International Journal of Approximate Reasoning 72 (2016) 95–112

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Causal compositional models in valuation-based systems with examples in specific theories

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

[a] Faculty of Management, University of Economics, Jindřichův Hradec, Czech Republic
[b] Institute of Information Theory and Automation, Academy of Sciences, Prague, Czech Republic
[c] School of Business, University of Kansas, Lawrence, KS, USA

ARTICLE INFO

Article history:
Received 30 December 2014
Received in revised form 29 June 2015
Accepted 13 October 2015
Available online 19 October 2015

Keywords:
Valuation-based systems
Causality
Conditioning
Intervention

ABSTRACT

We show that Pearl's causal networks can be described using causal compositional models (CCMs) in the valuation-based systems (VBS) framework. One major advantage of using the VBS framework is that as VBS is a generalization of several uncertainty theories (e.g., probability theory, a version of possibility theory where combination is the product *t*-norm, Spohn's epistemic belief theory, and Dempster–Shafer belief function theory), CCMs, initially described in probability theory, are now described in all uncertainty calculi that fit in the VBS framework. We describe conditioning and interventions in CCMs. Another advantage of using CCMs in the VBS framework is that both conditioning and intervention can be easily described in an elegant and unifying algebraic way for the same CCM without having to do any graphical manipulations of the causal network. We describe how conditioning and intervention can be computed for a simple example with a hidden (unobservable) variable. Also, we illustrate the algebraic results using numerical examples in some of the specific uncertainty calculi mentioned above.

1. Introduction

In many situations, we are faced with the question of what would happen if we made some changes, such as if we intervened by an action that changes the status quo. In [15], Pearl shows that such questions can be answered using *causal probabilistic models*, because of their ability *to represent and respond to external or spontaneous changes*. He also states that such questions cannot be answered with the help of non-causal probabilistic models, in which causal relations are not taken into consideration. The reason is obvious. Suppose that we have a joint probability distribution for two binary variables S and A. We can check whether they are related or not. But even if $S = s$ is strongly positively related to $A = a$, there is no way to determine that one is a cause of the other. We can only see that $P(A = a|S = s) > P(A = a)$, and $P(S = s|A = a) > P(S = s)$. However, if we are informed that $S = s$ denotes 'there is smoke in a room,' and $A = a$ denotes 'smoke alarm is sounding,' then we have more information about the situation than we had before. We now know that $S = s$ is a cause of $A = a$. In this case, we can compute not only the conditional probabilities as indicated above, but also the effects of interventions. Consider, e.g., two interventions: creating smoke in the room (e.g., by lighting a smoke bomb), and sounding the alarm (e.g., by pushing the test button on the smoke alarm). Using Pearl's *do-calculus*, we denote the first intervention by $do(S = s)$, and the second one by $do(A = a)$. Using this notation, for the effect of the first intervention,

* Corresponding author at: Faculty of Management, University of Economics, Jindřichův Hradec, Czech Republic.
E-mail addresses: radim@utia.cas.cz (R. Jiroušek), pshenoy@ku.edu (P.P. Shenoy).

http://dx.doi.org/10.1016/j.ijar.2015.10.003
0888-613X/

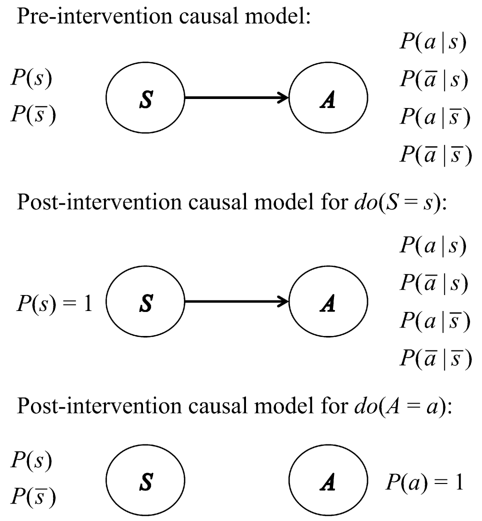

Fig. 1. Graphical modifications under interventions in the smoke-alarm causal model.

it is clear that $P(A = a|do(S = s)) = P(A = a|S = s) > P(A = a)$, because we know that $S = s$ is a cause of $A = a$. On the other hand, if we have $do(A = a)$, it activates the alarm, but we do not expect to find smoke in the room as a consequence. Clearly, $P(S = s|do(A = a)) = P(S = s)$ (see Fig. 1). In summary, when we have an intervention $do(X = x)$, where x is a state of variable X, we change the joint distribution such that all causal arcs that point to X disappear along with the conditional probability table for X (in the post–intervention model), we create a new probability distribution for X, $P(X = x) = 1$, and the remaining causal arcs and conditional probability tables remain unchanged.

In [8], causal probabilistic models, which are graphical probability models, were described by causal compositional models, which are algebraic probability models. Compositional models are based on the idea that a multidimensional probability model can be assembled, i.e., *composed*, from a collection of its lower-dimensional marginals. Naturally, it can be done only under some assumptions about the conditional independence relations between the variables for which the distribution is defined. Compositional models [7] are equivalent to Bayesian networks in the sense that a multidimensional probability distribution represented by a Bayesian network can be expressed also as a compositional model with practically the same number of parameters (probabilities), and vice-versa. The main difference stems from the fact that the building blocks of a Bayesian network are conditional probability distributions, whereas the building blocks of a compositional model are marginal distributions. These facts suggest that there are some computational advantages when computing marginals of a multidimensional model represented by a compositional model. As we will see in this paper, causal compositional models bring further advantages in comparison with causal probabilistic graphical models. The main advantage is that the effect of interventions can be described algebraically without having to do any graphical manipulations. One can compute the effect of an intervention to the original causal model without any graphical modifications, and the algebraic descriptions of conditioning and intervention differs only by a pair of parentheses. The disadvantages of causal compositional models are that these are relatively new, and consequently, not as well understood as causal graphical models.

The valuation-based system (VBS) framework was initially introduced in [20], and expanded further in [19,22,14,13,16]. VBS is an abstract framework that is able to capture many different systems such as propositional logic, uncertainty calculi, systems of equations, optimization problems, data-base systems, etc. VBS consists of objects called variables and valuations, and three operators called combination, marginalization, and removal, that operate on valuations. A valuation encodes knowledge about a subset of variables, and the subset of variables is referred to as the *domain* of the valuation. Combination allows us to combine two or more valuations resulting in a valuation that has aggregated knowledge from its constituents. In a specific VBS, if we combine all valuations in the system, we get what is called the *joint* valuation. Marginalization allows us to coarsen valuations allowing us to deduce the knowledge contained in a valuation for a variable in its domain. The coarsened valuation is called the *marginal* of the original valuation. Removal allows us to remove knowledge contained in a valuation from another valuation, and can be regarded as an inverse of combination. The three operators are required to obey some basic axioms, which enable us to compute marginals of the joint valuation for a particular variable of interest without having to explicitly compute the joint valuation. This is of interest in many systems as it may be computationally intractable to explicitly compute the joint valuation when we have many variables. The process of computing marginals of the joint valuation without explicitly computing the joint is called *local computation*.

The main contribution of this paper is a generalization of Pearl's probabilistic causal models to non-probabilistic uncertainty calculi. In [11], compositional models were described in the valuation-based system (VBS) framework, so that they are defined in all uncertainty calculi that the VBS framework is able to capture. In this paper, we extend compositional models described in the VBS framework further, so that we can represent the effects of conditioning and interventions. Thus, in doing so, we have extended Pearl's causal models to all uncertainty calculi that fit in the VBS framework. In particular, besides probability theory, possibility theory (but only when combination is the product t-norm), Spohn's epistemic belief theory (also known as kappa calculus), and Dempster–Shafer's belief function theory, fit in the VBS framework.

An outline of the remainder of the paper is as follows. Section 2 briefly reviews the VBS framework with examples from probability theory. Section 3 describes four different calculi that fit in the VBS framework. Also, it introduces the concept of *dominance* necessary to define composition of valuations, and this concept is described in greater detail in various theories that fit in the VBS framework. Section 4 reviews the composition operator and its basic properties in the VBS framework. Section 5 describes causal compositional models in the VBS framework, and making inferences in such models. We distinguish between conditioning and the effect of interventions. Finally, an algebraic example of hidden variables elimination in the VBS framework is presented in Section 6, and we provide some numerical examples in probability, possibility, and Spohn's theories.

2. Valuation-based systems

In this section we briefly review the basic concept of valuation-based systems. We use notation from [22] and [11] that have a detailed introduction to VBS, and to compositional models in VBS, respectively. Recall that in Section 3, we illustrate valuations and the operators of VBS for various uncertainty calculi that are captured by the VBS framework.

Let Φ denote a set whose elements are called *variables*. Elements of Φ are denoted by upper-case Roman alphabets, e.g., X, Y, and Z. Subsets of Φ are denoted by lower-case Roman alphabets, e.g., r, s, and t. Let Ψ denote a set whose elements are called *valuations*. Elements of Ψ are denoted by lower-case Greek alphabets, e.g., ρ, σ, and τ. 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.

It is useful to identify a subset of valuations $\Psi_n \subset \Psi$, whose elements are called *normal*. Normal valuations are those that are *coherent* in some sense. For example, in D-S belief function theory, normal valuations are basic probability assignment potentials (called b-valuations, see Section 3.4) whose values for non-empty subsets add to one.

For D-S belief function theory, it is useful to identify another subset of valuations $\Psi_p \subset \Psi$, whose elements are called *proper*. Proper valuations are those that are coherent in a sense that is different from normal. Proper valuations are b-valuations whose values are non-negative. The concept of proper valuation is not needed for probability, possibility, and Spohn's theory, and for these theories we assume that all valuations are proper. In [14], it is shown that for D-S belief function theory, the process of computation of marginals of joint belief functions involves intermediate valuations that are not proper.

We describe a specific VBS model by a pair (Φ_S, Ψ_S). This pair must be consistent in the sense that for each $X \in \Phi_S$ there exists a valuation $\rho \in \Psi_S$ for r such that $X \in r$, and that each valuation $\rho \in \Psi_S$ must be for variables $r \subseteq \Phi_S$. For example, consider $\Phi_S = \{U, X, Y, Z\}$. Then, in probability theory, Ψ_S may be specified in several ways. It may be specified, e.g., by a four-dimensional probability distribution $\tau(\{U, X, Y, Z\})$, or, it may be specified by four conditional probability distributions, one for each variable in Φ_S that constitute a factorization of $\tau(\{U, X, Y, Z\})$. A specific example is given in Section 6.

The VBS framework includes three basic operators — *combination, marginalization,* and *removal* — that are used to make inferences from the knowledge encoded in a VBS.

