compatible sequences.

Theorem 6. Let τ denote the joint BPA corresponding to a graphical belief model Ψ_G . Suppose r and s are disjoint subsets of variables such that the variables in r precede the variables in s in some G -compatible sequence, and that given r , the variables that follow all the variables in r are conditionally independent of the variables that precede some variable in r . Then, $\tau(s \mid r) = \tau(r \cup s) \ominus \tau(r)$ is a BPA.

Proof: Let g denote a G -compatible sequence such that r precedes s in it. Let $Pr(r)$ denote the predecessors of r in g , i.e.,

$$Pr(r) = \{X \in \mathbb{V} \setminus r \mid X \text{ precedes some variable in } r \text{ in } g\},$$

and let $An(r)$ denote $r \cup Pr(r)$. Because $\mathbb{V} \setminus An(r) \perp Pr(r) \mid r$, from Statement 5 of Proposition 3, it follows that $\tau = \tau(r \cup Pr(r)) \oplus \tau(\mathbb{V} \setminus An(r) \mid r)$, i.e., $\tau = \tau(An(r)) \oplus \tau(\mathbb{V} \setminus An(r) \mid r)$. But $\tau = \bigoplus_{X \in \mathbb{V}} \alpha_X$, and $\tau(An(r)) = \bigoplus_{X \in An(r)} \alpha_X$.

Thus, $\tau(\mathbb{V} \setminus An(r) \mid r) = \bigoplus_{X \in \mathbb{V} \setminus An(r)} \alpha_X$ is a combination of BPAs, and therefore, it is a BPA. Also, since r precedes s in g , $s \subset \mathbb{V} \setminus An(r)$, and $\tau(s \mid r)$ is a marginal of $\tau(\mathbb{V} \setminus An(r) \mid r)$. Therefore, $\tau(s \mid r)$ is a BPA. \square

Next, we will describe an example that illustrates why Theorem 6 is valid.

Example 3. Consider the graphical belief model Ψ_G as described in Example 2. There are two G -compatible sequences: DGB and DBG. Let $\tau = \delta \oplus \gamma \oplus \beta$ denote the joint BPA for $\{D, G, B\}$. Notice that

$$\begin{aligned}\tau(\{D\}) &= \tau^{\downarrow D} = \delta \oplus \gamma^{-G} \oplus \beta^{-B} = \delta, \\ \tau(\{D, G\}) &= \tau^{-B} = \delta \oplus \gamma \oplus \beta^{-B} = \delta \oplus \gamma, \\ \tau(\{D, B\}) &= \tau^{-G} = \delta \oplus \gamma^{-G} \oplus \beta = \delta \oplus \beta.\end{aligned}$$

Now consider, e.g., $\tau(\{G\} \mid \{D\}) = \tau(\{D, G\}) \ominus \tau(\{D\}) = (\delta \oplus \gamma) \ominus \delta$. Notice that the valuation being removed is included in the valuation it is being removed from. Thus, $\tau(\{G\} \mid \{D\}) = \gamma$, which is a BPA. Similarly, we can show that $\tau(\{B\} \mid \{D\})$, $\tau(\{G, B\} \mid \{D\})$, etc., are all BPA's.

8. Comparison with an Alternative Compositional Model for Belief Functions

For belief functions, the operator of composition was originally introduced in [16]. Since, as it will be shown in a simple example, it differs from the operator introduced here in Equation (1), we will use for the original operator a slightly different symbol.

Alternative composition for belief functions. Suppose ρ and σ are normal b -valuations for r and s , respectively. The old-composition of ρ and σ , written here as $\rho \succeq \sigma$, is defined for each $\mathbf{a} \subseteq \Omega_{r \cup s}$ by one of the following expressions:

$$\text{if } \sigma^{\downarrow r \cap s}(\mathbf{a}^{\downarrow r \cap s}) > 0 \text{ and } \mathbf{a} = \mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s} \text{ then } (\rho \succeq \sigma)(\mathbf{a}) = \frac{\rho(\mathbf{a}^{\downarrow r}) \cdot \sigma(\mathbf{a}^{\downarrow s})}{\sigma^{\downarrow r \cap s}(\mathbf{a}^{\downarrow r \cap s})}; \quad (11)$$

