Pardubice, Czech Republic
September 17–20, 2017

Proceedings of the 20th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty

Edited by
Vilém Novák
Masahiro Inuiguchi
Martin Štěpnička

IRAFM

UNIVERSITY OF OSTRAVA

Organizing institution:

Institute for Research and Applications of Fuzzy Modeling

Co-organizing institutions:

- Faculty of Mathematics and Physics of the Charles University
- Institute of Information Theory and Automation (ÚTIA) of the Czech Academy of Sciences

Title: **Proceedings of the 20th Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty**
CZECH-JAPAN SEMINAR 2017, Pardubice, Czech Republic, September 17–20, 2017

Editors: Vilém Novák, Masahiro Inuiguchi, Martin Štěpnička
Publisher: University of Ostrava

ISBN: 978-80-7464-932-5

Causality and Intervention in Business Process Management

Vladislav BÍNA[1] and Radim JIROUŠEK[1,2]

[1] *Faculty of Management, University of Economics, Prague*
Jarošovská 1117/II, 69121 Jindřichův Hradec
Czech Republic
E-mail: bina@fm.vse.cz

[2] *Inst. of Information Theory and Automation AV ČR*
Pod Vodárenskou věží 4, Praha 8
Czech Republic
E-mail: radim@utia.cas.cz

Abstract

Business processes, as well as various other fields of management and decision-making, involves causal relations. The paper presents an algebraic approach to the modeling of causality in systems of stochastic variables. The methodology is based on an operator of a composition that provides the possibility of composing a multidimensional distribution from low-dimensional building blocks taking advantage of the dependence structure of the problem variables. The authors formally define and demonstrate on a hypothetical example a surprisingly elegant unifying approach to conditioning by a single variable and to the evaluation of the effect of an intervention. Both operations are realized by the composition with a degenerated distribution and differ only in the sequence in which the operator of the composition is performed.

Keywords: Compositional models, operator of composition, causality, conditioning, intervention.

1 Introduction

The purpose of this paper is to foster a new method for causality modeling that can be incorporated into standard business process models. Without a doubt, causality and its proper treatment plays an important role in this modeling and may be used to solve a great variety of problems. For example, Feugas, Mosser, and Duchien [3] use it to predict the effect of business process evolution on the quality of service. They show the possibility of coping with one phenomenon that also will be discussed in this paper. Namely, with the fact that the characteristics of the quality of service, such as the response time, are influenced by *hidden factors.*

As Kirkwood argues in his book [9] *human beings are quick problem solvers,* because *quick problem solvers were the ones who survived.* For this, it is necessary to distinguish correctly between causes and effects. However, in many complex situations, it is not a simple task, as business processes are complex systems as a rule, especially when one considers systems with causal loops. To simplify the problem, we consider only the causal relations in this paper that can be well described by Markovian probabilistic models. In fact, we restrict our attention only to situations that are well described by Pearl's causal diagrams [15].

The exact approaches to causality modeling are often applied to the spheres of social sciences, e.g., applications in strategic management by Michael Ryall [16], in psychology by York Hagmayer et al. [4], in educational research and other social sciences [14]. For other examples the reader can also refer to an introduction to causal inference by Morgan and Winship [13] where the applications are aimed at the fields of sociology, political science, and economics.

Though graphical modeling for business process representation has been considered classical from its very beginning (flowcharts in the 20th, PERT in 50th, influence diagrams in 80th, . . .), the

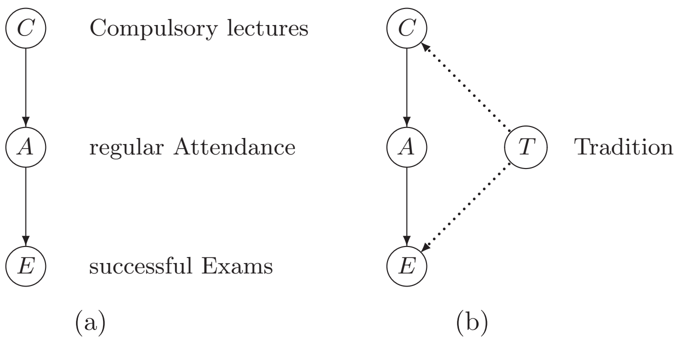

Figure 1: Causal models for the hypothetical example.

application of Pearl's causal diagrams (i.e., employment of directed graphs) to represent causality using arrows pointing from causes to effects, though quite intuitive and generally accepted, may bring some technical complications. Moreover, their advantage is diminishing when the diagrams containing hundreds of nodes are considered. In these situations, one may acknowledge an "algebraic" approach based on compositional models, in which computations of conditioning and interventions are computed in a similar way using the operator of a composition.