Combination. The combination operator $\oplus: \Psi \times \Psi \to \Psi_n$ represents aggregation of knowledge. We assume that it satisfies 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 first axiom states that after combination, the resulting valuation is normal. This suggests that the normalization operation is included in the combination operation. If ρ represents some knowledge about r, and σ represents some knowledge about s, then $\rho \oplus \sigma$ represents the aggregated knowledge about $r \cup s$. The commutativity and associativity axioms reflect the fact that the sequence in which knowledge is aggregated makes no difference in the aggregated result. Thus, we can write $\rho \oplus \sigma \oplus \tau$ without any parenthesis, and there is no ambiguity about the result of the combination regardless of the sequence used to combine the valuations.

In probability theory, for example, if ρ denotes a probability distribution of A, and σ denotes a conditional probability distribution of B given A, then $\rho \oplus \sigma$ constitutes the joint probability distribution of $\{A, B\}$. In this case, there is no need for normalization, i.e., the normalization constant is 1. For another example, if ρ denotes a joint probability distribution of $\{A, B\}$, and σ denotes the observation $B = b$, then $\rho \oplus \sigma$ constitutes the conditional distribution of $\{A, B\}$ given $B = b$. In this case, there is need for normalization and the normalization constant is equal to $P(B = b)$.

Marginalization. The marginalization operator $-X: \Psi \to \Psi$ allows us to coarsen knowledge by marginalizing X out of the domain of a valuation. We assume that it satisfies 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 says that if ρ represents some knowledge about r, then ρ^{-X} represents knowledge about $r \setminus \{X\}$ implied by ρ if we disregard X. The normal axiom states that marginalization preserves normality of knowledge. The third axiom says that the order in which variables are marginalized from a valuation makes no difference in the final marginal. The fourth axiom states that the computation of $(\rho \oplus \sigma)^{-X}$ can be done without having to compute $\rho \oplus \sigma$ first, it can be done by marginalizing X from ρ and then combining σ. The latter is computationally more efficient than the former.

Sometimes it is useful to use the notation $\rho^{\downarrow r \setminus \{X,Y\}}$ to denote $\rho^{-\{X,Y\}}$, when we wish to emphasize the variables that remain (instead of the variables that are marginalized out).

In probability theory, for example, if ρ denotes a joint distribution of $\{A, B\}$, then $\rho^{-B} = \rho^{\downarrow A}$ constitutes the marginal probability distribution of $\{A\}$, and $\rho^{-\{A,B\}} = \rho^{\downarrow \emptyset}$ is a valuation for $\emptyset$, i.e., a function that has only one value, whose value is equal to 1.

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

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 it matters. This can be ensured, e.g., when defining the composition operator in Section 4, by using the removal operator that is defined next.

Removal. The removal operator $\ominus: \Psi \times \Psi_n \to \Psi_n$ represents removing knowledge in the second valuation from the knowledge in the first valuation. We assume that it satisfies 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, $(\pi \oplus \theta) \ominus \rho = \pi \oplus (\theta \ominus \rho)$, and therefore we write $\pi \oplus \theta \ominus \rho$ without parenthesis to mean either $(\pi \oplus \theta) \ominus \rho$ or $\pi \oplus (\theta \ominus \rho)$.

We call $\sigma \ominus \rho$ the valuation resulting from removing ρ from σ. The domain axiom says that after removing ρ from σ, the result *must* be a valuation for $r \cup s$ as we cannot lose any of the domains of the two valuations. Therefore, $\sigma \oplus \rho \ominus \rho = \sigma \oplus \iota_\rho$, i.e., it is essentially σ that is extended from s to $r \cup s$. The identity axiom defines the removal operator as an inverse of the combination operator. In Section 3, we will describe explicitly the identity valuation ι_ρ for various uncertainty calculi that fit in the VBS framework.

In probability theory, for example, if σ denotes a joint probability distribution for $\{A, B\}$, and $\rho = \sigma^{-B}$ denotes the marginal probability distribution for A, then $\sigma \ominus \rho$ constitutes the conditional probability distribution of B given A assuming that the values of ρ are all positive.

For a detailed discussion and motivation for introducing these operators see [22], where a number of properties of combination, marginalization, and removal operators are stated and proved. For example, for valuations σ and θ for s and t, respectively, a normal valuation ρ for r, and $X \in s \setminus r$, it holds that

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

Again, we can write $\sigma \oplus \theta \ominus \rho$ without any parenthesis, and there is no ambiguity about the result regardless of whether ρ is removed from σ or from θ or from $\sigma \oplus \theta$.

In [14,13,16], the removal operator is defined using the algebraic theory of inverses in semigroups. This alternative way of defining the removal operator is mathematically more elegant than the way it is done here, but it involves a more sophisticated abstract algebraic theory (such as the theory of separative semigroups with inverses [2]) than the one we are using (the theory of regular semigroups [3]).

Conditional valuations. The removal operator allows us to define conditional valuations. Suppose τ is a normal valuation for t, and suppose r and s are disjoint subsets of t. Borrowing terminology from probability theory, $\tau^{\downarrow r \cup s} \ominus \tau^{\downarrow r}$ is called the *conditional* for s given r with respect to τ, and we denote it by $\tau(s|r)$.[1] Also, if $r = \emptyset$, let $\tau(s)$ denote $\tau(s|\emptyset)$.

[1] We caution the reader that $\tau(s|r)$ should be viewed as (the name of) a valuation, rather than the value of valuation τ for $s|r$. The latter makes no sense, of course. When we have to refer to values of $\tau(s|r)$, we will use the usual $\tau(s|r)(\cdot)$ notation.

Suppose τ is a normal valuation for t, and suppose r, s, and u are disjoint subsets of t. Then, the following statements (proved in [22]) hold.

1. $\tau(s|r)$ is a normal valuation for $r \cup s$.
2. $\tau(r) = \tau^{\downarrow r}$.
3. $\tau(r) \oplus \tau(s|r) = \tau(r \cup s)$.
4. $\tau(s|r) \oplus \tau(u|r \cup s) = \tau(s \cup u|r)$.
5. Suppose $s' \subset s$. Then $\tau(s|r)^{\downarrow r \cup s'} = \tau(s'|r)$.
6. $(\tau(s|r) \oplus \tau(u|r \cup s))^{-s} = \tau(u|r)$.
7. $\tau(s|r)^{-s} = \iota_{\tau(r)}$.

3. Examples of valuation-based systems

In this section, we will describe how probability theory, possibility theory with the product t-norm, Spohn's theory of epistemic beliefs, and Dempster–Shafer theory of belief functions fit in the framework of VBS.

One of the important concepts of this paper is a composition of two valuations that will be introduced in Section 4. However, one can compose only a pair of *non-conflicting* valuations. Therefore, another goal of this section is to describe, and illustrate by examples from the specific theories, which pairs of valuations are non-conflicting.

As defined in the identity property of removal, $\rho \oplus \iota_\rho = \rho$. 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, as a rule, there exists a class of normal valuations for r such that if ρ' is in this class, then $\rho' \oplus \iota_\rho = \rho'$. We will say that the valuations from this class are *dominated* by ρ. Thus, if ρ dominates ρ', written as $\rho \gg \rho'$, then $\rho' \oplus \iota_\rho = \rho'$.

States of variables. We use the symbol Ω_X for the set of possible values of a variable X, and we call Ω_X the state space of X. We assume that elements of Ω_X are mutually exclusive and exhaustive. We assume that all variables have finite state spaces.

Given a non-empty set s of variables, let Ω_s denote the Cartesian product of Ω_X for $X \in s$; $\Omega_s = \times_{X \in s} \Omega_X$. We call Ω_s the state space of s. We call the elements of Ω_s states of s. We use lower-case, bold-faced letters such as $\mathbf{a}, \mathbf{b}$, etc., to denote states. It is convenient to extend this terminology to the case where the set of variables s is empty. We adopt the convention that the state space for the empty set $\emptyset$ consists of a single state, and we use the symbol $\blacklozenge$ to denote that state; $\Omega_\emptyset = \{\blacklozenge\}$.

Projection of states. Projection simply means dropping extra coordinates; for example, if (w, x, y, z) is a state of $\{W, X, Y, Z\}$, then the projection of (w, x, y, z) to $\{W, Y\}$ is simply (w, y), which is a state of $\{W, Y\}$. If r and s are sets of variables, $r \subseteq s$, and $\mathbf{x}$ is a state of s, then $\mathbf{x}^{\downarrow r}$ denotes the projection of $\mathbf{x}$ to r. If $r = \emptyset$, then of course, $\mathbf{x}^{\downarrow r} = \blacklozenge$. If $\mathbf{x}$ is a state of r, $\mathbf{y}$ is a state of s, and $r \cap s = \emptyset$, then there is a unique state $\mathbf{z}$ of $r \cup s$ such that $\mathbf{z}^{\downarrow r} = \mathbf{x}$, and $\mathbf{z}^{\downarrow s} = \mathbf{y}$. Let $(\mathbf{x}, \mathbf{y})$ or $(\mathbf{y}, \mathbf{x})$ denote this $\mathbf{z}$. As per this notation, $(\mathbf{x}, \blacklozenge) = (\blacklozenge, \mathbf{x}) = \mathbf{x}$.

3.1. Probability theory

In the probabilistic framework, a valuation ρ for r is a mapping $\rho : \Omega_r \to [0, 1]$, where Ω_r denotes the set of all states of variables in r. Let $\mathbf{a}, \mathbf{b}$, etc., denote elements of Ω_r. Normal valuations are probability distributions, i.e. if ρ is a normal probability valuation for r, then

$$\sum_{\mathbf{b} \in \Omega_r} \rho(\mathbf{b}) = 1.$$

Marginalization. Marginalization in probability theory is summation. Suppose ρ is a probability valuation for r, and suppose $X \in r$. Then, ρ^{-X} is a probability valuation for $r \setminus \{X\}$ such that for each $\mathbf{a} \in \Omega_{r \setminus \{X\}}$,

$$\rho^{-X}(\mathbf{a}) = \sum_{b \in \Omega_X} \rho(\mathbf{a}, b). \tag{1}$$

Combination. Combination in probability theory is pointwise multiplication followed by normalization. Suppose ρ and σ are probability valuations for r and s, respectively. Then, $\rho \oplus \sigma$ is a normal probability valuation for $r \cup s$ such that for each $\mathbf{a} \in \Omega_{r \cup s}$,

$$(\rho \oplus \sigma)(\mathbf{a}) = \frac{\rho(\mathbf{a}^{\downarrow r})\, \sigma(\mathbf{a}^{\downarrow s})}{\sum_{\mathbf{b} \in \Omega_{r \cup s}} \rho(\mathbf{b}^{\downarrow r})\, \sigma(\mathbf{b}^{\downarrow s})}. \tag{2}$$

Thus, the combination $\rho \oplus \sigma$ is well defined only if the normalization constant in the denominator in Eq. (2) is positive. If the normalization constant is zero, then the valuations ρ and σ are said to be in *total conflict*, and cannot be combined.

