

Causal Bayes Nets and Token-Causation

Closing the Gap between Token-Level and Type-Level¹

Alexander Gebharter * Andreas Hüttemann

Abstract: Causal Bayes nets (CBNs) provide one of the most powerful tools for modelling coarse-grained type-level causal structure. As in other fields (e.g., thermodynamics) the question arises how such coarse-grained characterisations are related to the characterisation of their underlying structure (in this case: token-level causal relations). Answering this question meets what is called a “coherence-requirement” in the reduction debate: How are different accounts of one and the same system (or kind of system) related to each other. We argue that CBNs as tools for type-level causal inference are abstract enough to roughly fit any current token-level theory of causation as long as certain modelling assumptions are satisfied, but accounts of actual causation, i.e. accounts that attempt to infer token-causation based on CBNs, for the very same reason, face certain limitations.

Key words: causal inference, token-level causation, type-level causation, causal Bayes nets, actual causation

0 Conflict of interest statement

¹ This is a draft paper. Please do not cite or quote without permission. The final version of this paper is published under the following bibliographical data: Gebharter, A. & Hüttemann, A. (2023). Causal Bayes nets and token-causation: Closing the gap between token-level and type-level. *Erkenntnis*. Advance online publication. doi:10.1007/s10670-023-00684-5

The authors declare that there is no conflicts of interest.

1 Introduction

Causal Bayes nets (CBNs) provide one of the most powerful tools for causal inference on the market. They are useful for modelling coarse-grained type-level causal structure. As in other cases of coarse-grained characterisations (e.g., in thermodynamics) the question arises how the coarse-grained characterisation is related to the characterisation of the underlying structure (in this case: token-level causal relations). Answering this question on the one hand meets what has been called a “coherence-requirement” in the reduction debate: How are different accounts of one and the same system (or kind of system) related to each other.² In this paper, we take these issues as a starting point and tackle the question of how type-level and token-level causal structure are related. In particular, we look at how different accounts of token-causation may provide a suitable basis for CBNs. We argue that the type-level structures captured by CBNs can be generated by abstracting from token-level details. As a result, the type-level structure is robust vis-à-vis the details of the token-level as well as relative to the token-level account of causation. This stability licenses, on the one hand, CBNs as a powerful tool for type-level causal reasoning. On the other hand, the difficulties of accounts of actual causation (which we take to be accounts that attempt to infer token-level structure starting from CBNs) can be explained as well: Type-level structures generated by abstraction leave out many details that are relevant for which token-level causal relations obtain.

² Work on this paper started as a discussion of a worry raised by Cartwright (2007): Any methodological framework for causation needs to be legitimised by its underlying metaphysics. However, the resulting paper is largely independent of Cartwright’s particular framing. What we are interested in is how type-level causal accounts are related to token-level accounts.

The paper is structured as follows. In section 2 we briefly remind the reader how token-level accounts of causation differ with respect to their verdicts on two paradigmatic cases: late preemption and double-prevention. In section 3 we outline the main features of CBNs. Section 4 is devoted to elaborating how token-level causal facts give rise to type-level representations in terms of CBNs. In section 5 we argue that type-level structure is fairly robust with respect to the underlying details of token-level accounts. Once we have understood the relation between token-causation and CBNs, we are in a position to diagnose why accounts of actual causation face serious difficulties in section 6.

2 Token-causation

The purpose of this section is to briefly indicate how what is classified as a token-level causal relation differs depending on the underlying theory of token-causation. We will look at three token-level theories and two paradigmatic scenarios that are often discussed in the literature. The three accounts of token-causation shall serve as proxies for illustration, but we believe that the points we make roughly hold for all token-level theories currently on the market.

2.1 Throwing stones

Suzy and Billy both throw stones at a vase; Suzy's stone hits first and shatters the vase. Since Billy's stone arrives too late, it does not destroy the vase. This is a classical case of late pre-emption.

According to the simple *counterfactual account*, causation is the ancestral of counterfactual dependence between distinct (i.e., non-overlapping) events. e counterfactually depends on c iff had c not occurred, e would not have occurred either. c is then defined as a cause of e iff there is a (possibly empty) set of events d_1, \dots, d_n such that e counterfactually depends on d_n , each d_i (with $1 < i \leq n$) counterfactually depends

on d_{i-1} , and d_1 counterfactually depends on c (Lewis, 1973). As a consequence, on this account, neither Suzy's throw nor Billy's throw will be classified as a cause. Had Suzy not thrown, Billy's stone would have shattered the vase (and vice versa).

On the *conserved quantity account* causation holds in virtue of the fact that cause and effect are related by causal processes. Causal processes are characterised as world lines of an object that possess a conserved quantity such as energy and momentum (Dowe, 1995, 2007). The conserved quantity theory will identify Suzy's throw as a cause, but not Billy's. There was a causal process between Suzy's throw and the shattering, but not between Billy's and the shattering.

Finally, we will look at a third account, which captures aspects of both, production and dependence views and yields yet a different set of verdicts on our paradigmatic scenarios: the *disruption account* (Hüttemann, 2020). This account allows for several concepts of causation which, however, need to be linked, for example by the concept of a quasi-inertial process.

The first concept we are interested in is disruption causation:

(DC) A cause is something that accounts for an effect, where the effect consists of a system being in a state Z' at a certain time t that is different from or deviant relative to the quasi-inertial behaviour of the system at that time, i.e., relative to the state Z (where $Z \neq Z'$) that the system would have developed into in the absence of interferences. A cause is something that brought about this deviation.

(DC) gives us a contrastive notion of causation because the cause is a cause for e being in state Z' rather than in state Z .

While **(DC)** is the concept of causation that tracks most of our causal intuitions, another concept, closed system causation, evolved in 19th century physics (Hüttemann, 2020):

(CS) In a closed system, i.e., in a quasi-inertial system that is not interfered with, an earlier state of the system is the cause of any later state of the system.

The disruption account can be understood as the conjunction of **(CS)** and **(DC)**. Much needs to be said about the terms involved in **(CS)** and **(DC)**. For reasons of space, a few remarks must suffice. The notion of a quasi-inertial process refers to the temporally extended behaviour of a system that is not interfered with. The notion of an interference needs to be spelt out in terms of the sciences that are pertinent for the processes in question. Newton's first law, for example, describes the (quasi-)inertial behaviour of a massive particle and Galileo's law of free fall defines a quasi-inertial process: A free falling object in a vacuum displays a certain behaviour (as long as no interfering factors intervene). Newton's second law specifies possible interfering factors ("impressed forces").

The identification of quasi-inertial processes is relative to a prior identification or specification of the systems. When we are interested in causal explanation, it is often the contrast to the effect to be explained that determines the relevant quasi-inertial process. (This also holds in the case of preventions: When we are looking for a cause of why a dam did not break, for example, the relevant quasi-inertial process to which we compare candidate causes is a process that results in the breaking of the dam.) This is captured by **(DC)**. If in the absence of interferences we ask for a cause of a state of a system, on the other side, we rely on **(CS)** instead.

Let us finally come back to our example: According to the disruption account, Suzy's throw causes the shattering. Here is why: The effect to be explained is the shattering of the vase. The quasi-inertial behaviour (no interfering factors) of the vase would have led to the vase being intact. But an event (Suzy throwing the stone) obtained which led to an interference with the quasi-inertial process such that the vase is shattered instead of

remaining intact. By contrast, Billy's stone does not interfere with the vase. (For details, see Hüttemann, 2020.)

2.2 Bombing mission

We take this example directly from Hall (2004):

“Suzy is piloting a bomber on a mission to blow up an enemy target, and Billy is piloting a fighter as her lone escort. Along comes an enemy fighter plane, piloted by Enemy. Sharp-eyed Billy spots Enemy, zooms in, pulls the trigger, and Enemy's plane goes down in flames. Suzy's mission is undisturbed, and the bombing takes place as planned.” (ibid., p. 241)

This is a classical case of double-prevention. The counterfactual account classifies Billy's pulling the trigger as a cause. Had he not been on his mission, Suzy would not have succeeded in destroying the target. The conserved quantity theory cannot classify it as a cause because of the lack of a physical connection (Hall, 2007). The disruption account, on the other hand, will classify Billy's pulling the trigger as a cause because it disrupts the process of Enemy bringing down Suzy's plane (Hüttemann, 2020).

