Causal Compositional Models in Valuation-Based Systems

Radim Jiroušek[1] and Prakash P. Shenoy[2]

[1] Faculty of Management, University of Economics, Jarošovská 1117/II, 377 01 Jindřichův Hradec, Czech Republic
radim@utia.cas.cz

[2] University of Kansas School of Business, 1300 Sunnyside Ave., Summerfield Hall, Lawrence, KS 66045-7601, USA
pshenoy@ku.edu

Abstract. This paper shows that Pearl's causal networks can be described using compositional models in the valuation-based systems (VBS) framework. There are several advantages of using the VBS framework. First, 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). This implies that causal compositional models, initially described in probability theory, are now described in all uncertainty calculi that fit in the VBS framework. Second, using the operators of VBS, we describe how causal inference can be made in causal compositional models in an elegant and unifying algebraic way. This includes the computation of conditioning, and the computation of the effect of interventions.

Keywords: Valuation-based system, causality, conditionals, intervention, compositional model.

1 Introduction

In many situations we are faced with the question of what will happen if we make some changes, such as if we intervene by an action that changes the status quo. In [5], 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.* In [3], causal probabilistic models were described by causal compositional models in the probabilistic framework. In this paper we show that such causal compositional models can be described in the valuation-based systems (VBS) framework [7], so that they apply to all uncertainty calculi that fit in the VBS framework.

An outline of the paper is as follows. Section 2 reviews the VBS framework. Section 3 reviews the composition operator and its basic properties in the VBS framework. Section 4 describes causal compositional models in the VBS framework, and making inferences in such models. We distinguish between conditioning and the effect of interventions. We also describe a small illustrative example.

F. Cuzzolin (Ed.): BELIEF 2014, LNAI 8764, pp. 256–264, 2014.

2 Valuation-Based Systems

We use notation from [7] and [4] that have a detailed introduction to VBS and to compositional models in VBS. Φ denotes a set whose elements are called *variables* that 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). Ψ denotes 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 whose values for non-empty subsets add to one.

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

Combination. The combination operator $\oplus\colon \Psi \times \Psi \to \Psi_n$ represents aggregation of knowledge. 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*) $\rho \oplus (\sigma \oplus \tau) = (\rho \oplus \sigma) \oplus \tau$.

Marginalization. The marginalization operator $-X\colon \Psi \to \Psi$ allows us to coarsen knowledge by marginalizing X out of the domain of a valuation. 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 $(\rho \oplus \sigma)^{-X} = (\rho^{-X}) \oplus \sigma$.

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).

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 3, by the removal operator that is defined next.

Removal. This operator $\ominus\colon \Psi \times \Psi_n \to \Psi_n$ 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.

In [7], 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. $(\sigma \oplus \theta) \ominus \rho = (\sigma \ominus \rho) \oplus \theta$.
2. $(\sigma \ominus \rho)^{-X} = \sigma^{-X} \ominus \rho$.

Domination. 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, there may exist a class of normal valuations for r such that if ρ' is in this class, then $\rho' \oplus \iota_\rho = \rho'$. 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 $\rho' \oplus \iota_\rho = \rho'$.

3 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 $\rho \rhd \sigma$, is defined as follows:

$$\rho \rhd \sigma = \rho \oplus \sigma \ominus \sigma^{\downarrow r \cap s}. \tag{1}$$

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): *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 $\rho \rhd \sigma = \sigma \rhd \rho$.*
6. (Non-associativity): *Suppose τ is a normal valuation for t, and suppose $\tau^{\downarrow (r\cup s)\cap t} \gg (\rho \rhd \sigma)^{\downarrow (r\cup s)\cap t}$. Then, 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 composition): *If $(r \cap s) \subseteq t \subseteq s$ then, $(\rho \rhd \sigma^{\downarrow t}) \rhd \sigma = \rho \rhd \sigma$.*
10. (Exchangeability): *If $r \supset (s \cap t)$ then, $(\rho \rhd \sigma) \rhd \tau = (\rho \rhd \tau) \rhd \sigma$.*
11. (Simple marginalization): *If $(r \cap s) \subseteq t \subseteq r \cup s$ then, $(\rho \rhd \sigma)^{\downarrow t} = \rho^{\downarrow r\cap t} \rhd \sigma^{\downarrow s\cap t}$.*
12. (Irrelevant combination): *If $t \subseteq r \setminus s$ then, $\rho \rhd (\sigma \oplus \tau) = \rho \rhd \sigma$.*

Proof. All properties are proved in [4] with the exception of Properties 3, 7 and 12.