$$\text{if } \sigma^{\downarrow r \cap s}(\mathbf{a}^{\downarrow r \cap s}) = 0 \text{ and } \mathbf{a} = \mathbf{a}^{\downarrow r} \times \Omega_{s \setminus r} \text{ then } (\rho \succeq \sigma)(\mathbf{a}) = \rho(\mathbf{a}^{\downarrow r}); \quad (12)$$

$$\text{in all other cases } (\rho \succeq \sigma)(\mathbf{a}) = 0, \quad (13)$$

where $\mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s}$ denotes a join of the respective projections:

$$\mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s} = \{\mathbf{b} \in \Omega_{r \cup s} \mid \mathbf{b}^{\downarrow r} = \mathbf{a}^{\downarrow r} \ \& \ \mathbf{b}^{\downarrow s} = \mathbf{a}^{\downarrow s}\}.$$

Example 4. Consider Studený's example [1]. Suppose X, Y and Z are variables with state spaces $\Omega_X = \{x, \bar{x}\}$, $\Omega_Y = \{y, \bar{y}\}$, and $\Omega_Z = \{z, \bar{z}\}$. Consider two b -valuations ρ and σ for $\{X, Z\}$ and $\{Y, Z\}$, respectively, each having only two non-zero values: $\rho(\{x\bar{z}, \bar{x}z\}) = \rho(\{x\bar{z}, \bar{x}z\}) = 0.5$ and $\sigma(\{y\bar{z}, \bar{y}z\}) = \sigma(\{y\bar{z}, \bar{y}z\}) = 0.5$.

In [13], it is shown that $\rho \succeq \sigma$ has also only two non-zero values: $(\rho \succeq \sigma)(\{xy\bar{z}, \bar{x}\bar{y}z\}) = (\rho \succeq \sigma)(\{xy\bar{z}, \bar{x}\bar{y}z\}) = 0.5$. Thus, we see that $\rho \succeq \sigma$ is a proper normal b -valuation.

Also, $\rho \oplus \sigma$ is a normal b -valuation with value 0.25 for the following four sets: $\{xy\bar{z}, x\bar{y}\bar{z}\}$, $\{xy\bar{z}, \bar{x}\bar{y}\bar{z}\}$, $\{xy\bar{z}, \bar{x}\bar{y}z\}$, $\{xy\bar{z}, x\bar{y}z\}$. In contrast, $\rho \triangleright \sigma = \rho \oplus \sigma \ominus \sigma^{-Y}$ is a pseudo-BPA since $(\rho \triangleright \sigma)(\{\bar{x}\bar{y}z\}) = -0.25$ (the following are the remaining non-zero values of $\rho \triangleright \sigma$: $(\rho \triangleright \sigma)(\{xy\bar{z}, x\bar{y}\bar{z}\}) = 0.25$, $(\rho \triangleright \sigma)(\{xy\bar{z}, \bar{x}\bar{y}\bar{z}\}) = 0.25$, $(\rho \triangleright \sigma)(\{xy\bar{z}, \bar{x}\bar{y}z\}) = 0.5$, $(\rho \triangleright \sigma)(\{xy\bar{z}, x\bar{y}z\}) = 0.25$).

It is worth mentioning that the same result as $\rho \succeq \sigma$ is obtained also by the Srivastava-Cogger algorithm [38], but it need not be the case for different values of the ρ and σ b -valuations in this example.

To understand the differences between the two operators of composition, recall that a close connection exists between the combination operator \oplus and a notion of independence. Namely, after combining ρ for X and σ for Y , we get the valuation $\rho \oplus \sigma$ for $\{X, Y\}$, with respect to which variables X and Y are independent. Similarly, if ρ is a valuation for $\{X, Z\}$, and σ is a valuation for $\{Y, Z\}$, with respect to the valuation $\rho \oplus \sigma$ for $\{X, Y, Z\}$, variables X and Y are conditionally independent given Z . However, several other concepts of independence and conditional independence for belief functions exists in the literature. For a non-exhaustive survey, see [1, 3].