The verdicts listed are not meant to register the shortcomings of any of the accounts, but rather to point to the fact that they differ. They are clearly different accounts of token-causation — not only with respect to their claims what causation consists in, but also extensionally, i.e., with respect to the classification of the above cases.

3 Causal Bayes nets

Causal Bayes nets (CBNs) as tools for causal inference can, for example, be used to uncover causal structure on the basis of observational and experimental data. They can

also generate causal predictions based on observation and predict the outcomes of hypothetical interventions even in cases where no experimental data is available. Formally, a CBN is a triple $\langle \mathbf{V}, \mathbf{E}, Pr \rangle$, where \mathbf{V} is a set of random variables X_1, \dots, X_n representing the properties or events whose causal connections one wants to model, \mathbf{E} is a set of edges connecting pairs of variables in \mathbf{V} interpreted as direct causal dependence w.r.t. \mathbf{V} , and Pr is a probability distribution over \mathbf{V} . A CBN's graph $\mathbf{G} = \langle \mathbf{V}, \mathbf{E} \rangle$ is assumed to be directed and acyclic, meaning that all the edges in \mathbf{E} are directed edges (\longrightarrow) and the graph does not feature a path of the form $X_i \longrightarrow \dots \longrightarrow X_i$. CBNs are assumed to satisfy the causal Markov condition (Spirtes et al., 1993, p. 54):

Causal Markov condition: $\langle \mathbf{V}, \mathbf{E}, Pr \rangle$ satisfies the causal Markov condition iff every $X_i \in \mathbf{V}$ is probabilistically independent of its non-descendants conditional on its parents.

A variable X_i 's parents (their set is denoted as $\mathbf{Par}(X_i)$) are its direct causes w.r.t. \mathbf{V} , i.e., all the variables $X_j \in \mathbf{V}$ such that $X_j \longrightarrow X_i$. A variable X_i 's descendants are X_i itself and all the variables $X_j \in \mathbf{V}$ such that $X_i \longrightarrow \dots \longrightarrow X_j$. Finally, a variable X_i 's non-descendants are all the variables $X_j \in \mathbf{V}$ that are not descendants of X_i . For directed acyclic graphs, the causal Markov condition is equivalent to the factorisation

$$Pr(x_1, \dots, x_n) = \prod_{i=1}^n Pr(x_i | \mathbf{par}(X_i)),$$

where $\mathbf{par}(X_i)$ stands for the instantiation of the variables in $\mathbf{Par}(X_i)$ to their values on the left-hand side of the equation.

Another relevant condition is the causal minimality condition. It can be seen as a kind of difference-making requirement for causal relations, saying that each directed edge

connecting two variables produces a probabilistic dependence between these variables in some circumstances (Schurz & Gebharder, 2016). Formally, it can be stated as follows (Spirtes et al., 1993, p. 55):

Causal minimality condition: $\langle \mathbf{V}, \mathbf{E}, Pr \rangle$ satisfies the causal minimality condition iff it satisfies the causal Markov condition and there is no subgraph \mathbf{G}' of $\mathbf{G} = \langle \mathbf{V}, \mathbf{E} \rangle$ that satisfies the causal Markov condition.

Every graph \mathbf{G}' one gets from \mathbf{G} by deleting one or more edges is a subgraph of \mathbf{G} .

CBNs can be used to model the effects of interventions. An intervention sets a variable $X_i \in \mathbf{V}$ to a specific value x_i independently of X_i 's other causes in \mathbf{V} . The effects of an intervention that sets X_i to x_i can be computed by deleting all the arrows in the graph pointing at X_i and setting the probability of $X_i = x_i$ to 1. Next, one applies the Markov factorisation to the truncated graph to compute post intervention probabilities.

If the model's probability distribution is nice enough (see Spirtes et al., 2000, sec. 2.3.5 for details), the conditional probabilities $Pr(X_i | \mathbf{Par}(X_i))$ can be represented by structural equations of the form $X_i =_c f_i(\mathbf{V}', U_i)$ (where $\mathbf{V}' \subseteq \mathbf{V}$). Variables U_i are sometimes called error terms. They are introduced in order to model probabilistic causal systems. Interventions can be represented in structural equation models as in ordinary CBNs: The effects of an intervention that sets X_i to x_i can be computed by first replacing the original equation $X_i =_c f_i(\mathbf{V}', U_i)$ by $X_i = x_i$, and then using $X_i = x_i$ together with the original structural equations for all the other variables different from X_i in order to compute the value of any variable one is interested in.

Finally, CBNs and structural equations as introduced are first and foremost tools for modelling type-level causation. An arrow $X_i \longrightarrow X_j$ stands for the type-level causal claim

that the variable X_i (or the property it represents) is directly causally relevant for the variable X_j (or the property it represents) w.r.t. the set of variables \mathbf{V} under consideration. X_i could, for example, stand for *smoking behaviour* and X_j for *lung cancer*. Though the model can be used to make predictions about particular values, it does not tell us which particular X_i -value is a token-level cause of which particular X_j -value. It does also not tell us anything about whether a particular individual will get lung cancer as a result of her specific smoking behaviour. The model only makes predictions at the population level, for example, that in any subpopulation whose individuals all share this and that smoking behaviour the relative frequency of individuals developing lung cancer will be such and such high, or that changing the smoking behaviour in this and that way in a population (this would amount to an intervention) would lead to such and such changes in the lung cancer rate. To infer whether a token-level causal relation holds for specific individuals, additional assumptions are required. Depending on which specific assumptions one makes, one gets different accounts of *actual causation*. By an account of actual causation we always mean an account that relies on such additional assumptions in order to squeeze out token-level causal information based on type-level causal structure as represented by a CBN or structural equation model. Thus, we do not treat CBN-based accounts of actual causation as yet another theory of token-causation, but as an inference tool aiming at identifying token-level causal relations.

4 Closing the gap between type-level and token-level

In this section we make a proposal how the gap between CBNs as tools for modelling causal structure at the type-level and token-causation as characterised by theories like the

ones sketched in section 2 can be closed.³ In particular, we will ask how facts about token-level causation give rise to type-level causal structures in terms of CBNs.

First of all, note that we can formulate any token-level causal theory in terms of random variables. We assume that the different states systems s or their subsystems can be in are represented by the different values of random variables. Let X_1, \dots, X_n be the variables describing all the possible states of those systems in whose causal connections we are interested in and \mathbf{V} be the set of these variables. For illustration, let us take a brief look at a specific system s . Assume that this system is a game of billiard. Now let c, e be subsystems of s . Let c be the white and e be the black ball, both lying somewhere on the billiard table. Let $X_i(c)$ describe the behaviour of the white ball and $X_j(e)$ the behaviour of the black ball. Now assume that $X_i(c) = x_i$ describes a specific way how the white ball moves across the billiard table towards the black ball and let $X_j(e) = x_j$ describe the black ball falling into a specific pocket.⁴ We now assume that this specific movement of the white ball causes the black ball to fall into this pocket, i.e., that $X_i(c) = x_i$ causes $X_j(e) = x_j$. It can easily be checked that all three token-level theories introduced in section 2 support this result.

Now the question is how we can get a type-level claim $X_i \longrightarrow X_j$ in terms of CBNs out of token-level facts such as $X_i(c) = x_i$ causing $X_j(e) = x_j$. Before we present our answer to this question, let us introduce one more concept as well as a restriction in scope. Let

³ Recall that we pursue the project of investigating of how token-level accounts of causation and type-level accounts are related. We are not interested in uncovering causal structure from empirical data. While our project is about connecting the token with the type level, the latter project only concerns type-level to type-level inference. What one typically does is to use empirical data about how features are distributed among the individuals of a representative sample to infer the causal structure underlying the overall population.