Property 3 is a direct consequence of Property 2. To prove Property 7, it is sufficient to use the definition of the composition operator (Equation 1), simple marginalization (Property 11), the commutativity and associativity of combination, and the fact that under the specified condition $(r \cup s) \cap t = r \cap t$:

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

To prove Property 12 we use the definition of the composition operator (Equation 1), simple marginalization (Property 11), and the commutativity and associativity of combination:

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

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,

$$\rho \circledtriangleright_t \sigma = (\rho \oplus \sigma^{\downarrow (t \setminus r) \cap s}) \rhd \sigma. \tag{2}$$

Notice that, as explained above, this composition operator is parameterized by subset t. If $(t \setminus r) \cap s = \emptyset$ then $\rho \circledtriangleright_t \sigma = \rho \rhd \sigma$. 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*

$$(\tau \rhd \rho) \rhd \sigma = \tau \rhd (\rho \circledtriangleright_t \sigma). \tag{3}$$

Proof. The proof uses irrelevant combination (Property 12 of Theorem 1), and associativity under special condition I (Property 7 of Theorem 1):

$$\begin{aligned}(\tau \rhd \rho) \rhd \sigma &= (\tau \rhd (\rho \oplus \iota_{(t \setminus r) \cap s})) \rhd \sigma \\ &= \tau \rhd ((\rho \oplus \iota_{(t \setminus r) \cap s}) \rhd \sigma) = \tau \rhd (\rho \circledtriangleright_t \sigma).\end{aligned}$$ ■

4 Causal Compositional Models

Suppose $\Phi = \{X_1, X_2, \ldots, X_n\}$. For each variable X_i, let $\mathcal{C}(X_i)$ denote the subset of the variables that are causes of X_i. We assume that $X_i \notin \mathcal{C}(X_i)$. $\{\mathcal{C}(X_i)\}_{i=1}^n$ constitutes a *causal model*. Using Pearl's terminology [5], 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 $\mathcal{C}(X_1) = \emptyset$, and for $i = 2, 3, \ldots, n$, $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$. Markovian causal models are causal models without feedback relations.

Let r_i denote $\mathcal{C}(X_i) \cup \{X_i\}$. 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:

$$\tau = (\ldots((\rho_1 \rhd \rho_2) \rhd \rho_3) \rhd \ldots \rhd \rho_{n-1}) \rhd \rho_n = \rho_1 \rhd \rho_2 \rhd \ldots \rhd \rho_n. \tag{4}$$

(To increase legibility of the formulae, we will not include parentheses if the composition operator is successively performed from left to right.)

Notice that all the properties of the composition operator, including Property 10, describe Markovian preserving modifications. For example, if $\rho_1 \rhd \rho_2 \rhd \rho_3$ is a Markovian CCM, then $r_1 \supseteq r_2 \cap r_3$ guarantees that $\rho_1 \rhd \rho_3 \rhd \rho_2$ is also Markovian (it follows from the fact that under this assumption $r_3 \cap (r_1 \cup r_2) = r_3 \cap r_1$).

Readers familiar with Pearl's causal networks [5] have certainly noticed that for the probabilistic case, CCM τ defined by formula (4) is exactly the causal network represented by an acyclic directed graph $G = (V, E)$ with $V = \Phi$, and there is an edge $(X_j \rightarrow X_i) \in E$ *iff* $X_j \in \mathcal{C}(X_i)$. The conditional probability distributions necessary to define the probabilistic causal network are $\rho(X_i|\mathcal{C}(X_i))$ for $i = 1, \ldots, n$.

4.1 Conditioning and Intervention

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 $\mathcal{C}(S) = \emptyset$, $\mathcal{C}(Y) = \{S\}$, and $\mathcal{C}(C) = \{S\}$. 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.

To simplify the exposition, in the rest of this subsection, let s denote $r_1 \cup \ldots \cup r_n$ and t denote $s \setminus \{X\}$ for some $X \in s$. Thus, in CCM $\tau = \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n$, conditioning by $X = \mathbf{x}$ leads to a valuation $\tau(t|X = \mathbf{x})$ for t.