The goal of this paper is not to introduce a mathematical theory of compositional models. For this, the reader is referred to [6]. We want to convince the reader that incorporating of causal compositional models into business process models may help them to solve otherwise unsolvable problems. To make the following explanation as simple as possible, we present the results just using a trivial hypothetical example.

1.1 Motivation - Hypothetical Example

Analyzing data from universities, we found out that regular lecture attendance significantly increases the chances of passing exams on the first attempt. In connection with this, we were not surprised that students from universities where attendance at lectures was compulsory were slightly more successful in passing exams than students from universities where attendance was optional. These results led us to construct a simple and "natural" causal model with a causal diagram from Figure 1(a).

After publishing these research results, we succeeded in convincing the public to such an extent that eventually all universities excepted the rules that their lectures henceforth would be compulsory. We were quite satisfied that our research had made a real impact and expected that it would manifest itself in the form of better results at university exams. Unfortunately, the opposite proved to be true; on average, fewer students passed the exam on the first attempt. How could this be possible?

All this could happen because of a wrong causal graph. It was true that the new rules increased the percentage of regular attendance to lectures. It was also true that regular attendants were more successful at exams; yet, the impact of the law was negative because the causal model did not reflect another important cause - the fact that regulations and therefore also the behavior of students is different between traditional universities and new ones. Taking this fact into account, we can construct a new causal diagram (see Figure 1(b)), which describes the situation much better, and which can explain the decrease in success rates at the exams. Unfortunately, when we collected the data we did not recorded the type of university (traditional vs. new one), and therefore we have to treat the respective variable as a "hidden" one, which is why the respective arrows are dotted in Figure 1(b). As it will be shown later, if we had used the second causal model (even if we had to consider the type of university to be hidden - a non-measurable variable) to estimate the impact of an intervention realized by the new regulations, we would have really predicted a slight decline in the number of successful students.

In connection with this failure, one question arises. Was it possible to realize that the causal model from Figure 1(a) did not correspond to the situation described? The answer is positive and will be discussed in more details in Secion 4.

1.2 Causal Models

The purpose why we foster the employment of causal models in business process modeling is evident from the previous hypothetical example (for another example see, e.g., Váchová and Bína [18]). Prior to any serious intervention, the managerial board should take into consideration all possible ways of predicting a potential impact of the intended action.

In [15] (see page 29), Pearl says causal models may help to answer the following three types of queries:

- **predictions/conditioning** - are the students more successful if the lectures are compulsory?
- **interventions** - will the students be more successful if we introduce a regulation making the lectures compulsory?
- **counterfactuals** - would the students be more successful had the lectures been compulsory, given that they are not too successful and the lectures are optional?

In this paper, we will be interested only in the first two tasks: prediction (conditioning) and intervention. Let us stress the difference between them. While prediction gives evidence only about the students from universities where attendance at lectures is compulsory, intervention speaks about the success of all students when we make sure that lectures are compulsory at all universities. In what follows, we will show how to answer these questions with the help of compositional causal models. It will appear that, in contrast to causal diagrams, where the required results are computed in two totally dissimilar ways, we can use the same apparatus for causal compositional models. Syntactically, both these answers can be got as a composition of a model with a degenerated one-dimensional distribution. The only difference is just in a pair of parentheses. Unfortunately (but quite naturally), this pair of parentheses may bring some computational difficulties.

2 Compositional Models

In this paper we consider finite valued variables $A, E, C, T, X, Y, Z, W, ...$ that are denoted by upper case Latin characters. Boldface characters, such as $\mathbf{a}, \mathbf{b}, \mathbf{c}$ denote values, or combination of values of these variables. Sets of these variables are denoted by lower case characters $(x, y, ...)$, and their probability distributions are denoted using characters of a Greek alphabet, such as κ, λ, π. So, $\kappa(X_1, X_2, \ldots, X_n)$ denotes an n-dimensional probability distribution, and $\kappa(\mathbf{a})$ is a value of this distribution for state $\mathbf{a}$, which is a combination (vector) of values of variables $X_1, X_2, \ldots, X_n$. An $(n-1)$-dimensional marginal distribution of κ is denoted by κ^{-X_i}, or, denoting $x = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$, we use also the symbol $\kappa^{\downarrow x}$. The latter symbol $\kappa^{\downarrow y}$ can be used for any $y \subseteq \{X_1, \ldots, X_n\}$.

2.1 Operator of Composition

The term *compositional model* is derived from the fact that these multidimensional models (distributions) are *composed* (assembled) from a system of low-dimensional distributions using an *operator of composition* that was introduced first in [5]. For its rigorous mathematical definition and a detailed survey of its properties see [6]; here we will do just with the following brief summary.