⁴ We will ignore individual variables like c, e in expressions such as $X_i(c) = x_i$ and $X_j(e) = x_j$ when it is clear from the context to which individuals these expressions apply.

B_1, \dots, B_m be variables describing relevant background factors or initial conditions and \mathbf{B} be the set of these variables. In the following, we will assume that these background factors are fixed for all systems s under consideration,⁵ while the variables X_1, \dots, X_n are allowed to have different values in different systems s . In particular, we want to end up with a CBN for the variables X_1, \dots, X_n to the background of the fixed conditions $\mathbf{B} = \mathbf{b}$.

Finally, in order to get type-level causal structure out from token-level causal relations, we will quantify over all nomologically possible causal system. We will restrict the scope to those nomologically possible systems (i) whose token-level causal structures are acyclic and have tokens of the same types and (ii) whose causal relations can be disrupted. (i) implies that if $X_i = x_i$ is a token-level cause of $X_j = x_j$ in some system, then there is no system in which $X_j = x_j$ is a token-level cause of $X_i = x_i$. (ii) means that if $X_i = x_i$ is a token-level cause of $X_j = x_j$ in one system, then there is some other system in which $X_i = x_i$ is not a token-level cause of $X_j = x_j$.

Here is our proposal for grounding type-level structure in token-level causal facts. It consists of two steps.

Step 1: Identification. We take it to be a major constraint for an account of the relation between token-level and type-level causal structure to satisfy the following condition: If there is a value of a variable X_i which happens to be causally relevant for the value of a variable X_j at the token-level in some system under some circumstances, then this is already enough for X_i to be causally relevant for X_j at the type-level, and vice versa: We only want X_i to end up as a direct type-level cause of X_j if at least some X_i -value is a

⁵ We will keep this assumption in mind, but most of the time not mention the background factors explicitly.

token-level cause of some X_j -value in some system.⁶ This is the weakest possible connection between type- and token-causation we can think of. But as we will see shortly, it is already enough to do the job. Here we are clearly transcending the empirical approach a social scientist would pursue. While the latter will focus on actual occurrences and frequencies, we are interested in how — in general — type-level structure is related to token-level structure. Thus, when quantifying over systems, we always have the set of all nomologically possible systems in mind.

The goal of step 1 is to identify all the type-level causal relations among the variables of interest. In a bit more detail, what we have to do is to look at all nomologically possible systems s featuring subsystems c, e sharing the same background conditions $\mathbf{B} = \mathbf{b}$. We will find that the values of the variables X_1, \dots, X_n are differently distributed in different such systems. For each variable $X_j \in \mathbf{V}$ we now check whether X_j has a value x_j for which at least one value x_i of one of the other variables $X_i \in \mathbf{V}$ is a token-level cause of x_j in at least one of the systems s . If the answer to this question is yes, then X_i is an element of $\mathbf{C}(X_j)$. After applying this test to all variables $X_i \in \mathbf{V}$ different from X_j , we have determined the set $\mathbf{C}(X_j)$ of X_j 's type-level causes.

Let us briefly illustrate the first step of the procedure by a toy example. Assume we are interested in scenarios s involving a general, a hitman, a backup for the hitman, and the task of destroying a facility by detonating a bomb. Let us describe these possible events with the variables X_1, \dots, X_5 . In most of these systems s , the general does not order anyone to destroy the facility ($X_1 = 0$). But in some systems, she orders the hitman to

⁶ Note the restriction to “direct” type-level causation in the vice versa part. It stems from the fact that since there can be causal chains in CBNs such that the effect does not depend on the cause (such chains constitute a failure of causal faithfulness, cf. Spirtes et al., 2000, p. 31), we cannot assume that for each such type-level chain there must be a token-level causal connection. Thus, only direct type-level causal connections need to be backed up by corresponding token-level causal relations.

destroy the facility and the backup to blow up the bomb should the hitman fail ($X_1 = 1$). In most of these systems, the hitman presses the button ($X_2 = 1$), but in some she fails to do so ($X_2 = 0$).⁷ In most of the scenarios in which the hitman presses the button, the backup does nothing ($X_3 = 0$), and in most of the scenarios where the hitman fails to press the button, the backup presses the button instead ($X_3 = 1$). In most of the systems where at least one of the two presses the button, the bomb detonates ($X_4 = 1$), and in most systems in which the bomb detonates, the facility gets destroyed ($X_5 = 1$).

We have characterised the events that take place, but have not yet provided any information about what causes what. Accounts of token-causation allow us to add this type of information. Let us have a look at these dependencies through the lens of the simple counterfactual account, which we will use as a proxy for illustrating the abstraction procedure described in this section.⁸ What we need to figure out is whether among the systems mentioned above, there is one such that, say, $X_2 = 1$ counterfactually depends on $X_1 = 1$, etc. If thought through, we arrive at the conclusion that, according to the counterfactual account, in some systems $X_1 = 1$ is a cause of at least one of the events $X_2 = 1$, $X_3 = 1$, $X_4 = 1$, $X_5 = 1$, that in some $X_2 = 1$ is a cause of at least one of the events $X_4 = 1$, $X_5 = 1$ while in some $X_2 = 0$ is a cause of at least one of the events $X_3 = 1$, $X_4 = 1$, $X_5 = 1$, that in some $X_3 = 1$ is a cause of at least one of the events $X_4 = 1$, $X_5 = 1$, and that in some $X_4 = 1$ is a cause of $X_5 = 1$.

Using the method outlined above, we can now form the set $\mathbf{C}(X_j)$ for each of the variables X_1, \dots, X_5 . A variable X_i will be an element of this set if one of its values is a

⁷ We assume that every token-level cause's effect can, in principle, be prevented if another causal factor would have interfered in an appropriate way. Thus, the hitman fails to press the button in some of the systems s .

⁸ In section 5 we will argue that the abstraction procedure introduced above is rather robust in generating type-level causal models, meaning that different token-level accounts typically result in largely the same type-level representation.

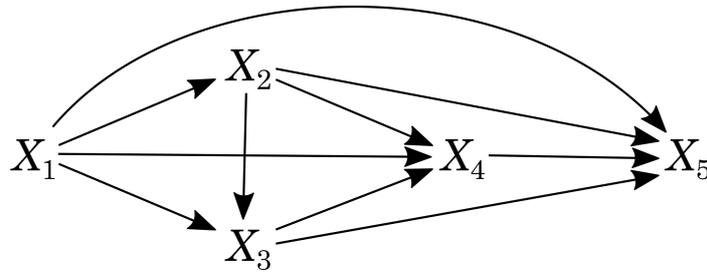


Figure 1

token-cause of one of X_j 's values in at least one system s . Thus, we end up with the following sets:

$$\mathbf{C}(X_1) = \emptyset$$

$$\mathbf{C}(X_2) = \{X_1\}$$

$$\mathbf{C}(X_3) = \{X_1, X_2\}$$

$$\mathbf{C}(X_4) = \{X_1, X_2, X_3\}$$

$$\mathbf{C}(X_5) = \{X_1, X_2, X_3, X_4\}$$

From this we get the structure depicted in Figure 1. Note that the arrows do not stand for direct type-level causal dependence (as is usually the case in such figures), but for direct *or* indirect type-level causal dependence.