Removal. The removal operator is pointwise division followed by normalization. Suppose ρ is a probability valuation for r, and σ is a normal probability valuation for s. Then, $\rho \ominus \sigma$ is a normal probability valuation for $r \cup s$ such that for each $\mathbf{a} \in \Omega_{r \cup s}$,

$$(\rho \ominus \sigma)(\mathbf{a}) = \begin{cases} \frac{\rho(\mathbf{a}^{\downarrow r})/\sigma(\mathbf{a}^{\downarrow s})}{\sum_{\mathbf{b} \in \Omega_{r \cup s}: \sigma(\mathbf{b}^{\downarrow s})>0} (\rho(\mathbf{b}^{\downarrow r})/\sigma(\mathbf{b}^{\downarrow s}))} & \text{if } \sigma(\mathbf{a}^{\downarrow s}) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Thus, $\rho \ominus \sigma$ is well-defined only if the normalization constant in the denominator is positive. If the normalization constant is zero, σ cannot be removed from ρ.

Domination. Suppose ρ is a normal probability valuation for r. From Eq. (3), the identity valuation $\iota_\rho = \rho \ominus \rho$ is a normal probability valuation for r such that for each $\mathbf{a} \in \Omega_r$,

$$\iota_\rho(\mathbf{a}) = \begin{cases} \frac{1}{|\{\mathbf{b} \in \Omega_r : \rho(\mathbf{b})>0\}|} & \text{if } \rho(\mathbf{a}) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, if ρ and ρ' are normal probability valuations for r, then $\rho' \oplus \iota_\rho = \rho'$ if and only if

$$\rho'(\mathbf{a}) > 0 \Rightarrow \rho(\mathbf{a}) > 0. \tag{4}$$

Thus, $\rho \gg \rho'$ if and only if (4) holds.

Identity valuations. The identity valuation ι_r for probability theory is the uniform probability distribution for r.

3.2. Possibility theory

Possibility theory [26,5] fits in the framework of VBS only when combination is based on the product t-norm. Similarly to probability theory, a possibility valuation ρ for r is a mapping $\rho : \Omega_r \to [0, 1]$. ρ is normal if

$$\max_{\mathbf{b} \in \Omega_r} \rho(\mathbf{b}) = 1.$$

Marginalization. Marginalization in possibility theory is maximization. Suppose ρ is a possibility valuation for r, and suppose $X \in r$. Then, ρ^{-X} is a possibility valuation for $r \setminus \{X\}$ such that for each $\mathbf{a} \in \Omega_{r \setminus \{X\}}$,

$$\rho^{-X}(\mathbf{a}) = \max_{b \in \Omega_X} \rho(\mathbf{a}, b). \tag{5}$$

Combination. Combination in possibility theory is pointwise multiplication followed by normalization. Suppose ρ and σ are possibility valuations for r and s, respectively. The combination $\rho \oplus \sigma$ is a normal possibility valuation for $r \cup s$ such that for each $\mathbf{a} \in \Omega_{r \cup s}$,

$$(\rho \oplus \sigma)(\mathbf{a}) = \frac{\rho(\mathbf{a}^{\downarrow r})\, \sigma(\mathbf{a}^{\downarrow s})}{\max_{\mathbf{b} \in \Omega_{r \cup s}} \rho(\mathbf{b}^{\downarrow r})\, \sigma(\mathbf{b}^{\downarrow s})}. \tag{6}$$

Thus, the combination is well defined only if the normalization constant in the denominator in Eq. (6) is positive. If the normalization constant is zero, then the valuations ρ and σ are said to be in *total conflict*, and cannot be combined.

Removal. Removal in possibility theory is division followed by normalization. Suppose ρ is a possibility valuation for r, and suppose σ is a normal possibility valuation for s. Then, $\rho \ominus \sigma$ is a normal possibility valuation for $r \cup s$ such that for each $\mathbf{a} \in \Omega_{r \cup s}$,

$$(\rho \ominus \sigma)(\mathbf{a}) = \begin{cases} \frac{\rho(\mathbf{a}^{\downarrow r})/\sigma(\mathbf{a}^{\downarrow s})}{\max_{\mathbf{b} \in \Omega_{r \cup s}: \sigma(\mathbf{b}^{\downarrow s})>0} \rho(\mathbf{b}^{\downarrow r})/\sigma(\mathbf{b}^{\downarrow s})} & \text{if } \sigma(\mathbf{a}^{\downarrow s}) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Thus, $\rho \ominus \sigma$ is well-defined only if the normalization constant in the denominator of Eq. (7) is positive. If the normalization constant is zero, then σ cannot be removed from ρ.

Domination. Suppose ρ is a normal possibility valuation for r. From Eq. (7), the identity valuation $\iota_\rho = \rho \ominus \rho$ is a normal possibility valuation for r such that for each $\mathbf{a} \in \Omega_r$,

$$\iota_\rho(\mathbf{a}) = \begin{cases} 1 & \text{if } \rho(\mathbf{a}) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, if ρ and ρ' are normal possibility valuations for r, then $\rho' \oplus \iota_\rho = \rho'$ if and only if

$$\rho'(\mathbf{a}) > 0 \Rightarrow \rho(\mathbf{a}) > 0. \tag{8}$$

Thus, $\rho \gg \rho'$ if and only if (8) holds, which is similar to that in probability theory.

Identity valuations. The identity valuation ι_r for possibility theory is the normal possibility valuation for r whose values are identically one for all $\mathbf{b} \in \Omega_r$.

3.3. Spohn's epistemic belief theory

In Spohn's epistemic belief theory [23,24,21], a Spohnian valuation ρ for r is a mapping $\rho : \Omega_r \to \mathbb{Z}^+$, where $\mathbb{Z}^+$ denotes the set of non-negative integers, $\{0, 1, 2, \ldots\}$. $\rho(\mathbf{a})$ represents the degree of disbelief in state $\mathbf{a}$. ρ is normal if

$$\min_{\mathbf{b} \in \Omega_r} \rho(\mathbf{b}) = 0.$$

Marginalization. Marginalization in Spohn's theory is minimization. Suppose ρ is a Spohnian valuation for r, and suppose $X \in r$. Then, ρ^{-X} is a Spohnian valuation for $r \setminus \{X\}$ such that for each $\mathbf{a} \in \Omega_{r\setminus\{X\}}$,

$$\rho^{-X}(\mathbf{a}) = \min_{b \in \Omega_X} \rho(\mathbf{a}, b). \tag{9}$$

Combination. Combination in Spohn's theory is pointwise addition followed by normalization. Suppose ρ and σ are Spohnian valuations for r and s, respectively. The combination $\rho \oplus \sigma$ is a normal Spohnian valuation for $r \cup s$ such that for each $\mathbf{a} \in \Omega_{r\cup s}$,

$$(\rho \oplus \sigma)(\mathbf{a}) = \rho(\mathbf{a}^{\downarrow r}) + \sigma(\mathbf{a}^{\downarrow s}) - \min_{\mathbf{b} \in \Omega r \cup s} (\rho(\mathbf{b}^{\downarrow r}) + \sigma(\mathbf{b}^{\downarrow s})) \tag{10}$$

The third term on the right hand side of Eq. (10) is a normalization constant that ensures that $\rho \oplus \sigma$ is normal. Unlike probability and possibility theories, we can always combine Spohnian valuations, i.e., no two Spohnian valuations can be in total conflict.

Removal. Removal in Spohn's theory is subtraction followed by normalization. Suppose ρ is a Spohnian valuation for r, and suppose σ is a normal Spohnian valuation for s. Then, $\rho \ominus \sigma$ is a normal Spohnian valuation for $r \cup s$ such that for each $\mathbf{a} \in \Omega_{r\cup s}$,

$$(\rho \ominus \sigma)(\mathbf{a}) = \rho(\mathbf{a}^{\downarrow r}) - \sigma(\mathbf{a}^{\downarrow s}) - \min_{\mathbf{b} \in \Omega r \cup s} (\rho(\mathbf{b}^{\downarrow r}) - \sigma(\mathbf{b}^{\downarrow s})). \tag{11}$$

The third term on the right hand side of Eq. (11) is a normalization constant that ensures that $\rho \ominus \sigma$ is normal. Unlike probability and possibility theories, we can always remove a normal Spohnian valuation from any Spohnian valuation.

Domination. Suppose ρ is a normal Spohnian valuation for r. From Eq. (11), the identity valuation $\iota_\rho = \rho \ominus \rho$ is a normal Spohnian valuation for r such that for each $\mathbf{a} \in \Omega_r$, $\iota_\rho(\mathbf{a}) = 0$. The identity valuation ι_r is the same as ι_ρ. Therefore, for any two normal Spohnian valuations ρ and ρ' for r, $\rho' \oplus \iota_\rho = \rho'$. Thus, $\rho \gg \rho'$ holds for any two Spohnian normal valuations.

Identity valuations. The identity valuation ι_r for Spohn's theory is the normal Spohnian valuation for r whose values are identically zero for all $\mathbf{b} \in \Omega_r$.

3.4. Dempster–Shafer belief function theory

In Dempster–Shafer (D-S) belief function theory [4,18], we can represent knowledge using either basic probability assignment valuations (b-valuations) or commonality valuations (c-valuations). A b-valuation μ for r is a function $\mu : 2^{\Omega_r} \to \mathbb{R}$, where 2^{Ω_r} denotes the set of all *non-empty* subsets of Ω_r,[2] and $\mathbb{R}$ denotes the set of real numbers. Let a, b, etc., denote elements of 2^{Ω_r}. A b-valuation μ for r is *normal* if

$$\sum_{\mathsf{b} \in 2^{\Omega_r}} \mu(\mathsf{b}) = 1,$$

and it is *proper* if for each $\mathsf{a} \in 2^{\Omega_r}$

$$\mu(\mathsf{a}) \geq 0.$$

Proper normal b-valuations are basic probability assignments (bpa) in D-S theory. If μ is a bpa for r, and $\mu(\mathsf{a}) > 0$, then a is said to be a *focal element* of μ. If μ has only one focal element a (so that $\mu(\mathsf{a}) = 1$), then μ is said to be *deterministic*.

A c-valuation θ for r is a function $\theta : 2^{\Omega_r} \to \mathbb{R}^+$. A c-valuation θ for r is *normal* if

$$\sum_{\mathsf{b} \in 2^{\Omega_r}} (-1)^{|\mathsf{b}|+1} \theta(\mathsf{b}) = 1,$$

and it is *proper* if for each $\mathsf{a} \in 2^{\Omega_r}$,

$$\sum_{\mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} \theta(\mathsf{b}) \geq 0.$$

Proper normal c-valuations correspond to commonality functions in D-S theory.