Consider two (non-empty) sets of variables x and y (notice that they may be, but need not to be, disjoint, one may be a subset of the other). Let κ and λ be distributions defined for x and y, respectively, such that the marginal distribution $\lambda^{\downarrow x\cap y}$ dominates $\kappa^{\downarrow x\cap y}$, i.e.,

$$\forall \mathbf{a} \;\; \lambda^{\downarrow x\cap y}(\mathbf{a}) = 0 \Longrightarrow \kappa^{\downarrow x\cap y}(\mathbf{a}) = 0.$$

In this case we can define composition of κ and λ by the formula (notice that we define $\frac{0\cdot 0}{0} = 0$)

$$\kappa \triangleright \lambda = \frac{\kappa \cdot \lambda}{\lambda^{\downarrow x\cap y}}.$$

Note that for disjoint x and y the marginal $\kappa^{\downarrow x\cap y} = \lambda^{\downarrow x\cap y} = 1$, and $\kappa \triangleright \lambda = \lambda \triangleright \kappa$ simplifies to a product of (independent) distributions. If $\lambda^{\downarrow x\cap y}$ does not dominate $\kappa^{\downarrow x\cap y}$ then the composition is not defined. However, to avoid technical problems, let us assume in this paper that whenever we use the operator of composition we assume it is defined.

It is known that the composition of distributions $\kappa(x)$ and $\lambda(y)$ is always a distribution of variables $x \cup y$. In [6], many properties of the operator of the composition are proven. In this paper, we need mainly the following three that are formulated for $\kappa(x)$ and $\lambda(y)$.

$$(\kappa \triangleright \lambda)^{\downarrow x} = \kappa; \tag{1}$$

$$z \supseteq x \cap y \Longrightarrow (\kappa \triangleright \lambda)^{\downarrow z} = \kappa^{\downarrow x \cap z} \triangleright \lambda^{\downarrow y \cap z}, \tag{2}$$

$$x \supseteq y \Longrightarrow \kappa \triangleright \lambda = \kappa; \tag{3}$$

$$z \subseteq x \Longrightarrow \kappa^{\downarrow z} \triangleright \kappa = \kappa. \tag{4}$$

2.2 Causal Compositional Models

Consider variables $\{X_1, X_2, \ldots, X_n\}$. For each variable X_i let $\mathcal{C}(X_i)$ denote the set of the variables that are causes of X_i. Naturally, we assume that $X_i \notin \mathcal{C}(X_i)$. In accordance with Pearl [15], we say that the causal model is *Markovian* if there exists an ordering of variables (without loss of generality we will assume that it is the ordering $X_1, X_2, \ldots, X_n$) such that $\mathcal{C}(X_1) = \emptyset$, and for all $i = 2, 3, \ldots, n$, $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$. It means that considering only Markovian models, we exclude feedback relations.

Denote $x_i = \mathcal{C}(X_i) \cup \{X_i\}$. If we have probability distributions $\mu_i(x_i)$ we can construct a compositional causal model (CCM) as

$$\kappa(X_1, \ldots, X_n) = \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n). \tag{5}$$

Therefore the proper causal models for the example from Section 1.1 (Fig. 1(b)) is just

$$\tau(T) \triangleright \omega(C, T) \triangleright \alpha(A, C) \triangleright \varepsilon(A, E, T).$$

To understand properly expression (5) (and therefore also the above-presented causal model for the example) we have to highlight some properties following from formulas (1)–(2). First of all, we have to realize that the composition is not commutative. The reader can get a simple counterexample when considering a composition of two distributions $\kappa(x)$, $\lambda(y)$ for which $\kappa^{\downarrow x \cap y} \neq \lambda^{\downarrow x \cap y}$. Though the absence of commutativity is often considered a drawback, for causal models it can be an advantage, because causal relation is also asymmetric. As we will see later, the operation of composition is neither associative. Therefore, to make expression (5) unambiguous, we have to make an agreement that if not specified by parentheses otherwise, in all formulas the operator is performed from left to right, i.e.,

$$\mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)$$
$$= (\ldots((\mu_1(x_1) \triangleright \mu_2(x_2)) \triangleright \mu_3(x_3)) \triangleright \ldots \triangleright \mu_{n-1}(x_{n-1})) \triangleright \mu_n(x_n).$$

There are several theorems in [6] saying under what conditions one can change the ordering of distributions in a compositional model without influencing the resulting joint distribution. It is important to stress that for causal models, most of such transformations are forbidden. For causal models, we can consider only the orderings guaranteeing their Markovianity, i.e., for which $\mathcal{C}(X_i) \subseteq \{X_1, \ldots, X_{i-1}\}$. And it is the result of Kratochvíl that says that all these orderings define the same joint probability distribution $\kappa(X_1, \ldots, X_n)$ (see [10]).