Step 2: Ordering. We do know each variable X_j 's type-level causes now, but we do not yet know their order. In other words, we do not yet know which of these type-level causes are the *direct* causes of X_j w.r.t. \mathbf{V} . We solve this problem in the second step. To this end, we first need to identify the conditional probabilities $Pr(X_j = x_j | \mathbf{C}(X_j) = \mathbf{c}(X_j))$. To do so, we need to have a look at all those systems s instantiating $\mathbf{C}(X_j) = \mathbf{c}(X_j)$ and determine the relative frequency of these systems in which $X_j = x_j$ is instantiated too. We identify the conditional probability $Pr(X_j = x_j | \mathbf{C}(X_j) = \mathbf{c}(X_j))$ with this frequency. Next,

we can identify X_j 's direct type-level causes by looking for the narrowest subset of $\mathbf{C}(X_j)$ that screens X_j off from all the variables in $\mathbf{C}(X_j)$. The relevant probabilities are determined as before. The elements of the narrowest screening off set will be the set of X_j 's direct type-level causes or causal parents $\mathbf{Par}(X_j)$.⁹ If we repeat the procedure outlined for every variable $X_j \in \mathbf{V}$, we can identify a directed acyclic graph $\mathbf{G} = \langle \mathbf{V}, \mathbf{E} \rangle$.

Let us illustrate how this second step works. We can look at all the systems s and determine the probabilities $Pr(X_j = x_j | \mathbf{R} = \mathbf{r})$ (where $\mathbf{R} \subseteq \mathbf{C}(X_j)$) as the ratio of the systems s in which $X_j = x_j, \mathbf{R} = \mathbf{r}$ relative to the systems s in which $\mathbf{R} = \mathbf{r}$. Based on these probabilities, we can now identify the causal parents of every variable X_j with the narrowest $\mathbf{R} \subseteq \mathbf{C}(X_j)$ that will screen X_j off from $\mathbf{C}(X_j)$. For X_1 this is trivially the empty set and for X_2 it is trivially $\{X_1\}$. For X_3 , it is $\{X_1, X_2\}$. Neither X_1 nor X_2 alone can do the job. If $X_2 = 1$, it might still happen that the backup believes the hitman did not press the button and, because of that, presses the button herself $X_3 = 1$. This becomes more likely, however, if the general gave the order ($X_1 = 1$). Otherwise, there would be no reason to press the button at all. So X_2 does not screen X_3 off from X_1 . But also X_1 does not screen X_3 off from X_2 . Granted, the probability that the backup presses the button is higher if the general gives the order. However, it still makes a huge difference whether the hitman presses the button or not.

For X_4 , the causal parents are X_2 and X_3 , but not X_1 . $\{X_2, X_3\}$ will screen X_4 off from X_1 because once it is decided what the hitman and the backup do (pressing vs. not pressing the button), whether the general gave the order becomes irrelevant for whether the bomb detonates. Again, neither X_2 nor X_3 can screen off X_4 from all the other members of $\mathbf{C}(X_4)$. If we know, for example, that the hitman pressed the button, then additional information about whether the backup did so still has an influence on the

⁹ Note that in some cases there might not be a unique screening off set. We will come back to this issue in due course.

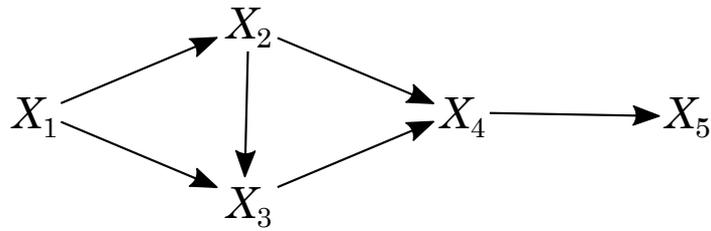


Figure 2

probability of the detonation. No cause is perfect; the effect might not occur due to other interfering factors. So the backup also pressing the button will further increase the probability of the bomb detonating. The same goes the other way round. Similar considerations can be made regarding X_5 . If thought through, we arrive at the following sets of type-level causal parents:

$$\mathbf{Par}(X_1) = \emptyset$$

$$\mathbf{Par}(X_2) = \{X_1\}$$

$$\mathbf{Par}(X_3) = \{X_1, X_2\}$$

$$\mathbf{Par}(X_4) = \{X_2, X_3\}$$

$$\mathbf{Par}(X_5) = \{X_4\}$$

Based on this information, we can finally identify the type-level causal structure as the one depicted in Figure 2.

Constraints

Let us briefly take a step back and have a look at the broader picture. Our aim is to understand how type-level structure as captured by CBNs depends on token-level causal structure. We proceeded by building graphs that can be used for type-level causal inference on the basis of facts about token-level causation so as to end up with a graph such that for each directed edge $X_i \longrightarrow X_j$ there are systems s such that a value x_i of X_i

and a value x_j of X_j are instantiated such that $X_i = x_i$ is a token-level cause of $X_j = x_j$. This is important to answer our initial question: How are CBNs related to the underlying token-level structure? We propose the following answer: If certain additional constraints are satisfied, then this will guarantee the intended fit. These constraints cannot be justified by facts about token-level causation, but rather tell us under which specific conditions a type-level structure that fits to the token-level causal structure according to the two-step procedure can be guaranteed to represent correctly. It is thus not the case that token-level structures on their own uniquely determine CBNs. Let us briefly walk through these constraints.

Ordering uniqueness. Step 2 of the method to get type-level structure out of token-level causation required us to identify the smallest subset of each variable's type-level causes that screens that variable off from all its type-level causes. Next, we identified this set with the variable's type-level causal parents. For the procedure to work it is required that there is only one such screening off set. Sometimes, however, there is no unique screening off set. This, for example, can happen if dependence patterns feature too many deterministic dependencies. Assume, for example, we are interested in the binary variables X_1, X_2, X_3 and that in all nomologically possible systems $X_1 = 1$ is a direct cause of $X_2 = 1$ and $X_3 = 1$, $X_1 = 0$ of $X_2 = 0$ and $X_3 = 0$, $X_2 = 1$ of $X_3 = 1$, and $X_2 = 0$ of $X_3 = 0$. Now step 1 of our procedure would tell us that both X_1 and X_2 are type-level causes of X_3 , but since both $\{X_1\}$ as well as $\{X_2\}$ screen X_3 off from all its type-level causes, step 2 does not give us a unique answer to the question of whether X_1 or X_2 or both are direct causes of X_3 .¹⁰ For these reasons, we need to assume ordering uniqueness: The values of the variables of our model need to be distributed among all nomologically possible systems s

¹⁰ X and Y are screened off by Z iff $Pr(y|x, z) = Pr(y|z)$ or $Pr(x, z) = 0$.

in such a way that the screening off part of step 2 leads to a unique result. Otherwise the type-level structure is underdetermined.

Markov distribution. Another requirement for step 2 is that a type-level effect's direct type-level causes screen it off from its indirect type-level causes. This on its own does not give us the causal Markov condition — which causally interpreted CBNs are assumed to satisfy by definition — introduced in section 3. It is rather a logically weaker consequence of the causal Markov condition. For this reason it cannot yet be guaranteed that the structure resulting from our procedure is in fact a CBN suitable for type-level reasoning. For example, nothing in our procedure guarantees that type-level common causes will screen off their effects, which is another logically weaker consequence of the Markov condition. Another such consequence is that causally unconnected variables are independent. Consequences like these, however, are essential in type-level reasoning with CBNs. Hence, we furthermore need to assume that the values of the variables to be modelled are distributed among systems s in a Markovian way, i.e., in such a way that the type-level structure resulting from step 2 satisfies the causal Markov condition.¹¹

Difference-making. Another problem can occur if causal differences at the token level do not show up anymore at the type level. Here is an example. Assume we are interested in the two variables X_1, X_2 . Assume further that in some nomologically possible systems s , some X_1 -values are token-level causes of some X_2 -values. Finally, assume that the two variables' values are distributed among all systems s in such a way that, when determining the conditional probabilities in step 2, we end up with the two variables being independent

¹¹ Whether all common causes satisfy the causal Markov condition has been contested (e.g., Cartwright, 1999). Such objections can be countered in many ways (e.g., Glymour, 1999; Gebharder & Retzlaff, 2020; Schurz, 2017).

of each other. In this case, the narrowest screening off set would be \emptyset and our procedure would tell us to not draw any arrows between X_1, X_2 . The resulting type-level causal structure would misrepresent token-level causal facts. Because of this, we need to assume difference making: If X_i is a direct token-level cause of X_j in some system s , then all involved variables' values are distributed among systems s in such a way that X_i and X_j come out as probabilistically dependent to the background of some instantiation of the type-level causes of X_j in step 2. This requirement will guarantee that the resulting type-level structure satisfies the causal minimality condition.