There is a one-to-one relationship between b-valuations and c-valuations. Suppose μ is a b-valuation for r. Then, the c-valuation θ_μ for r corresponding to b-valuation μ is as follows. For each $\mathsf{a} \in 2^{\Omega_r}$,

$$\theta_\mu(\mathsf{a}) = \sum_{\mathsf{b} \supseteq \mathsf{a}} \mu(\mathsf{b}). \tag{12}$$

Suppose θ is a c-valuation for r. Then the b-valuation μ_θ for r corresponding to θ is as follows. For each $\mathsf{a} \in 2^{\Omega_r}$,

$$\mu_\theta(\mathsf{a}) = \sum_{\mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} \theta(\mathsf{b}). \tag{13}$$

If μ is a normal b-valuation for r, then the c-valuation θ_μ corresponding to μ is also normal, and vice-versa. If μ is a proper b-valuation for r, then the c-valuation θ_μ corresponding to μ is also proper, and vice-versa. It follows from Eq. (12), that if θ is a commonality function for r, then $0 \leq \theta(\mathsf{a}) \leq 1$ for all $\mathsf{a} \in 2^{\Omega_r}$, and $\theta(\mathsf{a}) \geq \theta(\mathsf{b})$ whenever $\mathsf{a} \subseteq \mathsf{b}$.

For an example, consider the deterministic bpa μ for r such that $\mu(\mathsf{a}) = 1$ where $\mathsf{a} \in 2^{\Omega_r}$. Then the c-valuation θ_μ for r that corresponds to μ is such that for all $\mathsf{b} \in 2^{\Omega_r}$,

$$\theta_\mu(\mathsf{b}) = \begin{cases} 1 & \text{if } \mathsf{b} \subseteq \mathsf{a}, \\ 0 & \text{otherwise.} \end{cases}$$

Marginalization. Marginalization in D-S theory is addition of values of b-valuations. First, we generalize projection of states to subsets of states. Suppose $\mathsf{a} \in 2^{\Omega_r}$. Then $\mathsf{a}^{\downarrow r \setminus \{X\}} = \{\mathbf{b}^{\downarrow r \setminus \{X\}} : \mathbf{b} \in \mathsf{a}\}$. Notice that $\mathsf{a}^{\downarrow r \setminus \{X\}} \in 2^{\Omega_{r \setminus \{X\}}}$.

Suppose μ is a b-valuation for r, and suppose $X \in r$. Then, μ^{-X} is a b-valuation for $r \setminus \{X\}$ such that for each $\mathsf{a} \in 2^{\Omega_{r \setminus \{X\}}}$,

$$\mu^{-X}(\mathsf{a}) = \sum_{\mathsf{b} \in 2^{\Omega_r} : \mathsf{b}^{\downarrow r \setminus \{X\}} = \mathsf{a}} \mu(\mathsf{b}). \tag{14}$$

Combination. Combination in D-S theory is Dempster's rule of combination, which can be described using either b-valuations or using c-valuations. In terms of b-valuations, Dempster's rule is the product–intersection rule where the product of the b-valuation values is assigned to the intersection of the subsets, followed by normalization. In terms of c-valuations, Dempster's rule of combination is pointwise multiplication of c-valuations followed by normalization. The two rules are mathematically equivalent [18]. Here, we will formally describe Dempster's rule of combination in terms of c-valuations.

[2] Normally, 2^{Ω_r} denotes the set of all subsets of Ω_r, but here we exclude $\emptyset$.

We do this because consequently, we can easily define the removal operator, which can be considered an inverse of the combination operator, in terms of c-valuations.

Suppose θ_1 and θ_2 are c-valuations for r_1 and r_2, respectively. Then $\theta_1 \oplus \theta_2$ is a normal c-valuation for $r_1 \cup r_2$ such that for each $\mathsf{a} \in 2^{\Omega_{r_1 \cup r_2}}$,

$$(\theta_1 \oplus \theta_2)(\mathsf{a}) = \frac{\theta_1(\mathsf{a}^{\downarrow r_1})\,\theta_2(\mathsf{a}^{\downarrow r_2})}{\sum_{\mathsf{b} \in 2^{\Omega_{r_1 \cup r_2}}} (-1)^{|\mathsf{b}|+1} \theta_1(\mathsf{b}^{\downarrow r_1})\,\theta_2(\mathsf{b}^{\downarrow r_2})}. \tag{15}$$

Thus, combination is well defined only if the normalization constant in the denominator of Eq. (15) is positive. If the normalization constant is zero, then θ_1 and θ_2 are said to be in *total conflict*, and cannot be combined.

Removal. Removal in D-S theory is pointwise division of c-valuations followed by normalization. There is no easy definition of removal directly in terms of b-valuations. Suppose θ_1 is a proper c-valuation for r_1, and suppose θ_2 is a proper normal c-valuation for r_2. Then, $\theta_1 \ominus \theta_2$ is a normal c-valuation for $r_1 \cup r_2$ such that for each $\mathsf{a} \in 2^{\Omega_{r_1 \cup r_2}}$,

$$(\theta_1 \ominus \theta_2)(\mathsf{a}) = \begin{cases} \frac{\theta_1(\mathsf{a}^{\downarrow r_1})/\theta_2(\mathsf{a}^{\downarrow r_2})}{\sum_{\mathsf{b} \in 2^{\Omega_{r_1 \cup r_2}} : \theta_2(\mathsf{b}^{\downarrow r_2}) > 0} (-1)^{|\mathsf{b}|+1} \theta_1(\mathsf{b}^{\downarrow r_1})/\theta_2(\mathsf{b}^{\downarrow r_2})} & \text{if } \theta_2(\mathsf{a}^{\downarrow r_2}) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Like in probability and possibility theories, we get situations when we cannot remove one c-valuation from another. These are situations when the normalization factor in the denominator in Eq. (16) equals zero. Unlike probability and possibility theories, it should be noted that $\theta_1 \ominus \theta_2$ may not be a proper c-valuation. This may be true even if θ_1 and θ_2 are commonality functions, and θ_2 is a marginal of θ_1 (so that $r_2 \subset r_1$). This is one reason why we need the concept of proper valuations in the D-S theory.

Domination. Suppose θ is a commonality function for r. Then, it follows from Eq. (16) that $\iota_\theta = \theta \ominus \theta$ is a commonality function for r such that for each $\mathsf{a} \in 2^{\Omega_r}$,

$$\iota_\theta(\mathsf{a}) = \begin{cases} 1 & \text{if } \theta(\mathsf{a}) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Therefore, if θ and θ' are commonality functions for r, then $\theta' \oplus \iota_\theta = \theta'$ if and only if

$$\theta'(\mathsf{a}) > 0 \Rightarrow \theta(\mathsf{a}) > 0. \tag{18}$$

Thus, $\theta \gg \theta'$ if and only if Eq. (18) holds, which is similar to that in probability and possibility theories.

Domination can also be described for bpa's. Let μ_θ denote the bpa that corresponds to commonality function θ, and let a_{μ_θ} denote the union of all focal elements of μ_θ. If μ_{ι_θ} denotes the bpa that corresponds to ι_θ, then it follows from Eqs. (12) and (13) that μ_{ι_θ} is the deterministic bpa such that $\mu_{\iota_\theta}(\mathsf{a}_{\mu_\theta}) = 1$. And, if μ and μ' are bpa's for r, then $\mu' \oplus \mu_{\iota_\theta} = \mu'$ if and only if

$$\mu'(\mathsf{a}) > 0 \Rightarrow \mu(\mathsf{a}) > 0. \tag{19}$$

Thus, $\mu \gg \mu'$ if and only if Eq. (19) holds, which is also similar to that in probability and possibility theories.

Identity valuations. The identity valuation for r can be represented either by the deterministic bpa such that $\mu_{\iota_r}(\Omega_r) = 1$, or equivalently by the vacuous commonality function such that $\theta_{\iota_r}(\mathsf{a}) = 1$ for all $\mathsf{a} \in 2^{\Omega_r}$.

4. Composition operator

The composition operator aggregates knowledge encoded in two normal valuations while adjusting for the double counting of knowledge when it does matter. Suppose ρ and σ are normal valuations for r and s, respectively, and suppose that $\sigma^{\downarrow r\cap s} \gg \rho^{\downarrow r\cap s}$. The composition of ρ and σ, written as $\rho \triangleright \sigma$, is defined as follows:

$$\rho \triangleright \sigma = \rho \oplus \sigma \ominus \sigma^{\downarrow r\cap s}. \tag{20}$$

It follows from the third axiom for removal in Section 2 that the right-hand side of Eq. (20) can be written as either $(\rho \oplus \sigma) \ominus \sigma^{\downarrow r\cap s}$ or as $\rho \oplus (\sigma \ominus \sigma^{\downarrow r\cap s})$. For simplicity, we omit parenthesis.

The main idea behind compositional models is to construct a joint valuation from a set of marginals (normal valuations). For example, we may have a variety of data sources that provides marginals for $\{U\}$, $\{U, X\}$, $\{X, Y\}$, and $\{U, Y, Z\}$ subsets of variables. Let ρ_1, ρ_2, ρ_3, and ρ_4 be normal valuations for $\{U\}$, $\{U, X\}$, $\{X, Y\}$, and $\{U, Y, Z\}$, respectively. We are interested in constructing a joint distribution for $\{U, X, Y, Z\}$ from ρ_1, ρ_2, ρ_3, and ρ_4. One solution to this would be to compose these valuations in some sequence to construct joint valuation τ:

$$\tau = (((\rho_1 \rhd \rho_2) \rhd \rho_3) \rhd \rho_4 \quad (21)$$

In Eq. (21), we will assume that we have the dominance conditions to allow the three composition operations. For more information about the foundations of compositional model theory, see [6,7].

The following theorem summarizes the most important properties of the composition operator.

