A Causal Bayes Net Analysis of Glennan's Mechanistic Account of Higher-Level Causation (and Some Consequences)

Alexander Gebharter

One of Glennan's ([1996]) most prominent contributions to the new mechanist debate consists in his reductive analysis of higher-level causation in terms of mechanisms. In this article I employ causal Bayes nets (CBNs) to reconstruct his analysis. This allows for identifying general assumptions that have to be satisfied to get the analysis working. I show that once these assumptions are in place, they imply (against the background of the CBN machinery) that higher-level causation indeed reduces to interactions between component parts of mechanisms. I also briefly discuss the plausibility of these assumptions and some consequences for the mechanism debate.

1. Introduction

Mechanisms play an important role for many sciences such as biology, medicine, and neuroscience. Mechanisms are systems that can be described at different levels. They consist of several lower-level component parts whose causal interactions give rise to the overall behaviour of the system as a whole (Glennan [1996]; Machamer et al. [2000]; Illari and Williamson [2012]). Most modern mechanists are mainly interested in how mechanisms and models of mechanisms can be used for purposes of explanation and prediction (Bechtel and Abrahamsen [2005]; Craver [2007]). Mechanistic explanations come with the following key advantage over purely causal explanations: While purely causal knowledge only allows for an horizontal (or etiological) explanation of a phenomenon, the understanding of the mechanism underlying this phenomenon provides a basis for a vertical (or synchronic) explanation across levels

Electronically published March 22, 2022.

The British Journal for the Philosophy of Science, volume 73, number 1, March 2022.
 https://doi.org/10.1093/bjps/axz034

as well. So mechanisms provide a basis for causal and constitutional reasoning across different levels of organization.[1]

Since the concept of mechanism is intimately connected to that of causation, it seems promising to try to reduce (or explain) one of these two concepts to (or in terms of) the other. Attempts to analyse mechanisms in terms of causation (Gebharter [2014], [2017b]; Woodward [2002]) seem quite natural, simply because mechanisms are understood as systems consisting of causally interacting parts anyway. The other direction, namely, to analyse causation in terms of mechanisms, seems more challenging. Perhaps the most prominent mechanistic account of causation has been put forward by Glennan ([1996]). In a nutshell, Glennan proposes that causal relations (or laws) connecting higher-level phenomena occur (or hold) due to mechanisms underlying these phenomena.

Several philosophers, such as Craver ([2007]) and Psillos ([2004]), have attacked Glennan's ([1996]) mechanistic analysis of higher-level causation, and Glennan has defended his approach against these and other criticisms (see, for example, Glennan [2011]). The debate is still ongoing (Casini [2016]), and the question of whether a mechanistic analysis of higher-level causation can be successful is still controversial. This article is intended as a contribution to the debate about how mechanisms relate to higher-level causation. I use the causal Bayes net (CBN) machinery (Spirtes et al. [1993]; Pearl [2000]) as a tool for investigating Glennan's mechanistic analysis of higher-level causation. The framework allows for testing causal hypotheses, it provides the basis for a multitude of sophisticated procedures for uncovering causal structure (Spirtes et al. [1993]), and the approach satisfies standards that successful empirical theories satisfy as well. In particular, it provides the best explanation for why sets of variables describing different empirical systems show different probabilistic dependence and independence patterns, and the theory as a whole can be tested on empirical grounds (Gebharter [2017b]; Schurz and Gebharter [2016]). The hope is that the framework can thus also be used to get an empirically informed grasp on philosophical issues such as the question of how higher-level causation relates to mechanisms.[2]

Note that CBNs have already been used for modelling mechanisms (see, for example, Casini et al. [2011]; Clarke et al. [2014]; Gebharter [2014]; Murray-Watters and Glymour [2015]; Koch et al. [2017]).[3] I am also not the first to use causal modelling techniques to shed new light on the question of whether higher-level causation can be reduced to mechanisms. Casini ([2016]), for example, but also Glennan ([2011]) himself borrows from the causal modelling literature. The main difference between these approaches and my own approach is that I combine different levels as well as

[1] How to define and identify levels is an important but still controversial question (Craver and Bechtel [2007]; Eronen [2015]; Kästner [2018]). In this article I will bracket this problem and assume that it can be solved in some way.

[2] Note that alternative frameworks might be used for investigating how higher-level causation relates to mechanisms as well. However, I think that CBNs are especially nice for the reasons mentioned.