5 The robustness of CBNs

In section 4 we have presented an account telling us under which conditions CBNs are generated by token-level causal patterns at the type level. Most of the time, however, we simply spoke about token-level causation (in general) and not about token-level causation according to a specific account. But as we saw in section 2, the different token-level theories give rise to different causal claims in different causal scenarios. So naturally the question arises whether there is one token-level theory that is better than its competitors at licensing CBNs. After having developed the two-step account, the answer to that question becomes clearer: Given the assumptions outlined in the previous section, CBNs are relatively robust with respect to the specific token-level theory. Even in those cases in which token-level accounts disagree with respect to causal connections the type-level structures resulting from the two-step approach will almost always fit all token-level theories. We believe that this is a huge advantage of CBN methods. They allow us to make many causal type-level inferences without us having to know which is the right token-level account of causation. Let us illustrate this by walking through the exemplary

cases from section 2 again. This time, we will take a closer look at how the different token-level theories from section 2 give rise to type-level models.

Throwing stones. As we have discussed in the previous section, to arrive at a type-level structure, we need not only to consider the one specific case where Suzy and Billy throw stones from section 2, but all nomologically possible systems s like this specific one, i.e., all nomologically possible cases where two individuals independently throw (or not throw) stones at vases. For easier reference, let us refer to the first one of the individuals in all such systems by *Suzy*, to the second one by *Billy*, and to the respective vase by *Vase* (likewise for the bombing mission case discussed later). Now in some of these systems, *Suzy* will throw earlier, in others, *Billy* will. In some, *Suzy* will be closer to *Vase*, in others, *Billy* will be closer. In some, *Suzy* will throw with more force and, thus, her stone will fly faster, in others *Billy* will use more force. And in some systems, *Suzy's* and *Billy's* throws will be similar in all these respects. In some one of the two does not throw and in some neither will throw. In yet other systems, other events occur that influence the trajectory of one or both of the stones in such a way that they do not hit *Vase*, etc. If we now apply the three proxy token-level theories introduced in section 2, we sometimes arrive at the same results, and sometimes at different ones when going through all of these systems. Let us walk through two examples. Figure 3 summarises the results.¹²

Suzy's stone hits first: This is the classical late pre-emption case discussed in section 2. As we saw in section 2, the process and the disruption account both tells us that *Suzy's* throw is a cause, but *Billy's* is not, while according to the counterfactual account both throws come out as causally irrelevant.

¹² To make the difference between type-level and token-level causal relations clearer, we will represent the former by continuous arrows and the latter by dashed arrows in graphs.

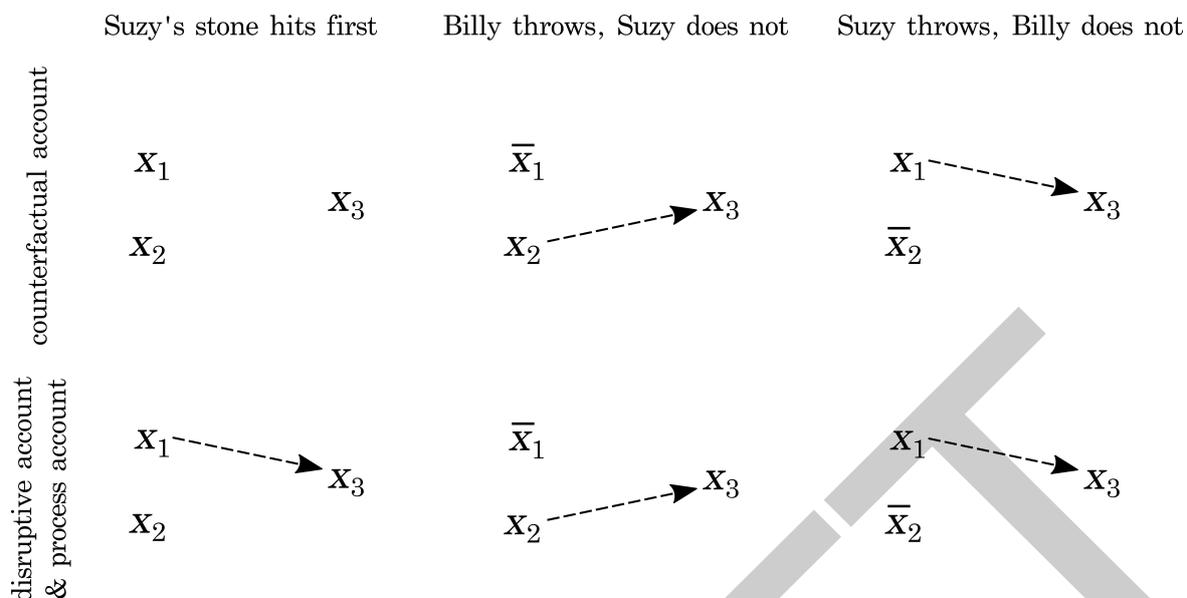


Figure 3: *Suzy* throws (X_1), *Billy* throws (X_2), *Vase* breaks (X_3).

Billy throws, *Suzy* does not: In this scenario, all three accounts of token-causation tell us that *Billy's* throw is a cause of the breaking of *Vase*. There is a suitable process connecting the two events, *Billy's* throw interferes and changes *Vase's* quasi-inertial behaviour, and the right counterfactual holds: Had *Billy* not thrown his stone, *Vase* would not have shattered. In the case where *Suzy* and *Billy* swap places, everything is inverted.

Bombing mission. Again, we need to have a look at all nomologically possible systems s similar to the one described in section 2. In some of these systems, double-prevention occurs as in section 2. But in others *Billy* will fail in his attempt to bring down *Enemy* in time. In others, *Suzy* will miss *Target*. In others, *Enemy* will not succeed in bringing down *Suzy* in time even if not intercepted by *Billy*. In still other systems, some of the persons involved might not even have started their mission, etc. Let us have a brief look at two specific cases. Figure 4 summarises the results.