In what follows, we will need the following three properties (for the respective proofs see [6]). For distributions $\mu_1(x_1), \mu_2(x_2), \mu_3(x_3)$

$$x_1 \supseteq x_2 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright \mu_3 \triangleright \mu_2; \tag{6}$$

$$x_1 \supseteq x_2 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright (\mu_2 \triangleright \mu_3); \tag{7}$$

$$x_2 \supseteq x_1 \cap x_3 \Longrightarrow \mu_1 \triangleright \mu_2 \triangleright \mu_3 = \mu_1 \triangleright (\mu_2 \triangleright \mu_3). \tag{8}$$

The reader certainly realized that all the above introduced properties, including property (6), describe Markovianity preserving modifications. It means that, for example, if $\mu_1 \triangleright \mu_2 \triangleright \mu_3$ is Markovian CCM then $x_1 \supseteq x_2 \cap x_3$ guarantees that $\mu_1 \triangleright \mu_3 \triangleright \mu_2$ is also Markovian (it follows from the fact that under this assumption $x_3 \cap (x_1 \cup x_2) = x_3 \cap x_1$).

2.3 Perfectization Procedure

Consider an arbitrary compositional model $\kappa(x_1 \cup \ldots \cup x_n) = \pi_1(x_1) \triangleright \ldots \triangleright \pi_n(x_n)$. Property 1 guarantees that $\pi_1 = \kappa^{\downarrow x_1}$ but generally $\pi_j \neq \kappa^{\downarrow x_j}$ for $j > 1$. This may be in some situations disadvantageous. For example, if we want to know probability of variable $X \in x_1$ then we can get it directly from π_1 without necessity to consider multidimensional model κ. However, if $X \notin x_1$ the only way how to ascertain the probability of X one has to consider $\kappa^{\downarrow x_1 \cup \ldots \cup x_j} = \pi_1 \triangleright \ldots \triangleright \pi_j$ such that $X \in x_1 \cup \ldots \cup x_j$. This disadvantage can easily be circumvented when restricting our attention to perfect sequences, for which, as it was proved in [6], all π_j are marginals of the multidimensional model κ.

We say that a compositional model $\kappa = \pi_1 \triangleright \ldots \triangleright \pi_n$ is *perfect* if

$$\begin{aligned}
&\pi_1 \triangleright \pi_2 = \pi_2 \triangleright \pi_1,\\
&\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_3 \triangleright (\pi_1 \triangleright \pi_2),\\
&\quad\vdots\\
&\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_n = \pi_n \triangleright (\pi_1 \triangleright \ldots \triangleright \pi_{n-1}).
\end{aligned}$$

Perhaps, it is not visible for the first sight, but it is not difficult to prove a stronger property than that mentioned above. In [6], it is proved that a compositional model $\kappa = \pi_1 \triangleright \ldots \triangleright \pi_n$ is perfect *if and only if* all distributions π_j are marginals of the multidimensional model κ. Therefore, considering that low-dimensional distributions π_k are carriers of local information, the constructed multidimensional model – if it is perfect – represents global information faithfully reflecting all of the local input.

It is important to realize that restricting our attention only to perfect models is not at the loss of generality. This is because any compositional model $\pi_1 \triangleright \ldots \triangleright \pi_n$ can be transformed into a perfect model $\kappa_1 \triangleright \ldots \triangleright \kappa_n$ using the following procedure:

$$\begin{aligned}
&\kappa_1 = \pi_1,\\
&\kappa_2 = \kappa_1^{\downarrow K_2 \cap K_1} \triangleright \pi_2,\\
&\kappa_3 = (\kappa_1 \triangleright \kappa_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} \triangleright \pi_3,\\
&\quad\vdots\\
&\kappa_n = (\kappa_1 \triangleright \ldots \triangleright \kappa_{n-1})^{\downarrow K_n \cap (K_1 \cup \ldots K_{n-1})} \triangleright \pi_n.
\end{aligned}$$

In this case, namely, $\kappa_1 \triangleright \ldots \triangleright \kappa_n$ is perfect and $\pi_1 \triangleright \ldots \triangleright \pi_n = \kappa_1 \triangleright \ldots \triangleright \kappa_n$. Moreover, since π_j and κ_j are defined for the same set of variables, it is evident that if $\pi_1 \triangleright \ldots \triangleright \pi_n$ is Markovian that also $\kappa_1 \triangleright \ldots \triangleright \kappa_n$ is Markovian.

Perhaps, we should mention that from the theoretical point of view the process of perfectization described above is simple. Unfortunately, it need not be valid from the point of view of computational complexity. The process requires marginalization that may be, in some situations, rather computationally expensive (for details on marginalization in compositional models see [1, 12]).