In their seminal papers, Dempster [6] and Walley and Fine [41] considered a type of independence that hold for variables X and Y with respect to BPA μ for $\{X, Y\}$ if

$$\mu(\mathbf{a}) = \begin{cases} \mu^{\downarrow X}(\mathbf{a}^{\downarrow X}) \cdot \mu^{\downarrow Y}(\mathbf{a}^{\downarrow Y}) & \text{if } \mathbf{a} = \mathbf{a}^{\downarrow X} \times \mathbf{a}^{\downarrow Y} \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } \mathbf{a} \in \Omega_{\{X, Y\}}.$$

Generalizing this idea, we define an alternative operation of combination, denoted by \oplus , for b -valuations ρ and σ (for r and s , respectively) as follows. Suppose K denotes $\sum\{\rho(\mathbf{a}^{\downarrow r}) \cdot \sigma(\mathbf{a}^{\downarrow s}) \mid \mathbf{a} \in \Omega_{r \cup s} \text{ s.t. } \mathbf{a} = \mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s}\}$. The combination $\rho \oplus \sigma$ is the b -valuation for $r \cup s$ given for all $\mathbf{a} \in \Omega_{r \cup s}$ by

$$(\rho \oplus \sigma)(\mathbf{a}) = \begin{cases} K^{-1} \rho(\mathbf{a}^{\downarrow r}) \sigma(\mathbf{a}^{\downarrow s}) & \text{if } K > 0, \mathbf{a} = \mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s}, \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

It is obvious that $\rho \oplus \sigma$ defined in Equation (14) is a proper normal b -valuation for $r \cup s$, and that \oplus satisfies all the three axioms of combination.

In a similar way, we define an alternative removal operator \ominus . Suppose ρ and σ are b -valuations for r and s , respectively, and suppose that ρ is normal. Let K denote $\sum\{\frac{\sigma(\mathbf{a}^{\downarrow s})}{\rho(\mathbf{a}^{\downarrow r})} \mid \mathbf{a} \in \Omega_{r \cup s} \text{ s.t. } \mathbf{a} = \mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s}, \rho(\mathbf{a}^{\downarrow r}) > 0\}$. $\sigma \ominus \rho$ is the b -valuation for $s \cup r$ given for all $\mathbf{a} \in \Omega_{s \cup r}$ by

$$(\sigma \ominus \rho)(\mathbf{a}) = \begin{cases} K^{-1} \left(\frac{\sigma(\mathbf{a}^{\downarrow s})}{\rho(\mathbf{a}^{\downarrow r})} \right) & \text{if } K > 0, \mathbf{a} = \mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s}, \rho(\mathbf{a}^{\downarrow r}) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, together with marginalization defined as in Section 7, we get an alternative VBS for belief functions. Let two normal b -valuations ρ and σ for r and s , respectively, be such that $\sigma^{\downarrow r \cap s} \gg \rho^{\downarrow r \cap s}$, i.e.,

$$\sigma^{\downarrow r \cap s}(\mathbf{x}) = 0 \implies \rho^{\downarrow r \cap s}(\mathbf{x}) = 0.$$

Consider $\mathbf{a} \subseteq \Omega_{r \cup s}$ for which $\mathbf{a} = \mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s}$. Then,

$$(\rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s})(\mathbf{a}) = \begin{cases} c \left(\frac{\rho(\mathbf{a}^{\downarrow r}) \sigma(\mathbf{a}^{\downarrow s})}{\sigma^{\downarrow r \cap s}(\mathbf{a}^{\downarrow r \cap s})} \right) & \text{if } \sigma^{\downarrow r \cap s}(\mathbf{a}^{\downarrow r \cap s}) > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (15)$$

(here c is a product of two normalization constants) which, due to the definition of old-composition, can be rewritten as

$$(\rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s})(\mathbf{a}) = c (\rho \supseteq \sigma)(\mathbf{a}).$$

Notice that because of the assumed domination, when computing $\rho \supseteq \sigma$, whenever the case described in Formula (12) of the definition of old composition applies, the value $\rho(\mathbf{a}^{\downarrow r}) = 0$.