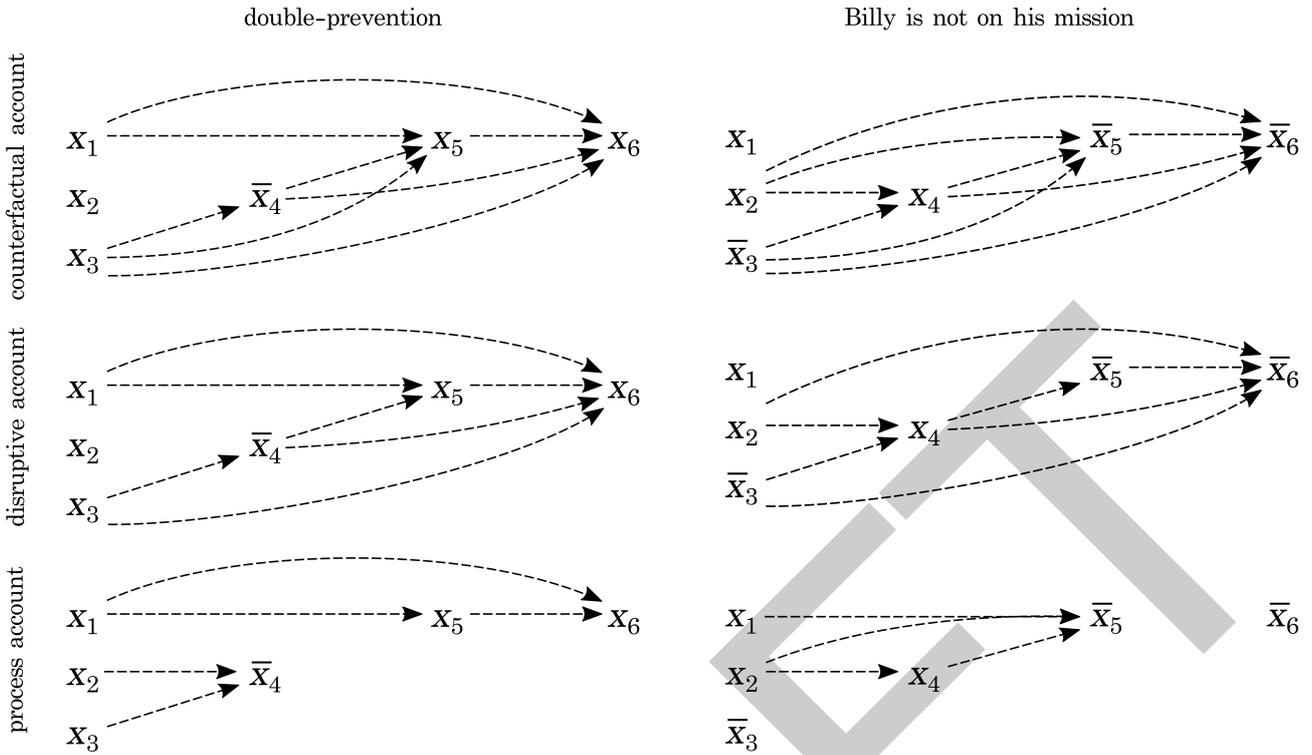


Figure 4: *Suzy* is on her mission (X_1), *Enemy* is on their mission (X_2), *Billy* is on his mission (X_3), *Enemy* pulls the trigger (X_4), *Suzy* pulls the trigger (X_5), *Target* gets destroyed (X_6).

Double-prevention: This is the case from section 2. *Suzy*, *Billy*, and *Enemy* are all on their mission. *Billy* intercepts *Enemy* and *Enemy* cannot bring down *Suzy* in time. Hence, *Suzy* pulls the trigger and *Target* gets destroyed. The following are the causal relations on which all three token-level accounts agree: *Suzy* being on her mission is causally relevant for her pulling the trigger and for *Target* getting destroyed. Also, *Suzy* pulling the trigger causes the destruction of *Target*. Finally, *Billy* being on his mission causes *Enemy* not pulling the trigger. Had one of the cited causes not occurred, then the cited effects would not have occurred either. Furthermore, the causes mentioned are closed system causes or disrupt the effects' quasi-inertial behaviour. Finally, there is a chain of causal processes and interactions connecting the effects to the causes. The token-level accounts disagree, however, on the following causal relations: While *Enemy* being on their mission is not a cause of *Enemy* not pulling the trigger on the counterfactual and disruption account, it is on the process account. By contrast, *Enemy* not pulling the trigger counts as a cause of *Suzy* pulling the trigger and also of *Target* getting destroyed on the counterfactual and the

disruption account, but not on the process account. Finally, *Billy* being on his mission comes out as causally relevant for *Suzy* pulling the trigger on the counterfactual account and for *Target* getting destroyed on the counterfactual and the disruption account, but not on the process account (since any causal process or chain of interactions initiated by *Billy* being on his mission ends with *Enemy* not pulling the trigger).¹³

Billy is not on his mission: This time *Billy* does not take part in the mission. As a consequence, *Enemy* is not intercepted and manages to shoot *Suzy*'s plane before she can pull the trigger, because of which *Target* does not get destroyed. The three token-level accounts agree on the following causal relations: *Enemy* being on their mission is causally relevant for *Enemy* pulling the trigger. Also, *Enemy* pulling the trigger comes out as a cause of *Suzy* not pulling the trigger. The accounts disagree on those causal relations: The counterfactual and the disruption account tell us that *Enemy* being on their mission, *Enemy* pulling the trigger, *Suzy* not pulling the trigger, and *Billy* not being on his mission are all causally relevant for *Target* not getting destroyed. While the counterfactual account results in *Billy* not being on his mission as a cause of *Suzy* not pulling the trigger, the disruption account does not. The process account denies all of these causal relations, but postulates an additional one between *Suzy* being on her mission and *Suzy* not pulling the trigger. The only processes and causal interactions involved in this scenario originate from *Enemy* being on their mission and end with *Suzy*'s plane getting intercepted.

¹³ When applying the process account to the double-prevention scenario we drew arrows from x_2 and x_3 to \bar{x}_4 . That requires a brief explanation: A causal process takes place from *Billy* being on his mission (x_3) to *Enemy* being shot down (which is not represented as a variable). *Enemy* being shot down implies that *Enemy* does not pull the trigger (\bar{x}_4). Likewise for \bar{x}_4 's other cause x_2 and \bar{x}_5 and its process causes in the scenario where *Billy* is not on his mission. We think that cases like these (with negative events as effects) can sometimes be integrated into a process account without being committed to endorsing quasi-causation (Dowe, 2000, ch. 6)

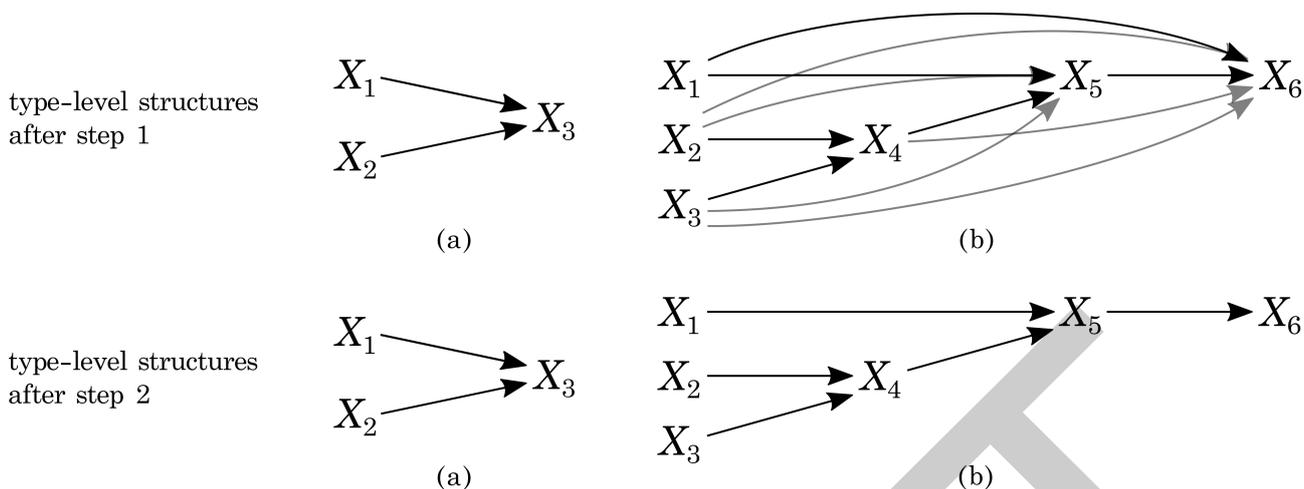


Figure 5: Type-level structures resulting from the two-step procedure applied to Throwing Stones (a) and Bombing Mission (b). Black arrows result from all token-level accounts, while grey arrows do not.

When going through all these examples, we find that the different token-level accounts of causation often disagree on what causes what. However, since the move from the token to the type level abstracts away from the specific causal dependencies among values of variables, these differences become largely irrelevant and we will end up with the same type-level structure regardless of the specific token-level causal differences. All that is needed is that when we are considering two variables and the question of whether one is a direct cause of the other, among the nomologically possible scenarios there is *at least one scenario* such that the event of the first variable having a certain value is a token-cause of the event of the second variable having a certain value. In particular, the type-level structures we will get are depicted in Figure 5. This confirms our earlier claim that one's specific token-level account does not matter that much for whether CBNs will be suitable tools for causal inference at the type level.