As shown in [3], we can realize both the conditioning and intervention as a composition of the causal compositional model $\rho_1 \triangleright \ldots \triangleright \rho_n$ with a valuation $\nu_{|X;\mathbf{x}}$, which is a valuation for variable X expressing knowledge that $X = \mathbf{x}$. Using this notation we can apply the following simple formulae that were proved for the probabilistic framework in [3]:

$$\tau(t|X = \mathbf{x}) = \left(\nu_{|X;\mathbf{x}} \triangleright (\rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n)\right)^{-X}, \tag{5}$$

and

$$\tau(t|do(X = \mathbf{x})) = \left(\nu_{|X;\mathbf{x}} \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n\right)^{-X}. \tag{6}$$

Notice the importance of the pair of brackets by which the formulae above differ from each other. This difference arises from the fact that the operator of composition is not associative.

To clarify these formulae, consider for a moment, again, probabilistic interpretation. Then, the expression in formula (5) equals

$$\nu_{|X;\mathbf{x}} \triangleright (\rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n) = \nu_{|X;\mathbf{x}} \triangleright \tau(s) = \frac{\nu_{|X;\mathbf{x}} \cdot \tau(s)}{\tau(X)},$$

which is a probability distribution for variables s, and equals $\tau(t|X = \mathbf{x})$ for those combinations of values of variables s for which $X = \mathbf{x}$, and 0 for all the remaining combinations of values. Therefore $\tau(t|X = \mathbf{x}) = \left(\nu_{|X;\mathbf{x}} \triangleright \tau(s)\right)^{-X}$.

To explain formula (6) we have to make a reference to Pearl's causal networks [5], and to consider CCM

$$\sigma = \rho_0 \triangleright \rho_1 \triangleright \rho_2 \triangleright \ldots \triangleright \rho_n, \tag{7}$$

for a one-dimensional distribution $\rho_0(X)$ (ρ_0 may be considered uniform). At the end of the preceding section we said that CCM τ defined by formula (4) corresponds to the causal network with an acyclic directed graph $G = (\Phi, E)$, where $(X_j \to X_i) \in E$ *iff* $X_j \in \mathcal{C}(X_i)$. Obviously, CCM σ defined by formula (7) 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., $\bar{E} = \{(X_j \to X_i) \in E : X_i \neq X\}$.

Following Definition 3.2.1 in [5] (or formula (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 σ:

$$\tau(t|do(X = \mathbf{x})) = \sigma(t|X = \mathbf{x}) = \big(\nu_{|X;\mathbf{x}} \triangleright \sigma(s)\big)^{-X}$$
$$= \big(\nu_{|X;\mathbf{x}} \triangleright (\rho_0 \triangleright \rho_1 \triangleright \ldots \triangleright \rho_n)\big)^{-X}.$$

Applying Property 8 of Theorem 1 n-times (it is possible because $\nu_{|X;\mathbf{x}}$ and ρ_0 are defined for the same variable X) we get:

$$\nu_{|X;\mathbf{x}} \triangleright (\rho_0 \triangleright \rho_1 \triangleright \ldots \triangleright \rho_{n-1} \triangleright \rho_n) = \nu_{|X;\mathbf{x}} \triangleright (\rho_0 \triangleright \rho_1 \triangleright \ldots \triangleright \rho_{n-1}) \triangleright \rho_n = \ldots$$
$$= \nu_{|X;\mathbf{x}} \triangleright \rho_0 \triangleright \rho_1 \triangleright \ldots \triangleright \rho_{n-1} \triangleright \rho_n,$$

from which the formula (6) is obtained using Property 3 of Theorem 1.

Readers familiar with the Pearl's causal networks [5] have certainly noticed an advantage of CCM. In CCM, we can compute both conditioning and intervention from one causal compositional model as shown above. 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.

4.2 An Example: Elimination of Hidden Variables

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

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

In the CCM above, $\rho_1(U)$ denotes a normal valuation for U, etc., and $\tau(U, Y, X, Z)$ denotes the joint normal valuation for $\{U, Y, X, Z\}$. As U is a hidden variable, only $\rho_3(Y, X)$ can be estimated from data, all others include U in their domains. To simplify notation, we will let, e.g., $\tau(Y, X, Z)$ denote $\tau(U, Y, X, Z)^{-U}$, etc.