Since for all $\mathbf{a} \neq \mathbf{a}^{\downarrow r} \bowtie \mathbf{a}^{\downarrow s}$, $(\rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s})(\mathbf{a}) = (\rho \supseteq \sigma)(\mathbf{a}) = 0$, we get

$$(\rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s})(\mathbf{a}) = c (\rho \supseteq \sigma)(\mathbf{a}), \quad \text{for all } \mathbf{a} \subseteq \Omega_{r \cup s}.$$

Since we know that both $\rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s}$ and $\rho \supseteq \sigma$ are normal b -valuations (for the former, it follows from the Proposition 1; for the latter, it is proved in [16]), it follows that $c = 1$.

Thus, we have shown that the operator of composition defined in [16] can be considered as a special case of composition in a VBS where combination is \oplus , removal is \ominus , and marginalization is the same as in the D-S theory. In other words, this VBS can be considered as an alternative uncertainty calculus that uses belief functions to represent knowledge, but with different rules for combination and removal than in the D-S theory, i.e., with an alternative definition of conditional independence, alternative semantics, etc.

9. VBS for Possibility Theory

The basic representation function in possibility theory is the possibility function [43, 7].

Possibility Function. A *possibility function* σ for s is a function $\sigma : 2^{\Omega_s} \rightarrow [0, 1]$ such that

$$\begin{aligned} \sigma(\{\mathbf{x}\}) &= 1 && \text{for some } \mathbf{x} \in \Omega_s; \text{ and} \\ \sigma(\mathbf{a}) &= \max\{\sigma(\{\mathbf{x}\}) \mid \mathbf{x} \in \mathbf{a}\} && \text{for all } \mathbf{a} \subseteq \Omega_s. \end{aligned}$$

Notice that a possibility function for s is completely specified by its values for singleton subsets of Ω_s . $\sigma(\mathbf{a})$ can be interpreted as the degree to which proposition \mathbf{a} is possible, and $1 - \sigma(\mathbf{a})$ is the degree to which proposition \mathbf{a} is not possible.

Valuations. A valuation σ for s is a function $\sigma : \Omega_s \rightarrow [0, 1]$. A valuation σ for s is *normal* if $\max\{\sigma(\mathbf{x}) \mid \mathbf{x} \in \Omega_s\} = 1$.

Combination. There are several definitions of combination in possibility theory. Any triangular norm can be used. However, the VBS framework requires combination to be associative. Also, the combination of two valuations is required to be a normal valuation (in case of no total conflict). For these reasons, we define combination as pointwise multiplication followed by normalization (assuming it is possible). Suppose ρ and σ are valuations for r and s , respectively. Let K denote $\max\{\rho(\mathbf{x}^{\downarrow r}) \sigma(\mathbf{x}^{\downarrow s}) \mid \mathbf{x} \in \Omega_{r \cup s}\}$. The combination $\rho \oplus \sigma$ is the valuation for $r \cup s$ given by

$$(\rho \oplus \sigma)(\mathbf{x}) = \begin{cases} K^{-1} \rho(\mathbf{x}^{\downarrow r}) \sigma(\mathbf{x}^{\downarrow s}) & \text{if } K > 0 \\ 0 & \text{if } K = 0 \end{cases} \quad \text{for all } \mathbf{x} \in \Omega_{r \cup s}. \quad (16)$$

It is easy to verify that the definition of combination in Equation (16) satisfies the three properties of combination.

Marginalization. Marginalization in possibility theory is maximization over the state space of the variable being marginalized. Suppose σ is a valuation for s and suppose $X \in s$. The marginal σ^{-X} for $s \setminus \{X\}$ is given by

$$\sigma^{-X}(\mathbf{y}) = \max\{\sigma(\mathbf{y}, x) \mid x \in \Omega_X\} \quad \text{for all } \mathbf{y} \in \Omega_{s \setminus \{X\}}. \quad (17)$$

It is easy to verify that the definition of marginalization in Equation (17) satisfies the four properties of marginalization.