Note that to show that type-level causation is robust (i.e., independent of the specific token-level causal account), we do not have to assume that the three constraints introduced are satisfied for particular cases. It suffices to show that the constraints are independent of the token-level theories. Ordering uniqueness and Markov distribution do

not require any explicit details about token-causation. They only make assumptions about the distribution of the values of the variables among the nomologically possible systems s . This distribution is independent of which causal relations the different token-level theories postulate in these systems. The third assumption, difference-making, requires reference to token-causation. In particular, it makes assumptions about the distribution of variable values if there is at least one system s featuring a direct token-causal relation. As we saw above: If there is a system in which some X_i -values cause some X_j -values according to one token-level account, it can be expected that there will also be some (possibly different) X_i -values that cause some (possibly different) X_j -value according to the other token-level accounts. This makes it highly probable that also difference-making will be independent of the specific token-level theory.

6 Actual Causation

By accounts of actual causation we understand accounts that try to squeeze token-level causation out of type-level causal structure. As Woodward (2003) puts it, the “question is what these type-level causal relationships and other background information imply about token-causation” (p. 75). Our claim is that type-level causal relations on their own do not imply very much about token-level relations — and that should not come as a surprise. As we saw in sections 4 and 5, certain features of token-level causation are washed out in the abstraction process that is constitutive for type-level causal structure, meaning that we potentially lose information when moving from the token to the type level. We thus cannot expect to get the correct token-level predictions from our type-level account on its own.

The question of what CBNs imply about token-level causation can be understood in at least two ways:

- (i) Is there a specific account of token-causation that is implied by the type-level structure?
- (ii) Does the type-level structure (together with the actual values of the variables featuring in this structure) imply whether a particular event $X_i = x_i$ causes another event $X_j = x_j$?

The answer to the first question follows from the robustness of CBNs. The fact that type-level structures as captured by CBNs are (largely) robust vis-à-vis the underlying accounts of token-causation means that accounts of token-causation are not implied by the type-level structure. In other words: Type-level structures are compatible with token-level causation consisting in counterfactual dependence, the transfer of conserved quantities, etc.

This can be substantiated if we look at how accounts of actual causation are justified. In Woodward's (2003) well-known account of actual causation (**AC***) all the work is done by condition (**AC*2**):

(AC*2) For each directed path P from X_i to X_j , fix by interventions all direct causes Z of X_j that do not lie along P at some combination of values within their redundancy range. Then determine whether, for each path from X_i to X_j and for each possible combination of values for the direct causes Z of X_j that are not on this route and that are in the redundancy range of Z , whether there is an intervention on X_i that will change the value of X_j . **(AC*2)** is satisfied if the answer to this question is “yes” for at least one route and possible combination of values within the redundancy range of the Z .

The details of this condition need not bother us. What is important is that it is not justified by deriving it from Woodward's (2003) account of type-level causation, but rather by appeal to causal intuitions. For example, **(AC*2)** generates the right verdict in cases of causal overdetermination. (The same holds for other accounts of actual causation, e.g., Halpern, 2016; Halpern & Pearl, 2005.) In other words: The account of actual causation is vindicated by evidence independent of the type level account — by the kind of evidence that traditional accounts of token-causation appeal to.

Let us now turn to the second issue: We are now not asking whether a specific account of actual causation designed to capture one specific type of token-level causal relation in mind is implied by the type-level structure. Instead, we are interested in the question of whether any account of actual causation can squeeze out the correct token-level causal relationships from a given type-level structure. As we will see, the fact that type-level causal relations presuppose abstraction leads to problems here as well. Our analysis so far helps us to understand why this is a difficult project and why such accounts only work if a lot of additional piecemeal information is added.

Let us once more walk through one of our test cases.¹⁴ In both cases we will use **(AC*)**¹⁵ as a proxy actual causation account, but the general points we illustrate hold regardless of the specific account of actual causation. To keep things simple, we ignore error terms U_i .

Throwing Stones. The type-level causal structure for this kind of causal scenario — i.e., where two agent's independently throw stones at vases — is the one depicted in Figure

¹⁴ We only consider throwing stones because it is simpler than bombing mission and suffices to make our point.

¹⁵ For the full account **(AC*)**, see the appendix.

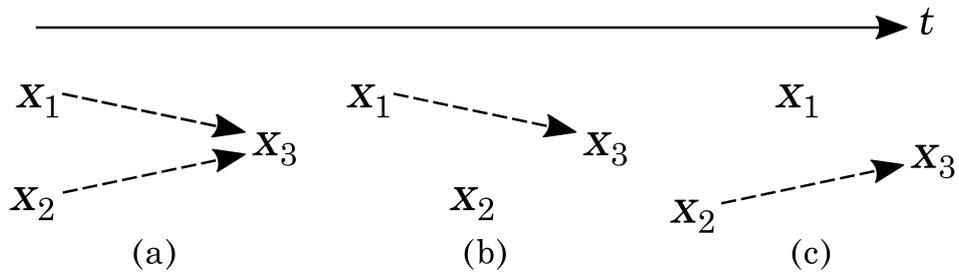


Figure 6

5(a). Variables are defined as before. We assume that one hit suffices for the vase to break:¹⁶

$$X_3 =_c X_1 + X_2 - X_1X_2 \quad (1)$$

The variables' actual values are:

$$X_1 = X_2 = X_3 = 1 \quad (2)$$

The token-level structures of a system s featuring *Suzy*, *Billy*, and *Vase* in which at least one of the two agents throws a stone and that could realise the type-level structure in Figure 5(a) are depicted in Figure 6. We assume that both *Suzy* and *Billy* stand equally far away from *Vase*, that the stones travel at the same speed, and that both have perfect aim. Figure 6(a) depicts a system where both *Suzy* and *Billy* throw at the same time, (b) the case where *Suzy* throws earlier, and (c) the situation where *Billy* does.

Now **(AC*)** applied to the type-level structure in Figure 5(a) tells us that *Suzy's* as well as *Billy's* throws were actual causes of *Vase's* breaking. In other words, it outputs the token-level structure in Figure 6(a), regardless of the specific time at which *Suzy* and *Billy* throw their stones. A similar result can be obtained, for example, when we assume that *Suzy* and *Billy* both throw at the same time, but their stones travel with different speeds, or when we assume that the distance to the target varies. The point here is that all the details

¹⁶ To keep things simple we assume that for both examples discussed in this section the background context $\mathbf{B} = \mathbf{b}$ is chosen in such a way that no error terms U_i are required in our structural equations.

that make a crucial difference for what causes what at the token-level are not captured by the type-level structure and the simple functional dependencies above. Each one of the situations in Figure 6(a)-(c) could in principle be the actual one, but only if it is (a) does **(AC*)** get things right.¹⁷ The point is, of course, that which one of the three situations is the actual one depends on the specific details of the situation. These details are not implied by the account but need to be added.

Now one might object that we used the wrong functional dependence for this specific case. But it seems hard to come up with something better than the dependence in Equation 1. After all, it remains true that *Vase* shatters exactly if at least one of the two agents throws. This goes for all the systems s correctly represented by the type-level structure in Figure 5(a), regardless of the additional details about time, distance between the target and the agents, the speeds of the stones, etc. We can also think about the situation in terms of subsets of systems s in which two agents independently throw (or not throw) stones at a target as follows: Let \mathbf{S} be the set of all systems s in which two agents independently throw (or not throw) stones at *Vase*. Now we introduce $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3 \subseteq \mathbf{S}$. Let \mathbf{S}_1 be the set of all s in which the two agents throw (if they throw both) at the same time (like in Figure 6(a)), \mathbf{S}_2 the set of all s in which the first agent throws earlier (like in Figure 6(b)) than the second (if both throw), and \mathbf{S}_3 the set of all s in which the second agent throws earlier (like in Figure 6(c)) than the first (if both throw). If we now apply the abstraction method outlined in section 4 to each of these subsets \mathbf{S}_i , we will in each case end up with the type-level structure in Figure 6(a) and the functional dependence in

¹⁷ Halpern & Pearl (2005) would also say that (a) is correct. If they are allowed to add two variables (for whether *Billy's/Suzy's* stone actually hits *Vase*), they get a more nuanced picture than Woodward (2003). But this seems a bit ad hoc: If we already know which result we want to get, then it becomes clear which additional variables and other structural elements need to be added. The problem is that if just starting from the type-level without any particular token-theory in mind we do not have this information.