Theorem 1. *Suppose ρ, σ and τ are normal valuations for r, s, and t, respectively, and suppose that $\sigma^{\downarrow r\cap s} \gg \rho^{\downarrow r\cap s}$, $\tau^{\downarrow (r\cup s)\cap t} \gg (\rho \rhd \sigma)^{\downarrow (r\cup s)\cap t}$ and $\tau^{\downarrow r\cap t} \gg \rho^{\downarrow r\cap t}$. Then the following statements hold:*

1. (Domain): *$\rho \rhd \sigma$ is a normal valuation for $r \cup s$.*
2. (Composition preserves first marginal): *$(\rho \rhd \sigma)^{\downarrow r} = \rho$.*
3. (Reduction:) *If $s \subseteq r$ then, $\rho \rhd \sigma = \rho$.*
4. (Non-commutativity): *In general, $\rho \rhd \sigma \neq \sigma \rhd \rho$.*
5. (Commutativity under consistency): *ρ and σ have a common marginal for $r \cap s$, i.e., $\rho^{\downarrow r\cap s} = \sigma^{\downarrow r\cap s}$, if and only if $\rho \rhd \sigma = \sigma \rhd \rho$.*
6. (Non-associativity): *In general, $(\rho \rhd \sigma) \rhd \tau \neq \rho \rhd (\sigma \rhd \tau)$.*
7. (Associativity under special condition I): *If $r \supset (s \cap t)$ then, $(\rho \rhd \sigma) \rhd \tau = \rho \rhd (\sigma \rhd \tau)$.*
8. (Associativity under special condition II): *If $s \supset (r \cap t)$ then, $(\rho \rhd \sigma) \rhd \tau = \rho \rhd (\sigma \rhd \tau)$.*
9. (Stepwise combination): *If $(r \cap s) \subseteq t \subseteq s$ then, $(\rho \oplus \sigma^{\downarrow t}) \rhd \sigma = \rho \oplus \sigma$.*
10. (Stepwise composition): *If $(r \cap s) \subseteq t \subseteq s$ then, $(\rho \rhd \sigma^{\downarrow t}) \rhd \sigma = \rho \rhd \sigma$.*
11. (Exchangeability): *If $r \supset (s \cap t)$ then, $(\rho \rhd \sigma) \rhd \tau = (\rho \rhd \tau) \rhd \sigma$.*
12. (Simple marginalization): *If $(r \cap s) \subseteq t \subseteq r \cup s$ then, $(\rho \rhd \sigma)^{\downarrow t} = \rho^{\downarrow r\cap t} \rhd \sigma^{\downarrow s\cap t}$.*
13. (Irrelevant combination): *If $t \subseteq r \setminus s$ then, $\rho \rhd (\sigma \oplus \tau) = \rho \rhd \sigma$.*

Proof. All properties are proved in [11] with the exception of Properties 3, 7, 9, and 13.

Property 3 is a direct consequence of Property 2. To prove Property 7, it is sufficient to use the definition of the composition operator (Eq. 20), simple marginalization (Property 12), the commutativity and associativity of combination, and the fact that under the specified condition $r \cap s \cap t = s \cap t$ and $(r \cup s) \cap t = r \cap t$:

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

The proof of Property 9 is based just on the definition of the operator of composition and commutativity and associativity of combination:

$$(\rho \oplus \sigma^{\downarrow t}) \rhd \sigma = \rho \oplus \sigma^{\downarrow t} \oplus \sigma \ominus \sigma^{\downarrow t} = \rho \oplus \sigma.$$

To prove Property 13 we use the definition of the composition operator (Eq. 20), simple marginalization (Property 12), and the commutativity and associativity of combination:

$$\begin{aligned}
\rho \rhd (\sigma \oplus \tau) &= \rho \oplus (\sigma \oplus \tau) \ominus (\sigma \oplus \tau)^{\downarrow r\cap(s\cup t)} \\
&= \rho \oplus (\sigma \oplus \tau) \ominus (\sigma^{\downarrow r\cap s} \oplus \tau) \\
&= \rho \oplus \sigma^{\downarrow r\cap s} \ominus \sigma^{\downarrow r\cap s} \oplus \sigma \oplus \tau \ominus (\sigma^{\downarrow r\cap s} \oplus \tau) \\
&= \rho \oplus \sigma \ominus \sigma^{\downarrow r\cap s} \oplus (\sigma^{\downarrow r\cap s} \oplus \tau) \ominus (\sigma^{\downarrow r\cap s} \oplus \tau) \\
&= \rho \rhd \sigma. \qquad \square
\end{aligned}$$

In designing computational procedures for probabilistic compositional models in [1], we compensated the lack of associativity of the composition operator by the so-called *anticipating composition operator*. Its name is suggestive from the fact that it introduces an additional conditional independence relation into the result of composition — it *anticipates* the independence relation that is necessary for associativity, and therefore it must take into account the set of variables, for which the preceding distribution is defined. In this paper we introduce the anticipating operator of composition for VBS in the following way. Suppose ρ and σ are normal valuations for r and s, respectively, and suppose t is a subset of variables. Then,

$$\rho \circledcirc_t \sigma = (\rho \oplus \sigma^{\downarrow (t\setminus r)\cap s}) \rhd \sigma. \quad (22)$$

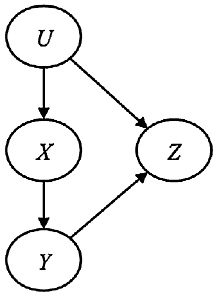

Fig. 2. A Markovian causal model.

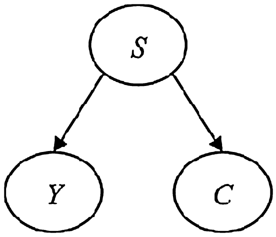

Fig. 3. A causal model for smoking (S), yellow teeth (Y), and lung cancer (C).

Notice that this composition operator is parameterized by subset t. If $(t \setminus r) \cap s = \emptyset$, then $\rho \circledtriangleright_t \sigma = \rho \triangleright \sigma$. The importance of this operator stems from the following assertion.

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

$$(\tau \triangleright \rho) \triangleright \sigma = \tau \triangleright (\rho \circledtriangleright_t \sigma). \tag{23}$$

Proof. The proof uses irrelevant combination (Property 13 of Theorem 1), and associativity under special condition II (Property 8 of Theorem 1):

$$\begin{aligned}(\tau \triangleright \rho) \triangleright \sigma &= (\tau \triangleright (\rho \oplus \sigma^{\downarrow (t\setminus r)\cap s})) \triangleright \sigma \\ &= \tau \triangleright ((\rho \oplus \sigma^{\downarrow (t\setminus r)\cap s}) \triangleright \sigma) \\ &= \tau \triangleright (\rho \circledtriangleright_t \sigma). \qquad \Box\end{aligned}$$

Notice that the marginal $\sigma^{\downarrow (t\setminus r)\cap s}$ appearing in Eq. (22) may be substituted by any valuation (on the same domain) that dominates $\sigma^{\downarrow (t\setminus r)\cap s}$ (e.g., by $\iota_\sigma{}^{\downarrow (t\setminus r)\cap s}$, or by $\iota_{(t\setminus r)\cap s}$). It is clear that if we do so, the proof of Theorem 2 remains valid.

How does the composition operator introduced above for VBS compare with the composition operators introduced in specific theories in [6,25,12]? For probability theory, both approaches are equivalent, they lead to the same formulae. In possibility theory, if we restrict only to the case where combination is the product t-norm, then both approaches also coincide. The situation is different for Dempster–Shafer belief function theory. The operator of composition introduced in [12] and the current approach differ substantially from each other. For a detailed description of these differences, we refer the reader to [9].

5. Causal compositional models

Suppose $\Phi_S = \{X_1, X_2, \ldots, X_n\}$. For each variable X_i, let $\mathscr{C}(X_i)$ denote the subset of the variables that are causes of X_i. We assume that $X_i \notin \mathscr{C}(X_i)$. $\{\mathscr{C}(X_i)\}_{i=1}^n$ constitutes a *causal model*. Using Pearl's terminology [15], we say that a causal model is *Markovian* if there exists an ordering of variables (without loss of generality we assume that it is the ordering $X_1, X_2, \ldots, X_n$) such that $\mathscr{C}(X_1) = \emptyset$, and for $i = 2, 3, \ldots, n$, $\mathscr{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$. Markovian causal models are causal models without feedback relations. An example of a Markovian causal model represented as a directed acyclic graph such that we have a directed arc from X_j to X_i when $X_j \in \mathscr{C}(X_i)$ is shown in Fig. 2.

Let r_i denote $\mathscr{C}(X_i) \cup \{X_i\}$. From here onwards, the symbol τ exclusively denotes causal models, i.e. if we have valuations ρ_i for r_i for $i = 1, \ldots, n$ a causal compositional model (CCM) τ is defined as follows:

$$\tau = (\ldots((\rho_1 \triangleright \rho_2) \triangleright \rho_3) \triangleright \ldots \triangleright \rho_{n-1}) \triangleright \rho_n = \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n. \tag{24}$$

To increase legibility of the formulae, we omit parentheses and assume that the composition operation is done successively from left to right as shown in Eq. (24).

Notice that all the properties of the composition operator, which are in Theorem 1, including Property 11, describe modifications that preserve the Markov property of the model, i.e., if any of these rules are applied to Markovian causal model, the resulting model is again Markovian. For example, if $\rho_1 \triangleright \rho_2 \triangleright \rho_3$ is a Markovian CCM, then $r_1 \supseteq r_2 \cap r_3$ guarantees that $\rho_1 \triangleright \rho_3 \triangleright \rho_2$ is also Markovian (it follows from the fact that under this assumption $r_3 \cap (r_1 \cup r_2) = r_3 \cap r_1$).

In causal models, there is a difference between conditioning and intervention. Suppose $S = 1$ denotes a person who smokes, $Y = 1$ denotes (nicotine-stained) yellow teeth, and $C = 1$ denotes presence of lung cancer. We assume $\mathscr{C}(S) = \emptyset$, $\mathscr{C}(Y) = \{S\}$, and $\mathscr{C}(C) = \{S\}$ (see Fig. 3). Conditioning on $Y = 0$ means including evidence that teeth are not stained (which lowers the chances that the person has cancer). On the other hand, the intervention denoted by $do(Y = 0)$ means a changed universe where the person gets his teeth whitened (e.g., from his dentist), but the chances of cancer remains unchanged.

Conditioning in CCMs. Suppose we have a CCM τ for Φ as in Eq. (24), and suppose we observe $X = a$ for some $a \in \Omega_X$, $X \in \Phi$. To simplify the exposition, in the rest of this section, let t denote $r_1 \cup \ldots \cup r_n$ and s denote $t \setminus \{X\}$ for some $X \in t$.

We model the observation $X = a$ as a normal valuation $\nu_{X=a}$ for $\{X\}$, which has the property that if ρ is a normal valuation for $\{X\}$, then $\nu_{X=a} \oplus \rho = \nu_{X=a}$ (assuming that the combination exists), and $\nu_{X=a} \ominus \rho = \nu_{X=a}$ (assuming that the removal exists). For example, in the D-S theory, we can represent the observation $X = a$ by a deterministic bpa $\mu_{X=a}$ such that $\mu_{X=a}(\{a\}) = 1$.

Suppose we wish to compute $\tau(s|X = a)$, the posterior valuation for s after observing $X = a$. In the VBS framework, we can compute $\tau(s|X = a)$ as follows:

$$\tau(s|X = a) = (\tau \oplus \nu_{X=a})^{-X} \tag{25}$$

We can achieve the same result using the composition operator as follows:

$$\tau(s|X = a) = (\nu_{X=a} \triangleright \tau)^{-X} = (\nu_{X=a} \triangleright (\rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n))^{-X}. \tag{26}$$

To see why, notice that $\nu_{X=a} \triangleright \tau = \nu_{X=a} \oplus \tau \ominus \tau^{\downarrow\{X\}} = \tau \oplus \nu_{X=a} \ominus \tau(\{X\}) = \tau \oplus \nu_{X=a}$.