Removal. Removal is division followed by normalization (assuming it is possible). Suppose σ is a valuation for s , and suppose ρ is a normal valuation for r . Let K denote $\max\{\frac{\sigma(\mathbf{x}^{\downarrow s})}{\rho(\mathbf{x}^{\downarrow r})} \mid \mathbf{x} \in \Omega_{s \cup r} \text{ s.t. } \rho(\mathbf{x}^{\downarrow r}) > 0\}$. The removal of ρ from σ is the valuation $\sigma \ominus \rho$ for $s \cup r$ given by

$$(\sigma \ominus \rho)(\mathbf{x}) = \begin{cases} K^{-1} \left(\frac{\sigma(\mathbf{x}^{\downarrow s})}{\rho(\mathbf{x}^{\downarrow r})} \right) & \text{if } K > 0, \rho(\mathbf{x}^{\downarrow r}) > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } \mathbf{x} \in \Omega_{s \cup r}. \quad (18)$$

It is easy to verify that the definition of removal in Equation (18) satisfies the three properties of removal.

Domination. Suppose ρ is a normal valuation for r . It is clear from the definition of removal in Equation (18) that $\rho \ominus \rho = \iota_\rho$ is a normal valuation for r whose values are as follows:

$$\iota_\rho(\mathbf{x}) = (\rho \ominus \rho)(\mathbf{x}) = \begin{cases} 1 & \text{if } \rho(\mathbf{x}) > 0 \\ 0 & \text{if } \rho(\mathbf{x}) = 0 \end{cases} \quad \text{for all } \mathbf{x} \in \Omega_r.$$

Now consider a normal valuation ρ' for r such that $\rho(\mathbf{x}) = 0 \Rightarrow \rho'(\mathbf{x}) = 0$. Then it is clear that $\rho' \oplus \iota_\rho = \rho'$. Thus, $\rho \gg \rho'$ if and only if $\rho(\mathbf{x}) = 0 \Rightarrow \rho'(\mathbf{x}) = 0$ for all $\mathbf{x} \in \Omega_r$.

10. Compositional Models for Possibility Theory

As mentioned in Section 9, in possibility theory, any (continuous) triangular norm can be used to define the combination operation. This fact is fully exploited by the definition of composition \triangleright_T introduced by Vejnarová in [40]. Let us see if it is possible to incorporate this generality in the VBS framework.

Triangular Norms. A *triangular norm* (or *t-norm*, in short) is a function $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ satisfying the following four properties. Suppose $a, b, c \in [0, 1]$. Then,

1. (*Commutativity*) $T(a, b) = T(b, a)$,
2. (*Associativity*) $T(T(a, b), c) = T(a, T(b, c))$,
3. (*Boundary condition*) $T(1, a) = a$,
4. (*Isotonicity*) if $a \leq b$ then $T(a, c) \leq T(b, c)$.

There are several well-known functions that qualify as a *t-norm*. One of them is the product *t-norm* $T_p(a, b) = a \cdot b$. Another is the Gödel *t-norm* $T_G(a, b) = \min\{a, b\}$.

It was already mentioned in [34] that normalization after application of the Gödel *t-norm* results in a loss of the associativity property. So, let us for a moment abandon assumption of normality for a combined valuation, and define a combination operation \oplus_T , parameterized by a *t-norm* T , as follows:

$$(\rho \oplus_T \sigma)(\mathbf{x}) = T(\rho(\mathbf{x}^{\downarrow r}), \sigma(\mathbf{x}^{\downarrow s})) \quad \text{for all } \mathbf{x} \in \Omega_{r \cup s},$$

Now, it is obvious that this operation meets all the properties required in Section 4: *domain* (with the exception of normality), *commutativity*, *associativity*. The definition of marginalization can remain the same as defined in Section 9:

$$\sigma^{-X}(\mathbf{y}) = \max\{\sigma(\mathbf{y}, x) \mid x \in \Omega_X\} \quad \text{for all } \mathbf{y} \in \Omega_{S \setminus \{X\}}.$$

However, in this case we would have to redefine the removal operation, which is an inverse of the combination operation. For this purpose, we use the inverse t -norm, also called *residuum*, defined as follows

$$T^{-1}(a, b) = \sup\{x \in [0, 1] \mid T(b, x) \leq a\}.$$

Thus, a removal operation \ominus_T , corresponding to the combination \oplus_T , can be defined as follows:

$$(\sigma \ominus_T \rho)(\mathbf{x}) = T^{-1}(\sigma(\mathbf{x}^{\downarrow S}), \rho(\mathbf{x}^{\downarrow r})) \quad \text{for all } \mathbf{x} \in \Omega_{r \cup S}.$$