3 Conditioning and Intervention

Let us recall the difference between conditioning and intervention. Going back to our example, while conditioning by $C = 1$ gives evidence only about the universities, where attendance at lectures is compulsory, the intervention $do(C = 1)$ speaks about *all* universities under the assumption that we make sure (e.g., by a legislative act) that attendance at lectures is compulsory at all universities.

To simplify the following formulas, denote in the rest of this subsection $y = x_1 \cup \ldots \cup x_n$ and $z = y \setminus \{X\}$ for some $X \in x_1 \cup \ldots \cup x_n$. It means that considering a causal compositional model $\kappa(y) = \mu_1(x_1) \triangleright \mu_2(x_2) \triangleright \ldots \triangleright \mu_n(x_n)$ the conditioning by $X = \mathbf{a}$ leads to a distribution of variables z: $\kappa(z|X = \mathbf{a})$.

As shown in [2, 7], we can realize both the conditioning and intervention as a composition of the causal compositional model $\mu_1 \triangleright \ldots \triangleright \mu_n$ with a degenerated one-dimensional distribution $\delta_{\mathbf{a}}(X)$, which is a distribution of variable X achieving probability 1 for value $X = \mathbf{a}$, i.e.,

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

Using this denotation we use the following simple formulas proven in [7]:

$$\kappa(z|X=\mathbf{a}) = (\delta_{\mathbf{a}}(X) \triangleright (\mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m))^{-X},$$

and

$$\kappa(z|do(X=\mathbf{a})) = (\delta_{\mathbf{a}}(X) \triangleright \mu_1 \triangleright \mu_2 \triangleright \ldots \triangleright \mu_m)^{-X}.$$

The goal of this paper is to illustrate the applicability of causal compositional models to business process modeling, and thus we do not go more deeply into theory of compositional models (for this the reader is referred to [6, 7, 8], nevertheless let us stress the importance of the pair of brackets in which the formulas above differ from each other. This difference arises from the fact that the operator of composition is not associative.

The reader familiar with Pearl's causal diagrams [15] has certainly noticed an advantage of the introduced compositional models. While here we can compute both conditioning and intervention from one causal compositional model, in causal diagrams we have to consider two different graphs. Conditioning is computed from the "full" causal diagram, from which, for the computation of intervention, we have to delete all the arrows heading to the intervention variable.

3.1 Anticipating Operator

In the next section we will illustrate on our example from Section 1.1 the possibility to compute results of intervention and conditioning even in the case of model with a hidden variable. For this task, it will appear advantageous to compensate the lack of associativity by introducing another operator, so called *anticipating operator*, defining a special type of composition. For a set of variables z and distributions $\kappa(x)$ and $\mu(y)$ it is defined by the formula:

$$\kappa \circledtriangleright_z \mu = (\mu^{\downarrow(z\setminus x)\cap y} \cdot \kappa) \triangleright \mu.$$

Its advantage is expressed by Theorem 9.4 in [6] saying that for $\pi(z)$, $\kappa(x)$ and $\mu(y)$

$$\pi(z) \triangleright \kappa(x) \triangleright \mu(y) = \pi(z) \triangleright (\kappa(x) \circledtriangleright_z \mu(y)). \tag{9}$$

3.2 Elimination of Hidden Variables

Let us now derive formulas to be used for computation of conditioning and intervention in the simple causal compositional model describing our example from Section 1.1. Assuming it is in a perfect form we can express it with the help of its marginals:

$$\kappa(A,C,E,T) = \kappa(T) \triangleright \kappa(C,T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T).$$

Computation of the conditional $\kappa(E|C=\mathbf{a})$ is simple.

$$\begin{aligned}
\kappa(E|C=\mathbf{a}) &= \left(\delta_{\mathbf{a}}(C) \triangleright (\kappa(T) \triangleright \kappa(C,T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T))\right)^{\downarrow\{E\}} \\
&= \left(\left(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C,E,T)\right)^{-T}\right)^{\downarrow\{E\}} \\
&\overset{(2)}{=} \left(\delta_{\mathbf{a}}(C) \triangleright (\kappa(A,C,E,T))^{-T}\right)^{\downarrow\{E\}} = \left(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C,E)\right)^{\downarrow\{E\}}.
\end{aligned}$$

Notice that during these computations we used the property (2) from Section 2.1. This is why the symbol (2) appears above the respective equality sign. This type of explanation will also be used in the subsequent computations.