Intervention in CCMs. As described in Section 1, we fully accept the philosophy of Pearl's causal networks [15], i.e., to compute the effects of intervention, we refer to Pearl's approach in which the effect of intervention $\tau(s|do(X = a))$ is computed as a conditioning, but in a different causal model. The new model differs from the original causal model $\{\mathscr{C}(X_i)\}_{i=1}^n$ only in one point: in the new model $\mathscr{C}(X)$ is replaced by $\emptyset$, which corresponds to deletion of edges heading to variable X. So the new CCM is defined[3]

$$\tau' = \tau^{\downarrow\{X\}} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n. \tag{27}$$

This is true, because CCM τ defined by Eq. (24) corresponds to the causal network with an acyclic directed graph $G = (\Phi, E)$, where $(X_j \to X_i) \in E$ *iff* $X_j \in \mathscr{C}(X_i)$. Obviously, CCM τ' defined by Eq. (27) corresponds to the causal network with an acyclic directed graph $\bar{G} = (\Phi, \bar{E})$, in which there is no edge heading to X and all the remaining edges from E are preserved; i.e., $\bar{E} = \{(X_j \to X_i) \in E : X_i \neq X\}$. Therefore, following Definition 3.2.1 in [15] (or Eq. (3.11) from the same source), we can see that the result of intervention performed in the causal model τ can be computed as a conditioning in the model τ'. Thus,

$$\tau(s|do(X = a)) = \left(\nu_{X=a} \triangleright \tau'\right)^{-X} = \left(\nu_{X=a} \triangleright \left(\tau^{\downarrow\{X\}} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n\right)\right)^{-X}. \tag{28}$$

As $\{X\} \cap r_n \subseteq (\{X\} \cup r_1 \cup \ldots \cup r_{n-1})$, we apply Property 8 of Theorem 1 to the right hand expression in (28) getting

$$\nu_{X=a} \triangleright \left(\tau^{\downarrow\{X\}} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n\right) = \nu_{X=a} \triangleright \left(\tau^{\downarrow\{X\}} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_{n-1}\right) \triangleright \rho_n.$$

Repeating this operation $n-1$ times we have

$$\nu_{X=a} \triangleright \left(\tau^{\downarrow\{X\}} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n\right) = \nu_{X=a} \triangleright \tau^{\downarrow\{X\}} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_{n-1} \triangleright \rho_n = \nu_{X=a} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_{n-1} \triangleright \rho_n$$

(the latter equality holds due to Property 3 of Theorem 1), which yields that

$$\tau(t|do(X = a)) = (\nu_{X=a} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n)^{-X}. \tag{29}$$

Notice the importance of the pair of brackets by which the Eqs. (26) and (29) differ from each other. This difference stems from the fact that the composition operator is not associative.

[3] Notice that, due to Property 3 of Theorem 1, we need not exclude the respective ρ_k (i.e. the ρ_k for which $X = X_k$) from the sequence.

Readers familiar with the Pearl's causal networks [15] have certainly noticed an advantage of CCM. In CCM, we can compute both conditioning and intervention from one causal compositional model (as in Eqs. (26) and (29)). In Pearl's causal networks, we have to consider two different networks. Conditioning is computed from the complete causal network. For the computation of intervention, we have to consider a reduced causal network where all the arrows heading to the intervention variable are deleted.

6. Elimination of hidden variables

In this section, as an illustration, we derive formulae for computation of conditioning and intervention in a simple causal compositional model with four variables U, X, Y, Z, the first of which is assumed to be hidden (unobservable). Suppose that $\mathscr{C}(U) = \emptyset$, $\mathscr{C}(X) = \{U\}$, $\mathscr{C}(Y) = \{X\}$, $\mathscr{C}(Z) = \{U, Y\}$, so that the causal model is Markovian (as depicted in Fig. 2). Also, suppose that the situation is described by a causal compositional model as follows:

$$\tau(\{U, X, Y, Z\}) = \rho_1(\{U\}) \triangleright \rho_2(\{U, X\}) \triangleright \rho_3(\{X, Y\}) \triangleright \rho_4(\{U, Y, Z\}). \tag{30}$$

In the CCM in Eq. (30), $\rho_1(\{U\})$ denotes a normal valuation for $\{U\}$, etc., and $\tau(\{U, X, Y, Z\})$ denotes the joint normal valuation for $\{U, X, Y, Z\}$. We will use the conditional valuation notation so that, e.g., $\tau(\{X, Y, Z\})$ denotes $\tau(\{U, X, Y, Z\})^{-U}$, etc.

As U is a hidden variable, we cannot estimate $\rho_1(U)$, $\rho_2(\{U, X\})$, and $\rho_4(\{U, Y, Z\})$ that include U in their domains. We will make the following assumptions. First, we can estimate any marginal of τ that includes non-hidden variables. Thus, as X, Y, and Z are non-hidden variables, we can estimate, e.g., the marginal $\tau(\{X, Y, Z\})$, which we will denote by $\hat{\tau}(\{X, Y, Z\})$. Second, without loss of generality, we can assume that $\rho_1(\{U\}), \ldots, \rho_4(\{U, Y, Z\})$ are marginals[4] of $\tau(\{U, X, Y, Z\})$. Therefore, we can estimate marginals of these valuations for non-hidden variables, and that these estimated marginals will agree with the corresponding marginals of $\hat{\tau}(\{U, X, Y, Z\})$. For example, $\rho_2(\{X\})$ can be estimated as $\hat{\tau}(\{X\})$, and $\rho_3(\{X, Y\})$ can be estimated as $\hat{\tau}(\{X, Y\})$.

Computation of the conditional $\tau(Z|X = x)$ is simple:

$$\begin{aligned}
\tau(Z|X = x) &= \left(\nu_{X=x} \triangleright \tau(\{U, X, Y, Z\})\right)^{\downarrow\{Z\}} \\
&\overset{(12)}{=} \left(\nu_{X=x} \triangleright \tau(\{U, X, Y, Z\})^{-U}\right)^{\downarrow\{Z\}} \\
&\overset{(12)}{=} \left(\nu_{X=x} \triangleright \tau(\{X, Y, Z\})^{-Y}\right)^{\downarrow\{Z\}} \\
&= \left(\nu_{X=x} \triangleright \tau(\{X, Z\})\right)^{\downarrow\{Z\}}.
\end{aligned}$$

Thus, we can estimate $\tau(Z|X = x)$ by $\left(\nu_{X=x} \triangleright \hat{\tau}(\{X, Z\})\right)^{\downarrow\{Z\}}$, which includes only observable variables. Notice that during these computations we used Property 12 of Theorem 1 twice. This is why the symbol (12) appears above the respective equality signs. This type of explanation will also be used in the subsequent computations.

To compute $\tau(Z|do(X = x))$, we use the properties of the composition and the anticipating operators defined in the preceding section. To simplify the exposition, we do just one elementary modification at each step, and thus the following computations may appear more cumbersome than they really are.

$$\begin{aligned}
\tau(Z|do(X = x)) &= \left(\nu_{X=x} \triangleright \rho_1(\{U\}) \triangleright \rho_2(\{U, X\}) \triangleright \rho_3(\{X, Y\}) \triangleright \rho_4(\{U, Y, Z\})\right)^{\downarrow\{Z\}} \\
&\overset{(3)}{=} \left(\nu_{X=x} \triangleright \rho_1(\{U\}) \triangleright \rho_3(\{X, Y\}) \triangleright \rho_4(\{U, Y, Z\})\right)^{\downarrow\{Z\}} \\
&\overset{(11)}{=} \left(\nu_{X=x} \triangleright \rho_3(\{X, Y\}) \triangleright \rho_1(\{U\}) \triangleright \rho_4(\{U, Y, Z\})\right)^{\downarrow\{Z\}} \\
&\overset{\text{Eq. (23)}}{=} \left(\nu_{X=x} \triangleright \rho_3(\{X, Y\}) \triangleright \left(\rho_1(\{U\}) \circledcirc_{\{X,Y\}} \rho_4(\{U, Y, Z\})\right)\right)^{\downarrow\{Z\}} \\
&\overset{(12)}{=} \left(\nu_{X=x} \triangleright \rho_3(\{X, Y\}) \triangleright \left(\rho_1(\{U\}) \circledcirc_{\{X,Y\}} \rho_4(\{U, Y, Z\})\right)^{-U}\right)^{\downarrow\{Z\}}.
\end{aligned}$$

To compute $\left(\rho_1(\{U\}) \circledcirc_{\{X,Y\}} \rho_4(\{U, Y, Z\})\right)^{-U}$ we take advantage of the idea of extension used by Pearl in [15]. It is one way of taking into account the mutual dependence of variables X, Y, and Z. It plays the same role as the inheritance of parents in Shachter's arc reversal rule [17]. Notice that the first of the following equalities holds because $(\{X, Y\} \setminus \{U\}) \cap \{U, Y, Z\} = (\{Y\} \setminus \{U\}) \cap \{U, Y, Z\} (= \{Y\})$. In the second equality, we also use the fact that ρ_1 is the marginal of ρ_2.

[4] The equations for recomputing valuations ρ_i so that they are marginals of τ are given in Theorem 10.9 in [7].

Table 1
An estimate of a probability distribution for $\{X, Y, Z\}$.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1
$\hat{\tau}(\{X, Y, Z\})$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$