[3] For possible problems with such approaches, see (Kaiser [2016]; Weber [2016]). For a possible solution to some of these problems, see (Gebharter and Koch [2021]).

causal and constitutional relationships in a single causal model.[4] The reconstruction of Glennan's ([1996]) mechanistic analysis to be developed in this article will show that a supporter of such an analysis needs to make three interrelated basic assumptions in order to reduce higher-level causation to mechanisms. Once one subscribes to these assumptions, it follows from the CBN apparatus that higher-level causation reduces to interactions between component parts of mechanisms. Though an in-depth evaluation of these assumptions would go beyond the scope of this article, I briefly discuss them in order to get a first impression of their putative plausibility.

The article is structured as follows: In section 2 I present Glennan's ([1996]) mechanistic account of higher-level causation. In section 3 I introduce the basics of the CBN framework required for subsequent sections. In section 4 I use the CBN framework as a basis for reconstructing Glennan's analysis. I also formulate and briefly discuss the three basic assumptions mentioned. In section 5 I discuss several interesting consequences of my analysis for the mechanism debate. In particular, I discuss (i) the worry that a mechanistic analysis of higher-level causation might just reduce higher-level causation to lower-level causation, (ii) several consequences for the debate about inter-level causation in the presence of mechanisms, and (iii) how interventions on mechanisms might work according to the analysis provided. I conclude in section 6.

Summarizing, the article aims at the following three main goals:

(1) Fleshing out Glennan's ([1996]) basic idea by providing a general formal framework for reducing higher-level causation in terms of mechanisms.

(2) Using the framework as a tool for identifying several assumptions one has to make (in addition to the framework) in order to get the reduction done.

(3) Sketching how the framework might shed new light on several other issues discussed in the mechanism literature.

2. Glennan's Analysis of Higher-Level Causation

Glennan ([1996]) starts his seminal article with Hume's problem. In a nutshell, Hume's problem consists in the question of how we can acquire causal knowledge even though we seem to be unable to qualitatively distinguish between regularities due to causal laws and spurious correlations. From a phenomenological point of view, the two kinds of regularities are indistinguishable simply because the 'secret connexion' that binds cause and effect together does not produce any specific impression or experience in the observer. While Hume's own skeptical solution to the problem consists in redefining causation as regular conjunction, Glennan proposes

[4] Casini et al. ([2011]) and Clarke et al. ([2014]) combine causal and constitutional relationships as well. They are, however, more interested in modelling mechanistic hierarchies than in analysing higher-level causation in terms of mechanisms.

that one might be able to distinguish between regularities due to causal laws and spurious correlations by checking whether there exists a mechanism that connects the phenomenon of interest with its purported cause. According to Glennan ([1996], p. 49), Hume's claim that 'no number of observations can yield an impression of a connection' between cause and effect is correct, but his claim that this connection is 'secret' is false. Actually, one can establish a causal relation by discovering a mechanism that connects the cause to the effect phenomenon, which will often require sophisticated scientific methods.

To further spell out the idea of reducing higher-level causation to mechanisms, Glennan ([1996], p. 52) provides the following definition of the concept of mechanism:

> A mechanism underlying a behaviour is a complex system which produces that behaviour by of the interaction of a number of parts according to direct causal laws.

According to this definition, a mechanism is always a mechanism for a specific behaviour. This nicely fits the idea that one and the same system might behave in quite different ways, where interactions between different parts might be responsible for these different behaviours. This observation leads to the distinction between the parts of a system that are constitutively relevant for the behaviour of interest and those that are not. Let us call the former, in accordance with Craver ([2007]), the mechanism's components. As the definition says, these components are supposed to bring about the behaviour of interest by causally interacting with each other. The part of the definition that requires them to interact 'according to direct causal laws' is intended to guarantee genuine causal interactions between component parts as well as that the overall 'behaviour of the aggregate stems from a series of local interactions' (Glennan [1996], p. S344). Note that later on, Glennan ([2002], p. S344) replaced the phrase 'according to direct causal laws' in the original definition with 'where the interactions between parts can be characterised by direct, invariant, change-relating generalisations', where 'direct, invariant, change-relating generalisations' should be understood in the sense of Woodward ([2003]). However, since whether one prefers the one or the other phrase to characterise mechanisms will play no role for the reconstruction of Glennan's analysis of higher-level causation in section 4, I will stick with the simpler definition from the 1996 article.

Glennan's ([1996], p. 64) mechanistic analysis of higher-level causation states that 'two events are causally connected when and only when there is a mechanism connecting them'. The analysis comes with a number of philosophical merits: First, it renders scientifically tangible questions about whether higher-level phenomena are causally connected. It suggests that questions like these can be answered on the basis of empirical investigations. This is, for example, also supported by the Russo–Williamson thesis (Russo and Williamson [2007]) that states that establishing causal claims requires that difference-making evidence is supported by an understanding of the mechanism responsible for the phenomenon of interest. Second, it provides a more

sophisticated solution to Hume's problem than regularity theories of causation have to offer. While regularity theories redefine causation as constant conjunction and, thus, run the risk of rendering too many regularities causal, a mechanistic analysis allows for phenomena that are not causally related though they regularly occur together. Finally, a mechanistic analysis of higher-level causation allows for multiple realisability. Smoking, for example, is a cause of premature death, but there are many different mechanisms for explaining how exactly smoking can lead to premature death.