Equation (1). After all, in all of these cases *Vase* shatters if at least one of the agents throws. This means that different token-level causal structures (e.g., Figure 6(a)-(c)) that give rise to one and the same type-level model (e.g., Figure 5(a) together with Equation (1)) will not be discriminable in terms of any account of actual causation. Vice versa, we can diagnose that an account of type-level causation has a chance to be adequate only if all the specific systems s covered by a type-level causal model share the same token-level causal structure.¹⁸

Here is another possible objection: One might worry that the system is not represented correctly by the variables X_1, X_2, X_3 . If one would introduce more variable values or additional variables covering all the missing details about the distance of each agent to *Vase*, the time at which each agent throws (or does not throw), etc., then one might be able to replace Equation 1 by one or more structural equations that would do a better job. We agree that this might work. However, what is required is additional information about the situation at hand, additional information that is typically captured in terms of token-level accounts of causation. Because type-level causation is coarse-grained and robust, an account of actual causation, in order to get from equation 1 to one of the figures in (6), has to look at exactly those details that are irrelevant for the type-level causal relations but significant for token-level accounts.

Furthermore, note that any type-level structure satisfying the assumptions introduced in section 4 will be suitable for type-level causal inference, regardless of how exactly one specifies their variables. For the throwing stones case, for example, there are many such successful type-level representations. The problem is that it seems that most of them, for example the one featuring the variables X_1, X_2, X_3 as specified earlier, is — as we saw above — not informative about the token-causal relations involved. But this emphasises

¹⁸ Note that splitting \mathbf{S} into subsets $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$ leads to the same result as packing who throws first into the background conditions $\mathbf{B} = \mathbf{b}$.

the point we want to make in this section: The robustness and abstractness of the type-level explains why it is so difficult for an account of actual causation to reliably squeeze out token-level causation from type-level structure. Different accounts such as **(AC*)** (or Halpern & Pearl, 2005) will be lucky in some cases (given some token-level interpretation of causation) and perform horribly in others. We also note that one's variables can be fine-grained in multiple ways and that not every fine-graining of X_1, X_2, X_3 would give us the right result (to the background of a specific token-level account).

7 Conclusion

In this paper we explored how CBN methods for type-level and token-level causal reasoning are related. In particular, we chose the simple counterfactual account, the conserved quantity theory, and the disruption account as proxies for the whole plethora of token-causation accounts. We presented two results.

(i) It can be explained that the success of CBN's as a type-level account of causation is not tied to a particular conception of token-level causation. CBNs turn out to be relatively robust w.r.t. one's specific token-level causal theory. In section 4, we suggested an abstraction procedure that constructs type-level causal structures based on token-level causal facts. The procedure works whenever three additional modelling assumptions, ordering uniqueness, Markov distribution, and difference-making, are satisfied. In section 5 we then found that though different token-level accounts might drastically disagree on what causes what in specific situations, mostly CBNs are abstract enough so that they can represent token-level causal patterns regardless of these disagreements as long as the three conditions mentioned are satisfied. As we also saw, it is highly likely that once one causal scenario satisfies these conditions to the background of one token-level account, then it also does so to the background of all the other token-level accounts. The upshot of all this is that the success of CBNs for type-level reasoning can be explained without

having to know which token-level theory provides the correct account, which we believe is a huge advantage.

(ii) A similar result does, alas, not follow for CBN-based accounts of actual causation. CBNs obtain their robustness when it comes to type-level inference by ignoring many token-level details. The downside of this is that type-level structure alone cannot tell us much about token-level causation. In particular, type-level structure alone cannot tell us which token-causation account gets things right in the end. Depending on which specific intuitions (e.g., counterfactual intuitions) and which specific cases (e.g., double-prevention) one wants to capture, one needs to make additional assumptions leading to different accounts of actual causation. Since these assumptions are tailor-made for specific intuitions and causal scenarios, one cannot hope that they will excel in different scenarios. Another problem is that so many token-level details are ignored by the type-level structure represented by a CBN that it seems impossible to get the correct results for all nomologically possible systems. What this means in the end is that not only tracking one specific token-causal account seems hopeless, but also that a more principled (or axiomatic) treatment of actual causation such as, for example, suggested by Glymour et al. (2010) does not seem too promising.

Appendix

In this appendix we provide details about Woodward's (2003) account of actual causation (**AC***). For a more detailed explanation, see (ibid., sec. 2.7).

(AC*1) The actual value of X_i is x_i and the actual value of X_j is x_j .

(AC*2) For each directed path P from X_i to X_j , fix by interventions all direct causes Z of X_j that do not lie along P at some combination of values within their

redundancy range. Then determine whether, for each path from X_i to X_j and for each possible combination of values for the direct causes Z of X_j that are not on this route and that are in the redundancy range of Z , whether there is an intervention on X_i that will change the value of X_j . **(AC*2)** is satisfied if the answer to this question is “yes” for at least one route and possible combination of values within the redundancy range of the Z .

(AC*) $X_i = x_i$ is an actual cause of $X_j = x_j$ if and only if **(AC*1)** and **(AC*2)** are satisfied.

The notion of a redundancy range is borrowed from Hitchcock (2001) and is defined as follows:

Redundancy Range: Let P be a directed path from X_i to X_j and Z be the direct causes of X_j not lying on P . Then the values z_1, \dots, z_n are in the redundancy range w.r.t. P if and only if, given the actual value of X_i , there is no intervention setting the values of Z to z_1, \dots, z_n that will change the actual value of X_j .

From this definition it follows that the actual values of the direct causes Z of X_j are always in the redundancy range of P .

Acknowledgements: We would like to thank Thomas Blanchard, Bram Vaassen, and two anonymous reviewers for helpful comments on an earlier version of this paper.

References

- Cartwright, N. (1999). Causal diversity and the Markov condition. *Synthese*, 121(1/2), 3–27.
- Cartwright, N. (2007). *Hunting causes and using them*. Cambridge: Cambridge University Press.
- Dowe, P. (2000). *Physical causation*. Cambridge: Cambridge University Press.
- Gebharder, A., & Retzlaff, N. (2020). A new proposal how to handle counterexamples to Markov causation à la Cartwright, or: fixing the chemical factory. *Synthese*, 197(4), 1467–1486.
- Glymour, C. (1999). Rabbit Hunting. *Synthese*, 121(1/2), 55–78.
- Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., Spirtes, P., Teng, M. T., & Zhang, J. (2010). Actual causation: A stone soup essay. *Synthese*, 175(2), 169–192.
- Hall, N. (2007). Structural equations and causation. *Philosophical Studies*, 132(1), 109–136.
- Halpern, J. Y. (2016). *Actual causality*. Cambridge: MIT Press.
- Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach, part I: Causes. *British Journal for the Philosophy of Science*, 56(4), 843–887.
- Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy* 98, 273–99.
- Hüttemann, A. (2020). Processes, pre-emption and further problems. *Synthese* 197, 1487–1509
- Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.
- Schurz, G. (2017). Interactive causes: Revising the Markov condition. *Philosophy of Science*, 84(3), 456–479.

Schurz, G., & Gebharder, A. (2016). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese*, 193(4), 1073–1103.

Woodward, J. F. (2003). *Making things happen*. Oxford: Oxford University Press.

DRAFT