Unfortunately, the removal operation \ominus_T defined in this way satisfies neither the *identity* property, nor the *combination and removal* property. For example, for the Gödel t -norm T_G ,

$$(\rho \oplus_{T_G} \rho) \ominus_{T_G} \rho = \rho \ominus_{T_G} \rho = \iota_\rho \neq \rho \oplus_{T_G} (\rho \ominus_{T_G} \rho) = \rho \oplus_{T_G} \iota_\rho,$$

which disproves both the mentioned properties. The only t -norm for which we know that both the properties hold is the product t -norm T_p , which leads to the operator of combination \oplus and its inverse \ominus studied in Section 9.

In this case,

$$(\rho \triangleright \sigma)(\mathbf{x}) = (\rho \oplus (\sigma \ominus \sigma^{\downarrow r \cap S}))(\mathbf{x}) = c \cdot T_p(\rho(\mathbf{x}^{\downarrow r}), T_p^{-1}(\sigma(\mathbf{x}^{\downarrow S}), \sigma(\mathbf{x}^{\downarrow r \cap S}))), \quad (19)$$

where $T_p(\rho(\mathbf{x}^{\downarrow r}), T_p^{-1}(\sigma(\mathbf{x}^{\downarrow S}), \sigma(\mathbf{x}^{\downarrow r \cap S})))$ is exactly the definition of the operator of composition $(\rho \triangleright_{T_p} \sigma)(\mathbf{x})$ introduced by Vejnarová in [40]. Notice that the constant c appearing in Equation (19) comes from normalization constants in Equations (16) and (18). However, since both $\rho \triangleright \sigma$ defined in Section 9 and $\rho \triangleright_{T_p} \sigma$ defined by Vejnarová are normalized possibility distributions, $c = 1$.

Thus, we conclude that the VBS framework can model possibility theory only for the case where combination is defined using the product t -norm, and in this case the operator of composition introduced in Section 4 and that defined by Vejnarová coincide.

11. Summary and Conclusions

We have described the VBS framework in general, and described the composition model in the VBS framework using the semantics of no double counting of knowledge. Since the VBS framework applies to many different uncertainty calculi, we have effectively defined the compositional model for any calculi that fits in the VBS framework.

We have shown that conditioning, which is defined using the combination, marginalization, and removal operators of VBS, can be expressed in terms of the composition operator. We have defined a decomposable compositional model as a special case of a compositional model in the VBS framework. We have shown that for decomposable compositional models, conditional valuations can be computed efficiently using local computation. All of this is done in the general VBS framework. Since the VBS framework applies to many different uncertainty calculi, we have effectively defined decomposable compositional models, and efficient computation of conditionals in decomposable compositional models, for any calculi that fits in the VBS framework. For example, because Spohn's epistemic theory fits in the VBS framework, all results described in this paper applies to this calculus.

We have compared the compositional model defined in this paper for D-S belief function theory with the one described in [16] for belief functions. Our conclusion is that although both of these compositional models are defined for belief functions and its alternative representations (BPA, commonality, etc.), the former is defined for *the* D-S belief function theory (that necessarily entails Dempster's rule of combination), and the latter for an alternative belief function theory that has \oplus as the rule of combination. Both of these theories fit in the VBS framework, but they have different semantics, different notions of conditional independence, etc.

We have compared the VBS compositional model defined in this paper with the compositional model described by Vejnarová [40] for possibility theory. The former is only defined for the version of possibility theory where

combination is the product t -norm, and the latter is defined more generally for any t -norm. For the case where combination is the product t -norm, the two coincide.

There are several issues related to the topics discussed in this paper that is not addressed here. First, we have an alternative theory of belief functions that has the alternative combination operator defined in Section 8. It is obvious that the alternative combination operator is computationally less complex than Dempster's rule of combination. But we do not know exactly how this alternative theory of belief functions compares with the D-S theory. In particular, one needs to develop semantics for this alternative theory that are tied to the alternative combination rule, \oplus .