Table 2
Computation of $\hat{\tau}(Z|do(X = 0))$ for the probability distribution in Table 1.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1
$\hat{\tau}(\{X, Y, Z\})$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$
$\hat{\tau}(\{X\})$	$\frac{9}{16}$				$\frac{7}{16}$			
$\hat{\tau}(\{Y\})$	$\frac{7}{16}$		$\frac{9}{16}$					
$\hat{\tau}(\{X, Y\})$	$\frac{5}{16}$		$\frac{1}{4}$		$\frac{1}{8}$		$\frac{5}{16}$	
$\hat{\tau}(\{X\}) \mathbin{\textcircled{$\triangleright$}}_{\{Y\}} \hat{\tau}(\{X, Y, Z\}) = \sigma(\{X, Y, Z\})$	$\frac{63}{1280}$	$\frac{63}{320}$	$\frac{81}{512}$	$\frac{81}{512}$	$\frac{49}{512}$	$\frac{49}{512}$	$\frac{63}{1280}$	$\frac{63}{320}$
$\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\})$	$\frac{5}{9}$		$\frac{4}{9}$		0		0	
$\sigma(\{Y, Z\})$	$\frac{371}{2560}$	$\frac{749}{2560}$	$\frac{531}{2560}$	$\frac{909}{2560}$				
$\sigma(\{Y\})$	$\frac{7}{16}$		$\frac{9}{16}$					
$(\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\})) \triangleright \sigma(\{Y, Z\})$	$\frac{53}{288}$	$\frac{107}{288}$	$\frac{59}{360}$	$\frac{101}{360}$	0	0	0	0
$\hat{\tau}(Z|do(X = 0))$	$\frac{167}{480}$	$\frac{313}{480}$						

$$
\begin{aligned}
\left(\rho_1(\{U\}) \mathbin{\textcircled{$\triangleright$}}_{\{X,Y\}} \rho_4(\{U, Y, Z\})\right)^{-U} &= \left(\rho_1(\{U\}) \mathbin{\textcircled{$\triangleright$}}_{\{Y\}} \rho_4(\{U, Y, Z\})\right)^{-U} \\
&\overset{(12)}{=} \left(\left(\rho_2(\{U, X\}) \mathbin{\textcircled{$\triangleright$}}_{\{Y\}} \rho_4(\{U, Y, Z\})\right)^{-X}\right)^{-U} \\
&\overset{\text{Eq. (22)}}{=} \left((\rho_4(\{Y\}) \oplus \rho_2(\{U, X\})) \triangleright \rho_4(\{U, Y, Z\})\right)^{\downarrow\{Y,Z\}} \\
&\overset{(9)}{=} \left((\rho_4(\{Y\}) \oplus \rho_2(\{X\})) \triangleright \rho_2(\{U, X\}) \triangleright \rho_4(\{U, Y, Z\})\right)^{\downarrow\{Y,Z\}} \\
&\overset{(3)}{=} \left((\rho_4(\{Y\}) \oplus \rho_2(\{X\})) \triangleright \rho_2(\{U, X\}) \triangleright \rho_3(\{X, Y\}) \triangleright \rho_4(\{U, Y, Z\})\right)^{\downarrow\{Y,Z\}} \\
&\overset{(7)}{=} \left((\rho_4(\{Y\}) \oplus \rho_2(\{X\})) \triangleright (\rho_2(\{U, X\}) \triangleright \rho_3(\{X, Y\})) \triangleright \rho_4(\{U, Y, Z\})\right)^{\downarrow\{Y,Z\}} \\
&\overset{(8)}{=} \left((\rho_4(\{Y\}) \oplus \rho_2(\{X\})) \triangleright (\rho_2(\{U, X\}) \triangleright \rho_3(\{X, Y\}) \triangleright \rho_4(\{U, Y, Z\}))\right)^{\downarrow\{Y,Z\}} \\
&\overset{\text{Eq. (30)}}{=} \left((\rho_4(\{Y\}) \oplus \rho_2(\{X\})) \triangleright \tau(\{U, X, Y, Z\})\right)^{\downarrow\{Y,Z\}} \\
&\overset{(12)}{=} \left((\rho_4(\{Y\}) \oplus \rho_2(\{X\})) \triangleright \tau(\{X, Y, Z\})\right)^{\downarrow\{Y,Z\}}.
\end{aligned}
$$

In what follows, we will take advantage of the fact that $\rho_1(\{U\}), \ldots, \rho_4(\{U, Y, Z\})$ are marginals of $\tau(\{U, X, Y, Z\})$. Therefore,

$$
\begin{aligned}
\hat{\tau}(Z|do(X = x)) &= \left(\nu_{X=x} \triangleright \hat{\rho}_3(\{X, Y\}) \triangleright \left((\hat{\rho}_4(\{Y\}) \oplus \hat{\rho}_2(\{X\})) \triangleright \hat{\tau}(\{X, Y, Z\})\right)^{\downarrow\{Y,Z\}}\right)^{\downarrow\{Z\}} \\
&= \left(\nu_{X=x} \triangleright \hat{\tau}(\{X, Y\}) \triangleright \left((\hat{\tau}(\{Y\}) \oplus \hat{\tau}(\{X\})) \triangleright \hat{\tau}(\{X, Y, Z\})\right)^{\downarrow\{Y,Z\}}\right)^{\downarrow\{Z\}} \\
&= \left(\nu_{X=x} \triangleright \hat{\tau}(\{X, Y\}) \triangleright \left(\hat{\tau}(\{X\}) \mathbin{\textcircled{$\triangleright$}}_{\{Y\}} \hat{\tau}(\{X, Y, Z\})\right)^{-X}\right)^{\downarrow\{Z\}}.
\end{aligned}
$$

6.1. A probabilistic example

Suppose that for the CCM in Eq. (30), we get estimates of a three-dimensional probability distribution $\hat{\tau}(\{X, Y, Z\})$ as given in Table 1.

Table 3
An estimate of a possibility distribution for $\{X, Y, Z\}$.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1
$\hat{\tau}(\{X, Y, Z\})$	$\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

In the probabilistic framework, $\nu_{X=a}$ is a normal valuation for $\{X\}$ representing the one-dimensional degenerate probability distribution for variable X as follows:

$$\nu_{X=a}(x) = \begin{cases} 1 & \text{if } x = a, \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

To evaluate expression for $\hat{\tau}(Z|do(X=0))$ we have to first compute $\hat{\tau}(\{X\}) \circledR_{\{Y\}} \hat{\tau}(\{X, Y, Z\})$, which is denoted by $\sigma(\{X, Y, Z\})$:

$$\begin{aligned}\sigma(\{X, Y, Z\}) = \hat{\tau}(\{X\}) \circledR_{\{Y\}} \hat{\tau}(\{X, Y, Z\}) &= (\hat{\tau}(\{X\}) \oplus \hat{\tau}(\{Y\})) \triangleright \hat{\tau}(\{X, Y, Z\}) \\ &= \hat{\tau}(\{X\}) \oplus \hat{\tau}(\{Y\}) \oplus \hat{\tau}(\{X, Y, Z\}) \ominus \hat{\tau}(\{X, Y\}),\end{aligned}$$

the computation of which is represented in the upper part of Table 2. Thus, e.g.,

$$\sigma(\{X, Y, Z\})(0, 0, 0) = \frac{\frac{9}{16} \cdot \frac{7}{16} \cdot \frac{1}{16}}{\frac{5}{16}} = \frac{63}{1280}.$$

The rest of the computation of $\hat{\tau}(Z|do(X=0))$ is simple:

$$\begin{aligned}\hat{\tau}(Z|do(X=0)) &= \left(\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\}) \triangleright \sigma(\{Y, Z\})\right)^{\downarrow\{Z\}} \\ &= \left(\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\}) \oplus \sigma(\{Y, Z\}) \ominus \sigma(\{Y\})\right)^{\downarrow\{Z\}},\end{aligned}$$

which is shown in the lower part of Table 2. Thus, e.g.,

$$(\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\}) \triangleright \sigma(\{Y, Z\}))(0, 0, 0) = \frac{\frac{5}{9} \cdot \frac{371}{2560}}{\frac{7}{16}} = \frac{53}{288},$$

etc., and

$$\hat{\tau}(Z|do(X=0))(0) = \frac{53}{288} + \frac{59}{360} + 0 + 0 = \frac{167}{480}.$$

Notice that $\hat{\tau}(Z|do(X=0))$ is different from $\hat{\tau}(Z|X=0)$:

$$\hat{\tau}(Z|X=0)(0) = \frac{\frac{3}{16}}{\frac{9}{16}} = \frac{1}{3}.$$

6.2. A possibilistic example

Consider again the same CCM given in Eq. (30) with the estimates of a three-dimensional possibility distribution $\hat{\tau}(\{X, Y, Z\})$ as given in Table 3.

In the possibilistic framework, $\nu_{X=a}$ is a normal valuation for $\{X\}$ representing the one-dimensional degenerate possibility distribution for variable X as follows (notice that it is exactly the same as in the probabilistic case):

$$\nu_{X=a}(x) = \begin{cases} 1 & \text{if } x = a, \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

The computation of $\hat{\tau}(Z|X=0)$ is shown in Table 4.

Computation of $\hat{\tau}(Z|do(X=0))$ is similar to the probabilistic case. First, we compute $\hat{\tau}(\{X\}) \circledR_{\{Y\}} \hat{\tau}(\{X, Y, Z\})$ (again, denoted by $\sigma(\{X, Y, Z\})$):

$$\begin{aligned}\sigma(\{X, Y, Z\}) = \hat{\tau}(\{X\}) \circledR_{\{Y\}} \hat{\tau}(\{X, Y, Z\}) &= (\hat{\tau}(\{X\}) \oplus \hat{\tau}(\{Y\}) \triangleright \hat{\tau}(\{X, Y, Z\}) \\ &= \hat{\tau}(\{X\}) \oplus \hat{\tau}(\{Y\}) \oplus \hat{\tau}(\{X, Y, Z\}) \ominus \hat{\tau}(\{X, Y\}).\end{aligned}$$

The details of the computation are shown in the upper part of Table 5. Thus, e.g.,

Table 4
Computation of $\hat{\tau}(Z|X=0)$ for the possibility distribution in Table 3.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1
$\hat{\tau}(\{X,Y,Z\})$	$\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1
$\hat{\tau}(\{X,Z\})$	$\frac{1}{2}$	1			$\frac{1}{4}$	1		
$\nu_{X=0} \oplus \hat{\tau}(\{X,Z\})$	$\frac{1}{2}$	1			0	0		
$\hat{\tau}(Z\|X=0)$	$\frac{1}{2}$	1						

Table 5
Computation of $\hat{\tau}(Z|do(X=0))$ for the possibility distribution in Table 3.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1
$\hat{\tau}(\{X,Y,Z\})$	$\frac{1}{4}$	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1
$\hat{\tau}(\{X\})$	1				1			
$\hat{\tau}(\{Y\})$	1		1					
$\hat{\tau}(\{X,Y\})$	1		$\frac{1}{2}$		$\frac{1}{4}$		1	
$\hat{\tau}(\{X\}) \circledcirc\!\!\triangleright_{\{Y\}} \hat{\tau}(\{X,Y,Z\}) = \sigma(\{X,Y,Z\})$	$\frac{1}{4}$	1	1	1	1	1	$\frac{1}{4}$	1
$\nu_{X=0} \triangleright \hat{\tau}(\{X,Y\})$	1		$\frac{1}{2}$		0		0	
$\sigma(\{Y,Z\})$	1	1	1	1				
$\sigma(\{Y\})$	1		1					
$(\nu_{X=0} \triangleright \hat{\tau}(\{X,Y\})) \triangleright \sigma(\{Y,Z\})$	1	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
$\hat{\tau}(Z\|do(X=0))$	1	1						

$$\sigma(\{X,Y,Z\})(0,0,0) = \frac{1 \cdot 1 \cdot \frac{1}{4}}{1} = \frac{1}{4}.$$

The rest of the computation of $\hat{\tau}(Z|do(X=0))$ is simple:

$$\hat{\tau}(Z|do(X=0)) = \left(\nu_{X=0} \triangleright \hat{\tau}(\{X,Y\}) \triangleright \sigma(\{Y,Z\})\right)^{\downarrow\{Z\}}$$
$$= \left(\nu_{X=0} \triangleright \hat{\tau}(\{X,Y\}) \oplus \sigma(\{Y,Z\}) \ominus \sigma(\{Y\})\right)^{\downarrow\{Z\}},$$

which is shown in the lower part of Table 5. Thus, e.g.,

$$(\nu_{X=0} \triangleright \hat{\tau}(\{X,Y\}) \triangleright \sigma(\{Y,Z\}))(0,0,0) = \frac{1 \cdot 1}{1} = 1,$$
$$(\nu_{X=0} \triangleright \hat{\tau}(\{X,Y\}) \triangleright \sigma(\{Y,Z\}))(0,1,0) = \frac{\frac{1}{2} \cdot 1}{1} = \frac{1}{2},$$

etc., and

$$\hat{\tau}(Z|do(X=0))(0) = \max\{1, \frac{1}{2}, 0, 0\} = 1,$$
$$\hat{\tau}(Z|do(X=0))(1) = \max\{1, \frac{1}{2}, 0, 0\} = 1.$$

Notice that $\hat{\tau}(Z|do(X=0))$ is different from $\hat{\tau}(Z|X=0)$.