Computation of the conditional $\tau(Z|Y = \mathbf{y})$ is simple.

$$\tau(Z|Y = \mathbf{y}) \;=\; \big(\nu_{|Y;\mathbf{y}} \triangleright \tau(U, Y, X, Z)\big)^{\downarrow\{Z\}} \stackrel{(11)}{=} \Big(\nu_{|Y;\mathbf{y}} \triangleright \tau(U, Y, X, Z)^{-\{U\}}\Big)^{\downarrow\{Z\}}$$
$$\stackrel{(11)}{=} \Big(\nu_{|Y;\mathbf{y}} \triangleright \tau(Y, X, Z)^{-\{X\}}\Big)^{\downarrow\{Z\}} = \big(\nu_{|Y;\mathbf{y}} \triangleright \tau(Y, Z)\big)^{\downarrow\{Z\}}.$$

Thus we can estimate $\tau(Z|Y=\mathbf{y})$ by $\left(\nu_{|Y;\mathbf{y}} \triangleright \hat{\tau}(Y,Z)\right)^{\downarrow\{Z\}}$, which includes only observable variables. Notice that during these computations we used Property 11 of Theorem 1 twice. This is why the symbol (11) appears above the respective equality signs. This type of explanation will also be used in the subsequent computations.

To compute $\tau(Z|do(Y=\mathbf{y}))$ 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.

$$\begin{aligned}
\tau(Z|do(Y=\mathbf{y})) &= \left(\nu_{|Y;\mathbf{y}} \triangleright \rho_1(U) \triangleright \rho_2(U,Y) \triangleright \rho_3(Y,X) \triangleright \rho_4(U,X,Z)\right)^{\downarrow\{Z\}} \\
&\overset{(3)}{=} \left(\nu_{|Y;\mathbf{y}} \triangleright \rho_1(U) \triangleright \rho_3(Y,X) \triangleright \rho_4(U,X,Z)\right)^{\downarrow\{Z\}} \\
&\overset{(10)}{=} \left(\nu_{|Y;\mathbf{y}} \triangleright \rho_3(Y,X) \triangleright \rho_1(U) \triangleright \rho_4(U,X,Z)\right)^{\downarrow\{Z\}} \\
&\overset{\text{Th 2}}{=} \left(\nu_{|Y;\mathbf{y}} \triangleright \rho_3(Y,X) \triangleright \left(\rho_1(U) \circledtriangleright_{\{Y,X\}} \rho_4(U,X,Z)\right)\right)^{\downarrow\{Z\}} \\
&\overset{(11)}{=} \left(\nu_{|Y;\mathbf{y}} \triangleright \rho_3(Y,X) \triangleright \left(\rho_1(U) \circledtriangleright_{\{Y,X\}} \rho_4(U,X,Z)\right)^{-U}\right)^{\downarrow\{Z\}}.
\end{aligned}$$

To express $\left(\rho_1(U) \circledtriangleright_{\{Y,X\}} \rho_4(U,X,Z)\right)^{-U}$ we take advantage of the idea of extension used by Pearl in [5]. 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 property of Shachter's arc reversal rule [6].