To compute $\kappa(E|do(C=\mathbf{a}))$ we will need to apply most of the properties introduced above, as well as the anticipating operator defined in the preceding section. Since we cannot expect the reader is accustomed to the computations with the operator of composition we will perform just one elementary modification at each of the following steps. This is why the following computations look even more cumbersome than they really are.

$$\begin{aligned}
\kappa(E|do(C=\mathbf{a})) &= \left(\delta_{\mathbf{a}}(C) \triangleright \kappa(T) \triangleright \kappa(C,T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T)\right)^{\downarrow\{E\}} \\
&\overset{(3)}{=} \left(\delta_{\mathbf{a}}(C) \triangleright \kappa(T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T)\right)^{\downarrow\{E\}} \\
&\overset{(6)}{=} \left(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C) \triangleright \kappa(T) \triangleright \kappa(A,E,T)\right)^{\downarrow\{E\}} \\
&\overset{(9)}{=} \left(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C) \triangleright \left(\kappa(T) \circledast_{\{C,A\}} \kappa(A,E,T)\right)\right)^{\downarrow\{E\}} \\
&\overset{(2)}{=} \left(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C) \triangleright \left(\kappa(T) \circledast_{\{C,A\}} \kappa(A,E,T)\right)^{-T}\right)^{\downarrow\{E\}}
\end{aligned}$$

To express $\left(\kappa(T) \circledast_{\{C,A\}} \kappa(A,E,T)\right)^{-T}$ we will take advantage of the idea of extension used for this purpose by Pearl in [15]. The reader can realize that it is the way how to take into account the mutual dependence of variables A, C, E (notice that it plays the same role of what is realized by inheritance of parents in Shachter's edge reversal rule [17]).

$$\begin{aligned}
&\left(\kappa(T) \circledast_{\{C,A\}} \kappa(A,E,T)\right)^{-T} = \left(\kappa(T) \circledast_{\{A\}} \kappa(A,E,T)\right)^{-T} \\
&= \left(\left(\kappa(C,T) \circledast_{\{A\}} \kappa(A,E,T)\right)^{-C}\right)^{-T} = \left(\kappa(A) \cdot \kappa(C,T) \triangleright \kappa(A,E,T)\right)^{\downarrow\{A,E\}} \\
&= \left(\kappa(A) \cdot \kappa(C) \triangleright \kappa(C,T) \triangleright \kappa(A,E,T)\right)^{\downarrow\{A,E\}} \\
&\overset{(3)}{=} \left(\kappa(A) \cdot \kappa(C) \triangleright \kappa(C,T) \triangleright \kappa(A,C) \triangleright \kappa(A,E,T)\right)^{\downarrow\{A,E\}} \\
&\overset{(7)}{=} \left(\kappa(A) \cdot \kappa(C) \triangleright (\kappa(C,T) \triangleright \kappa(A,C)) \triangleright \kappa(A,E,T)\right)^{\downarrow\{A,E\}} \\
&= \left(\kappa(A) \cdot \kappa(C) \triangleright \kappa(A,C,T) \triangleright \kappa(A,E,T)\right)^{\downarrow\{A,E\}} \\
&\overset{(8)}{=} \left(\kappa(A) \cdot \kappa(C) \triangleright (\kappa(A,C,T) \triangleright \kappa(A,E,T))\right)^{\downarrow\{A,E\}} \\
&= \left(\kappa(A) \cdot \kappa(C) \triangleright \kappa(A,C,E,T)\right)^{\downarrow\{A,E\}} \overset{(2)}{=} \left(\kappa(A) \cdot \kappa(C) \triangleright \kappa(A,C,E)\right)^{\downarrow\{A,E\}} \\
&= \left(\kappa(C) \circledast_{\{A\}} \kappa(A,C,E)\right)^{-C},
\end{aligned}$$

which eventually leads to

$$\kappa(E|do(C=\mathbf{a})) = \left(\delta_{\mathbf{a}}(C) \triangleright \kappa(A,C) \triangleright \left(\kappa(C) \circledast_{\{A\}} \kappa(A,C,E)\right)^{-C}\right)^{\downarrow\{E\}}.$$

3.3 Hypothetical Example - Numerical Computations

Let us illustrate the above-derived formulas on data corresponding to the example from Section 1.1. For the sake of simplicity, we consider all the variables to be binary. The data concerning 200 students from 16 universities are presented together with corresponding percentages as estimates of corresponding probabilities (see Table 1).

From Table 1 we can immediately see that the total probability of a success at the first attempt is $\kappa(E=1) = 0.68$.

Let us first consider the inapt model without the hidden variable T (recall that it assumes the conditional independence of C and E given A). To construct it, we need three distributions, namely a distribution for variable C, a distribution for variables A, C and the last one for variables A, E. Then we compute these distributions from Table 1, and denote them $\kappa_c(C), \kappa_{ac}(A,C), \kappa_{ae}(A,E)$, respectively. Constructing the compositional model corresponding to the graph from Figure 1(a)

$$\kappa(A,C,E) = \kappa_c(C) \triangleright \kappa_{ac}(A,C) \triangleright \kappa_{ae}(A,E) = \kappa_{ac}(A,C) \triangleright \kappa_{ae}(A,E)$$

we cannot expect to get a perfect model (the reader can see that it is really not perfect from the last row in Table 1), and therefore keep in mind that the estimated distributions κ_c, κ_{ac} and κ_{ae} are not marginals of κ.