6.3. A Spohnian example

Suppose that for the CCM in Eq. (30), we get estimates of a three-dimensional Spohnian disbelief function $\hat{\tau}(\{X,Y,Z\})$ as given in Table 6.

In Spohn's theory of epistemic beliefs, $\nu_{X=a}$ is a normal valuation for $\{X\}$ representing the one-dimensional Spohn's disbelief function for variable X as follows:

$$\nu_{X=a}(x) = \begin{cases} 0 & \text{if } x = a, \\ k & \text{otherwise,} \end{cases} \tag{33}$$

where k is a large integer, say an order of magnitude larger than any value in $\hat{\tau}(\{X,Y,Z\})$. For our example in Table 6, we can assume $k = 10$ or higher.

Table 6
An estimate of a Spohnian disbelief function for $\{X, Y, Z\}$.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1
$\hat{\tau}(\{X, Y, Z\})$	2	0	1	1	2	2	2	0

Table 7
Computation of $\hat{\tau}(Z|X=0)$ for the Spohnian disbelief function in Table 6.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1
$\hat{\tau}(\{X, Y, Z\})$	2	0	1	1	2	2	2	0
$\hat{\tau}(\{X, Z\})$	1	0			2	0		
$\nu_{X=0} \oplus \hat{\tau}(\{X, Z\})$	1	0			12	10		
$\hat{\tau}(Z|X=0)$	1	0						

Table 8
Computation of $\hat{\tau}(Z|do(X=0))$ for the Spohnian disbelief function in Table 6.

X	0	0	0	0	1	1	1	1
Y	0	0	1	1	0	0	1	1
Z	0	1	0	1	0	1	0	1
$\hat{\tau}(\{X, Y, Z\})$	2	0	1	1	2	2	2	0
$\hat{\tau}(\{X\})$	0				0			
$\hat{\tau}(\{Y\})$	0		0					
$\hat{\tau}(\{X, Y\})$	0		1		2		0	
$\hat{\tau}(\{X\}) \circledtriangleright_{\{Y\}} \hat{\tau}(\{X, Y, Z\}) = \sigma(\{X, Y, Z\})$	2	0	0	0	0	0	2	0
$\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\})$	0		1		12		10	
$\sigma(\{Y, Z\})$	0	0	0	0				
$\sigma(\{Y\})$	0		0					
$(\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\})) \triangleright \sigma(\{Y, Z\})$	0	0	1	1	12	12	10	10
$\hat{\tau}(Z|do(X=0))$	0	0						

The computation of $\hat{\tau}(Z|X=0)$ is shown in Table 7.

To evaluate $\hat{\tau}(Z|do(X=0))$, first we compute $\hat{\tau}(\{X\}) \circledtriangleright_{\{Y\}} \hat{\tau}(\{X, Y, Z\})$. This is shown in the upper part of Table 8. Thus, e.g.,

$$\sigma(\{X, Y, Z\})(0, 0, 0) = 0 + 0 + 2 - 0 = 2.$$

The rest of the computation of $\hat{\tau}(Z|do(X=0))$ is shown in the lower part of Table 8. Thus, e.g.,

$$(\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\}))(0, 0) = 0 + 0 - 0 = 0,$$

etc., and

$$(\nu_{X=0} \triangleright \hat{\tau}(\{X, Y\}) \triangleright \sigma(\{Y, Z\}))(0, 0, 0) = 0 + 0 - 0 = 0,$$

etc., and

$$\hat{\tau}(Z|do(X=0))(0) = \min\{0, 1, 12, 10\} = 0,$$

$$\hat{\tau}(Z|do(X=0))(1) = \min\{0, 1, 12, 10\} = 0.$$

Notice that $\hat{\tau}(Z|do(X=0))$ is different from $\hat{\tau}(Z|X=0)$.

7. Conclusions

We have described causal compositional models, originally introduced in [8] in the probabilistic framework, in the VBS framework. We have shown that both conditioning and interventions can easily be described using the composition operator. The composition operation requires a dominance relation for the valuations being composed, which is a new concept in the

VBS framework. We have described the dominance relation in probability, possibility, Spohn's disbelief function, and D-S belief function theories.

In Section 6, a simple example with a hidden variable illustrates the use of causal compositional models for computing the effects of conditioning and intervention. We have provided numerical examples with details of the computation of the effects of conditioning and intervention for probability theory, possibility theory, and Spohn's theory of epistemic beliefs.

We have not provided a numerical example for the case of D-S belief function theory as showing the details of computation of conditioning and intervention for this case would require too much space. Even if we chose a bpa with a few focal elements, doing removal would require the use of c-valuations, which may be non-zero for many (or all) of the 255 elements of $2^{\Omega_{\{X,Y,Z\}}}$. We do not know how to represent removal directly in terms of b-valuations.

Acknowledgements

This work has been supported in part by funds from grant Czech Science Foundation 15-00215S to the first author, and from the Ronald G. Harper Distinguished Professorship at the University of Kansas to the second author. This paper is an extended version of [10].

References

[1] V. Bína, R. Jiroušek, Marginalization in multidimensional compositional models, Kybernetika 42 (4) (2006) 405–422.
[2] A.H. Clifford, G.B. Preston, The Algebraic Theory of Semigroups, American Mathematical Association, Providence, RI, 1967.
[3] R. Croisot, Demi-groupes inversifs et demi-groupes reunions de demi-groupes simples, Ann. Sci. Éc. Norm. Super. 70 (4) (1953) 361–379.
[4] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Stat. 38 (2) (April 1967) 325–339.
[5] D. Dubois, H. Prade, Possibility Theory: An Approach to Computerized Processing of Uncertainty, Plenum Press, New York, 1988.
[6] R. Jiroušek, Composition of probability measures on finite spaces, in: D. Geiger, P.P. Shenoy (Eds.), Uncertainty in Artificial Intelligence: Proceedings of the 13th Conference, UAI-97, Morgan Kaufmann, San Francisco, CA, 1997, pp. 274–281.
[7] R. Jiroušek, Foundations of compositional model theory, Int. J. Gen. Syst. 40 (6) (2011) 623–678.
[8] R. Jiroušek, On causal compositional models: simple examples, in: A. Laurent, O. Strauss, B. Bouchon-Meunier, R.R. Yager (Eds.), IPMU 2014, Part I, in: CCIS, vol. 442, Springer, Heidelberg, 2014, pp. 517–526.
[9] R. Jiroušek, On two composition operators in Dempster–Shafer theory, in: T. Augustin, S. Doria, E. Miranda, E. Quaeghebeur (Eds.), Proceedings of the 9th International Symposium on Imprecise Probability: Theories and Applications, Aracne editrice, Rome, Italy, 2015, pp. 157–165.
[10] R. Jiroušek, P.P. Shenoy, Causal compositional models in valuation-based systems, in: F. Cuzzolin (Ed.), Belief Functions: Theory and Applications, in: Lect. Notes Comput. Sci., vol. 8764, Springer International Publishing, Switzerland, 2014, pp. 256–264.
[11] R. Jiroušek, P.P. Shenoy, Compositional models in valuation-based systems, Int. J. Approx. Reason. 55 (1) (2014) 277–293.
[12] R. Jiroušek, J. Vejnarová, M. Daniel, Compositional models of belief functions, in: G. de Cooman, J. Vejnarová, M. Zaffalon (Eds.), Proceedings of the 5th Symposium on Imprecise Probabilities and Their Applications, ISIPTA-07, Charles University Press, Prague, Czech Republic, 2007, pp. 243–252.
[13] J. Kohlas, Information Algebras: Generic Structures for Inference, Discrete Math. Theor. Comput. Sci., Springer-Verlag, London, UK, 2003.
[14] S.L. Lauritzen, F.V. Jensen, Local computation with valuations from a commutative semigroup, Ann. Math. Artif. Intell. 21 (1) (1997) 51–69.
[15] J. Pearl, Causality: Models, Reasoning, and Inference, Cambridge University Press, Cambridge, UK, 2009.
[16] M. Pouly, J. Kohlas, Generic Inference: A Unifying Theory for Automated Reasoning, Wiley, Hoboken, NJ, 2011.
[17] R.D. Shachter, Evaluating influence diagrams, Oper. Res. 34 (6) (1986) 871–882.
[18] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[19] G. Shafer, An axiomatic study of computation in hypertrees, Working Paper 232, School of Business, University of Kansas, Lawrence, KS, 1991.
[20] P.P. Shenoy, A valuation-based language for expert systems, Int. J. Approx. Reason. 3 (2) (September 1989) 383–411.
[21] P.P. Shenoy, On Spohn's rule for revision of beliefs, Int. J. Approx. Reason. 5 (2) (1991) 149–181.
[22] P.P. Shenoy, Conditional independence in valuation-based systems, Int. J. Approx. Reason. 10 (3) (1994) 203–234.
[23] W. Spohn, Ordinal conditional functions: a dynamic theory of epistemic states, in: W.L. Harper, B. Skyrms (Eds.), Causation in Decision, Belief Change, and Statistics, vol. 2, D. Reidel, Netherlands, 1988, pp. 105–134.
[24] W. Spohn, A general non-probabilistic theory of inductive reasoning, in: R.D. Shachter, T.S. Levitt, J.F. Lemmer, L.N. Kanal (Eds.), Uncertainty Artif. Intell., vol. 4, North-Holland, Amsterdam, 1990, pp. 149–158.
[25] J. Vejnarová, Composition of possibility measures on finite spaces: preliminary results, in: B. Bouchon-Meunier, R.R. Yager (Eds.), Proceedings of the 7th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, IPMU-98, 1998, pp. 25–30.
[26] L.A. Zadeh, A theory of approximate reasoning, in: J. Hayes, D. Mikulich (Eds.), Mach. Intell., vol. 9, Elsevier, Amsterdam, 1979, pp. 149–194.