$$\begin{aligned}
\left(\rho_1(U) \circledtriangleright_{\{Y,X\}} \rho_4(U,X,Z)\right)^{-U} &= \left(\rho_1(U) \circledtriangleright_{\{X\}} \rho_4(U,X,Z)\right)^{-U} \\
&\overset{(11)}{=} \left(\left(\rho_2(U,Y) \circledtriangleright_{\{X\}} \rho_4(U,X,Z)\right)^{-Y}\right)^{-U} \\
&= \left((\rho_4(X) \oplus \rho_2(U,Y)) \triangleright \rho_4(U,X,Z)\right)^{\downarrow\{X,Z\}} \\
&= \left((\rho_4(X) \oplus \rho_2(Y)) \triangleright \rho_2(U,Y) \triangleright \rho_4(U,X,Z)\right)^{\downarrow\{X,Z\}} \\
&\overset{(3)}{=} \left((\rho_4(X) \oplus \rho_2(Y)) \triangleright \rho_2(U,Y) \triangleright \rho_3(Y,X) \triangleright \rho_4(U,X,Z)\right)^{\downarrow\{X,Z\}} \\
&\overset{(7)}{=} \left((\rho_4(X) \oplus \rho_2(Y)) \triangleright (\rho_2(U,Y) \triangleright \rho_3(Y,X)) \triangleright \rho_4(U,X,Z)\right)^{\downarrow\{X,Z\}} \\
&\overset{(8)}{=} \left((\rho_4(X) \oplus \rho_2(Y)) \triangleright (\rho_2(U,Y) \triangleright \rho_3(Y,X) \triangleright \rho_4(U,X,Z))\right)^{\downarrow\{X,Z\}} \\
&= \left((\rho_4(X) \oplus \rho_2(Y)) \triangleright \tau(U,Y,X,Z)\right)^{\downarrow\{X,Z\}} \\
&\overset{(11)}{=} \left((\rho_4(X) \oplus \rho_2(Y)) \triangleright \tau(Y,X,Z)\right)^{\downarrow\{X,Z\}} \\
&= \left((\tau(X) \oplus \tau(Y)) \triangleright \tau(Y,X,Z)\right)^{\downarrow\{X,Z\}} = \left(\tau(Y) \circledtriangleright_{\{X\}} \tau(Y,X,Z)\right)^{-Y},
\end{aligned}$$

which eventually leads to

$$\begin{aligned}
\hat{\tau}(Z|do(Y=\mathbf{y})) \\
&= \left(\nu_{|Y;\mathbf{y}} \triangleright \rho_3(Y,X) \triangleright \left((\rho_4(X) \oplus \rho_2(Y)) \triangleright \tau(Y,X,Z)\right)^{\downarrow\{X,Z\}}\right)^{\downarrow\{Z\}} \\
&= \left(\nu_{|Y;\mathbf{y}} \triangleright \hat{\tau}(Y,X) \triangleright \left((\hat{\tau}(X) \oplus \rho_2(Y)) \triangleright \hat{\tau}(Y,X,Z)\right)^{\downarrow\{X,Z\}}\right)^{\downarrow\{Z\}} \\
&= \left(\nu_{|Y;\mathbf{y}} \triangleright \hat{\tau}(Y,X) \triangleright \left(\hat{\tau}(Y) \circledtriangleright_{\{X\}} \hat{\tau}(Y,X,Z)\right)^{-Y}\right)^{\downarrow\{Z\}} \\
&= \left(\nu_{|Y;\mathbf{y}} \triangleright \hat{\tau}(Y,X) \triangleright \left(\hat{\tau}(Y) \circledtriangleright_{\{X\}} \hat{\tau}(Y,X,Z)\right)^{-Y}\right)^{\downarrow\{Z\}}.
\end{aligned}$$

5 Conclusions

We have described causal compositional models, originally introduced in [3] in the probabilistic framework, in the VBS framework. Both conditioning and interventions can be described easily using the composition operator. A simple example illustrates the use of the composition operator for conditioning and intervention.

Acknowledgments. 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.

References

1. Bína, V., Jiroušek, R.: Marginalization in multidimensional compositional models. Kybernetika 42(4), 405–422 (2006)
2. Jiroušek, R.: Foundations of compositional model theory. Int. J. of General Systems. 40(6), 623–678 (2011)
3. Jiroušek, R.: On causal compositional models: Simple examples. In: Laurent, A., Strauss, O., Bouchon-Meunier, B., Yager, R.R., et al. (eds.) IPMU 2014, Part I. CCIS, vol. 442, pp. 517–526. Springer, Heidelberg (2014)
4. Jiroušek, R., Shenoy, P.P.: Compositional models in valuation-based systems. Int. J. of Approximate Reasoning 55(1), 277–293 (2014)
5. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge Univ. Press, NY (2009)
6. Shachter, R.: Evaluating influence diagrams. Operations Research 34(6), 871–882 (1986)
7. Shenoy, P.P.: Conditional independence in valuation-based systems. Int. J. of Approximate Reasoning 10(3), 203–234 (1994)
8. Shenoy, P.P.: No double counting semantics for conditional independence. In: Cozman, F.G., Nau, R., Seidenfeld, T. (eds.) Proc. of the 4th Int. Symposium on Imprecise Probabilities and Their Applications (ISIPTA 2005), pp. 306–314. SIPTA (2005)