We have described conditioning in decomposable compositional models. One natural question is what if the compositional model is not decomposable? The VBS framework does not require decomposability of the valuations. The Shenoy-Shafer architecture [35] or the Lauritzen-Spiegelhalter architecture [20] for computing marginals of the joint valuation in a VBS works whether the set of valuations are decomposable or not (with respect to the combination operator). Transforming a non-decomposable model to a decomposable model must therefore be possible. An efficient way of accomplishing this for compositional models is an open question. In this regard, a result that characterizes equivalent structures of decomposable models [19] seems to be promising.

Acknowledgements

This work has been supported in part by funds from grant GAČR 403/12/2175 to the first author, and from the Ronald G. Harper Distinguished Professorship at the University of Kansas to the second author. We are grateful to Milan Studený for valuable discussions and comments. A portion of this paper appeared in [14] and [15].

References

- [1] B. Ben-Yaghlane, P. Smets, and K. Mellouli. Belief function independence: II. The conditional case. *International Journal of Approximate Reasoning*, 31(1–2):31–75, 2002.
- [2] U. Bertele and F. Brioschi. *Nonserial Dynamic Programming*. Academic Press, New York, NY, 1972.
- [3] I. Couso, S. Moral, and P. Walley. Examples of independence for imprecise probabilities. In G. D. Cooman, F. G. Cozman, S. Moral, and P. Walley, editors, *Proceedings of the First International Symposium on Imprecise Probabilities and Their Applications (ISIPTA'99)*, pages 121–130, 1999.
- [4] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3(1):146–158, 1975.
- [5] W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, 11(4):427–444, 1940.
- [6] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, April 1967.
- [7] D. Dubois and H. Prade. *Possibility Theory; An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.
- [8] D. Geiger and J. Pearl. Logical and algorithmic properties of independence and their application to Bayesian networks. *Annals of Mathematics and Artificial Intelligence*, 2(1–4):165–178, 1990.
- [9] S. J. Haberman. *The Analysis of Frequency Data*. University of Chicago Press, Chicago, IL, 1974.
- [10] R. Jiroušek. Composition of probability measures on finite spaces. In D. Geiger and P. P. Shenoy, editors, *Uncertainty in Artificial Intelligence: Proceedings of the 13th Conference (UAI-97)*, pages 274–281, San Francisco, CA, 1997. Morgan Kaufmann.
- [11] R. Jiroušek. Foundations of compositional model theory. *International Journal of General Systems*, 40(6):623–678, 2011.
- [12] R. Jiroušek. Local computations in Dempster-Shafer theory of evidence. *International Journal of Approximate Reasoning*, 53(8):1155–1167, 2012.
- [13] R. Jiroušek and P. P. Shenoy. Compositional models in valuation-based systems. Working Paper 325, University of Kansas School of Business, Lawrence, KS, 2011.
- [14] R. Jiroušek and P. P. Shenoy. Compositional models in valuation-based systems. In T. Denoeux and M.-H. Masson, editors, *Belief Functions: Theory and Applications*, Advances in Intelligent and Soft Computing 164, pages 221–228, Heidelberg, 2012. Springer.
- [15] R. Jiroušek and P. P. Shenoy. Conditioning in decomposable compositional models in valuation-based systems. In S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. Yager, editors, *Advances in Computational Intelligence*, volume 4 of *Lecture Notes in Computer Science 300*, pages 676–685, Berlin, 2012. Springer.
- [16] R. Jiroušek, J. Vejnarová, and M. Daniel. Compositional models of belief functions. In G. de Cooman, J. Vejnarová, and M. Zaffalon, editors, *Proceedings of the 5th Symposium on Imprecise Probabilities and Their Applications (ISIPTA-07)*, pages 243–252, Prague, Czech Republic, 2007. Charles University Press.
- [17] J. Kohlas. *Information Algebras: Generic Structures for Inference*. Discrete Mathematics and Theoretical Computer Science. Springer-Verlag, London, UK, 2003.
- [18] J. Kohlas and N. Wilson. Semiring induced valuation algebras: Exact and approximate local computation algorithms. *Artificial Intelligence*, 172(11):1360–1399, 2008.
- [19] V. Kratochvíl. Characteristic properties of equivalent structures in compositional models. *International Journal of Approximate Reasoning*, 52(5):599–612, 2011.