Table 1: Estimates of probabilities

C	0	0	0	0	1	1	1	1
A	0	0	1	1	0	0	1	1
E	0	1	0	1	0	1	0	1
frequencies	18	42	22	38	12	12	12	44
percentages	0.09	0.21	0.11	0.19	0.06	0.06	0.06	0.22
$\kappa_{ace} = \kappa_{ac} \triangleright \kappa_{ae}$	0.11	0.19	0.9	0.21	0.04	0.08	0.08	0.20

Now, it is an easy task to compute

$$\kappa(E|do(C=1)) = \kappa(E|C=1) = ((\delta_1(C) \triangleright \kappa_{ac}(A,C)) \triangleright \kappa_{ae}(A,E))^{\downarrow\{E\}},$$

from which we immediately get

$$\kappa(E=0|do(C=1)) = \kappa(E=0|C=1) = 0.31,$$
$$\kappa(E=1|do(C=1)) = \kappa(E=1|C=1) = 0.69.$$

Therefore, as it was said in Introduction, this model expects a slight increase of success after the intervention making attendance at lectures compulsory at all universities.

Consider, now, a more complex model including the hidden variable T, i.e.,

$$\lambda(A,C,E,T) = \lambda_t(T) \triangleright \lambda_{ct}(C,T) \triangleright \lambda_{ac}(A,C) \triangleright \lambda_{aet}(A,E,T).$$

As said above, since the type of a university is for us a hidden variable, for this model we can estimate only one distribution from the data, namely $\lambda_{ac} = \kappa_{ac}$.

Fortunately, as it was shown in the preceding section, in this very model we can eliminate the hidden variable getting

$$\lambda(E|C=1) = \left(\delta_1(C) \triangleright \lambda_{ace}(A,C,E)\right)^{\downarrow\{E\}},$$

and

$$\lambda(E|do(C=1)) = \left(\delta_1(C) \triangleright \lambda_{a,c}(A,C) \triangleright \left(\lambda_c(C) \circledcirc_{\{A\}} \lambda_{a,c,e}(A,C,E)\right)^{-C}\right)^{\downarrow\{E\}},$$

from which we get

$$\lambda(E=0|C=1) = 0.30,$$
$$\lambda(E=1|C=1) = 0.70,$$
$$\lambda(E=0|do(C=1)) = 0.33,$$
$$\lambda(E=1|do(C=1)) = 0.67.$$

4 Model verification

When discussing the hypothetical example, in Fig. 1 we introduced two possible causal models that were denoted κ and λ in Section 3.3. Each of them yielded different results for the intervention $do(C=1)$. In this section we are going to show that based on the data from Section 3.3 we should not have taken the simpler model from Fig. 1(a) (i.e., model κ) into consideration.

When setting up a model, one should always consider, and verify, all (or at least most of) the conditional independence relations that can be read from the model. For this, several tools were described in the literature. When the model is in a form of a Bayesian network, one can use either a *d-separation rule* or the rule based on separation in a *moralized ancestral graph* [11]. For the compositional models, *persegrams* were designed [10], from which one can read all the independence relations that must hold in the considered model.

Table 2: Conditional frequencies of C and E for $A = 0$ (left) and $A = 1$ (right)

	E=0	E=1
C=0	18	42
C=1	12	12

	E=0	E=1
C=0	22	38
C=1	12	44

We do not describe here the mentioned procedures. Just let us recall how the conditional independence relation is defined. Consider a distribution $\pi(x)$, and variables $X, Y \in x$, and a subset (possibly empty) of variables $y \subseteq x \setminus \{X, Y\}$. We say that for distribution π variables X and Y are *conditionally independent given set of variables* y (in symbol $X \perp\!\!\!\perp_\pi Y|x$) if

$$\pi^{\downarrow y \cup \{X,Y\}} \cdot \pi^{\downarrow y} = \pi^{\downarrow y \cup \{X\}} \cdot \pi^{\downarrow y \cup \{Y\}}.$$

Either of the above-mentioned tools yields the following systems of conditional independence relations:

- for model κ described in Fig. 1(a) it is just one relation: $C \perp\!\!\!\perp_\kappa E|A$;
- for model λ described in Fig. 1(b) there are two relations: $A \perp\!\!\!\perp_\lambda T|C$, and $E \perp\!\!\!\perp_\lambda C|\{A, T\}$.

In the latter model, both the independence relations contain the unobserved variable T, so we do not have a possibility to verify (or, better said, to exclude) this model.