- [20] S. L. Lauritzen and F. V. Jensen. Local computation with valuations from a commutative semigroup. *Annals of Mathematics and Artificial Intelligence*, 21(1):51–69, 1997.
- [21] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, 50(2):157–224, 1988.
- [22] J. Pearl and A. Paz. Graph-based logic for reasoning about relevance relations. In B. D. Boulay, D. Hogg, and L. Steele, editors, *Advances in Artificial Intelligence - II*, pages 357–363. North-Holland, Amsterdam, 1987.
- [23] A. Perez. ε -admissible simplification of the dependence structure of a set of random variables. *Kybernetika*, 13(6):439–449, 1977.
- [24] M. Pouly and J. Kohlas. *Generic Inference: A Unifying Theory for Automated Reasoning*. Wiley, Hoboken, NJ, 2011.
- [25] D. J. Rose. A graph-theoretic study of the numerical solution of sparse positive definite systems of linear equations. In R. C. Read, editor, *Graph Theory and Computing*, pages 183–217. Academic Press, New York, NY, 1973.
- [26] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.
- [27] G. Shafer. An axiomatic study of computation in hypertrees. Working Paper 232, School of Business, University of Kansas, Lawrence, KS, 1991.
- [28] P. P. Shenoy. A valuation-based language for expert systems. *International Journal of Approximate Reasoning*, 3(2):383–411, September 1989.
- [29] P. P. Shenoy. Valuation-based systems for propositional logic. In Z. Ras, M. Zemankova, and M. L. Emrich, editors, *Methodologies for Intelligent Systems*, 5, pages 305–312. North-Holland, New York, NY, 1990.
- [30] P. P. Shenoy. On Spohn’s rule for revision of beliefs. *International Journal of Approximate Reasoning*, 5(2):149–181, March 1991.
- [31] P. P. Shenoy. Valuation-based systems for discrete optimization. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 385–400. North-Holland, Amsterdam, 1991.
- [32] P. P. Shenoy. Valuation-based systems: A framework for managing uncertainty in expert systems. In L. A. Zadeh and J. Kacprzyk, editors, *Fuzzy Logic for the Management of Uncertainty*, chapter 4, pages 83–104. Wiley-Interscience, New York, NY, 1992.
- [33] P. P. Shenoy. Valuation-based systems for Bayesian decision analysis. *Operations Research*, 40(3):463–484, May–June 1992.
- [34] P. P. Shenoy. Conditional independence in valuation-based systems. *International Journal of Approximate Reasoning*, 10(3):203–234, April 1994.
- [35] P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, pages 169–198. North-Holland, 1990.
- [36] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publication Institute Statistique, Université Paris 8*, pages 229–231, 1959.
- [37] W. Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W. L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics*, volume 2, pages 105–134. D. Reidel, Netherlands, 1988.
- [38] R. P. Srivastava and K. O. Cogger. Beliefs on individual variables from a single source to beliefs on the joint space under Dempster-Shafer theory: An algorithm. In *Proceedings of the First International Conference on Agents and Artificial Intelligence (ICAART 2009)*, pages 191–197, Porto, Portugal, 2009. INSTICC Press.
- [39] M. Studený. *Probabilistic Conditional Independence Structures*. Springer-Verlag, 2005.
- [40] J. Vejnarová. Composition of possibility measures on finite spaces: Preliminary results. In B. Bouchon-Meunier and R. R. Yager, editors, *Proceedings of the 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU-98)*, pages 25–30, 1998.
- [41] P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *The Annals of Statistics*, 10(3):741–761, September 1982.
- [42] S. K. M. Wong, C. J. Butz, and Y. Xiang. A method for implementing a probabilistic model as a relational database. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence: Proceedings of the 11th Conference*, pages 556–564, San Francisco, CA, 1995. Morgan Kaufmann.
- [43] L. A. Zadeh. A theory of approximate reasoning. In J. Hayes and D. Mikulich, editors, *Machine Intelligence*, 9, pages 149–194. Elsevier, Amsterdam, 1979.