But, having data from Table 1 we can test whether the conditional independence $C \perp\!\!\!\perp E|A$ holds true, which is necessary for model κ. The test of this relation can be performed using well-known Cochran-Mantel-Haenszel test or in a simplified manner using the sum of partial chi-squares for two conditional probabilities in Table 2. Now we get value of chi-square for $A = 0$ equal to $\chi_0^2 = 2.987$ and for $A = 1$ equal to $\chi_1^2 = 3.246$ both with one degree of freedom. Therefore we yield summary test statistics $\chi^2 = 6.233$ which is with two degrees of freedom above the five percent critical value, now equal to 5.99. And thus we should had rejected the null hypothesis of conditional independence of C and E given A, and should not had considered this causal model.

5 Conclusions

On a slightly paradoxical example we tried to convince the reader that considering causal influence in business process models may be not only useful but sometimes inevitable, and that having such a causal model one should do their best to verify the model (or more precisely, to test whether the model is not excluded by the data). The necessity of having a simple example led us to illustrate the proposed approach on artificially generated data. However, we tried to generate realistically looking (non-degenerated) data and this is why the numerical results may seem not so convincing. Naturally, when more mutually dependent effects are taken into consideration, or when considering a more complex model with a greater number of variables, the results obtained from descriptive and causal models can be substantially different.

Another purpose of this paper was to show that causal models need not be only causal diagrams. The employment of causal compositional models can be used to describe causal relations among considered variables. The composition of model distributions with a special (degenerated) distribution surprisingly provides a possibility to evaluate both conditional probabilities and the effects of causal interventions using a unifying apparatus of compositional models.

Acknowledgement
The research was partially supported by GA ČR under the project number 15-00215S. The second author wants to express his sincere thanks to Judea Pearl for the motivation to start employing causal models in applications.

References

[1] V. Bína, R. Jiroušek. Marginalization in Multidimensional Compositional Models. *Kybernetika,* 42, 4, 405–422, 2006.

[2] V. Bína, R. Jiroušek. On computations with causal compositional models. *Kybernetika,* 51, 3, 525–539, 2015.

[3] A. Feugas, S. Mosser, L. Duchien. A Causal Model to predict the Effect of Business Process Evolution on Quality of Service. In *Proceedings of the 9th International ACM Sigsoft Conference on the Quality of Software Architectures,* pages 143–152. ACM, New York, 2013.

[4] Y. Hagmayer, S. A. Sloman, D. A. Lagnado, M. R. Waldmann. Causal reasoning through intervention. In *Causal learning: Psychology, philosophy, and computations,* Alison Gopnik and Laura Schulz (eds.) Oxford University Press, Oxford, New York, 2007.

[5] R. Jiroušek. Composition of Probability Measures on Finite Spaces. In *13th Conf. on Uncertainty in Artificial Intelligence,* pages 274–281. Morgan Kaufman, San Francisco, 1997.

[6] R. Jiroušek. Foundations of compositional model theory. *Int. J. General Systems,* 40, 6, 623–678, 2011.

[7] R. Jiroušek. On Causal Compositional models: simple examples. In *Proc. of the 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems,* A. Laurent et al. (eds.), Part I, pages 517–526, Springer International Publishing, CCIS 442, Switzerland, 2014.

[8] R. Jiroušek, V. Kratochvíl. Foundations of compositional models: structural properties. *Int. J. General Systems* 44, 1, 2–25, 2015.

[9] C. W. Kirgwood. *System Dynamics Methods: A Quick Introduction.* Arizona State University, 2013. [http://www.public.asu.edu/~kirkwood/sysdyn/SDIntro/SDIntro.htm]

[10] V. Kratochvíl. Probabilistic Compositional Models: solution of an equivalence problem. *Int. J. Approximate Reasoning,* 54, 5, 590–601, 2013.

[11] S. L. Lauritzen. *Graphical Models.* Oxford University Press, 1996.

[12] F. M. Malvestuto. Marginalization in Models Generated by Compositional Expressions. *Kybernetika* 51, 4, 541–570, 2015.

[13] S. Morgan, C. Winship. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, Cambridge University Press, New York, 2007.

[14] R. J. Murnane, J. B. Willett. *Methods Matter: Improving Causal Inference in Educational and Social Science Research.* Oxford University Press, Oxford, 2011.

[15] J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, New York, 2009.

[16] M. Ryall. Causal Ambiguity, Complexity, and Capability-Based Advantage, *Management Science,* 55, (3), 389–403, 2009.

[17] R. D. Shachter. Evaluating Influence Diagrams. *Operations Res.,* 34, 871–882, 1986.

[18] L. Váchová, V. Bína. A Note on Computation of Causal Effect in Economic Models. In *Proceedings of the 31st Mathematical models in Economy,* pages 974–979, College of Polytechnics, Jihlava, 2013.