There are, of course, also problems with Glennan's ([1996]) mechanistic analysis of higher-level causation. One of the most frequently discussed problems is that a mechanistic account does not provide any possibility for analysing causal interactions among component parts at the fundamental physical level. While the account seems to be capable of solving Hume's problem for higher-level causation, the problem remains a threat at the fundamental physical level. Glennan ([2002], [2011]) argues that counterfactuals stating what would happen under interventions can be used to achieve knowledge of the causal relations among component parts at the fundamental physical level. He emphasises, however, that though interventions can be used for such epistemic purposes, they do not provide a basis for a thorough metaphysical analysis of the concept of causation.[5] Another problem with Glennan's mechanistic analysis might be that it seems to simply reduce higher-level causation to causation at the fundamental physical level. Glennan ([1996], p. 66) counters this worry by pointing out that 'it is not [the fundamental causal] laws which make the causal claim true; rather, it is the structure of the higher level mechanism and the properties of its parts'. The idea here is that the presence of the mechanism as a structured whole makes—in addition to the causal structure at the fundamental physical level—a substantial contribution in bringing about cause-effect relationships at higher levels. I will come back to this and other issues in section 5. But first I will briefly introduce the basics of the CBN formalism relevant for subsequent sections.

3. Causal Bayes Nets

Bayes nets (BNs) were originally developed to graphically store independence information and to simplify reasoning under uncertainty. The formalism can, however, be used to represent all kinds of relations and dependencies that have just the right formal properties. In particular, they must conform to the Markov condition (Spirtes et al. [1993], p. 33), which establishes a connection between probabilistic dependence and graphical structure. Before I can introduce the Markov condition properly, a few preliminaries are required. In the following, **V** shall be a set of random variables, $X_1, \ldots, X_n$, which might represent events or properties; **E** a set of arrows ($\rightarrow$) connecting pairs of variables in **V**; and P a probability distribution over **V**.

[5] For a discussion of the limitations of interventions for testing causal relations underpinned by complex mechanisms, see (Casini [2016]).

$\mathbf{G} = \langle \mathbf{V}, \mathbf{E} \rangle$ is called the BN's graph, and $\mathbf{Par}(X_j)$ shall stand for the set of a variable X_j's parents, which is the set of all $X_i \in \mathbf{V}$, with $X_i \rightarrow X_j$. Variables X_i that are connected to X_j via a path of the form $X_i \rightarrow \ldots \rightarrow X_j$ are called X_j's ancestors, and variables X_j connected to a variable X_i over such a path are called X_i's descendants. The Markov condition can be formulated as follows:

> Markov Condition: $\langle \mathbf{V}, \mathbf{E}, P \rangle$ satisfies the Markov condition if and only if $\mathbf{G} = \langle \mathbf{V}, \mathbf{E} \rangle$ and probability distribution P conforms to
>
> $$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathbf{Par}(X_i)). \quad (1)$$

The conditional probabilities $P(X_i|\mathbf{Par}(X_i))$ appearing in equation 1 are called X_i's parameters. The Markov condition is basically a screening off condition. If satisfied, it guarantees that every variable $X_i \in \mathbf{V}$ is independent of its non-descendants (which are the variables X_j not connected via a path $X_i \rightarrow \ldots \rightarrow X_j$ to X_i) conditional on its direct ancestors (which are the variables in $\mathbf{Par}(X_i)$).[6] Now one kind of relation that conforms to equation 1, and thus could be represented by a BN's arrows, is direct causal dependence (Spirtes et al. [1993]; Pearl [2000]). We shall call a BN in which some (or all) arrows are causally interpreted a causal Bayes net (CBN). For a model in which all arrows are causally interpreted, the Markov condition implies that every variable becomes probabilistically independent of all its non-effects given its direct causes. It also implies Reichenbach's ([1956]) insights that common causes screen off their effects and that an effect's direct causes screen it off from its indirect causes.

Other relations that seem to conform to equation 1 are, for example, supervenience and constitution (Gebharter [2017a], [2017c]). Recent work by Schaffer ([2016]) suggests that the grounding relation might conform to the Markov condition as well. In the following, however, I will only be interested in BNs that represent causal and/or constitutional relationships. I will, from now on, represent direct causal connections by continuous arrows ($\rightarrow$) and direct constitutional relationships by dashed arrows (---›).[7] I will refer to a variable X_i's direct causes as its causal parents and to X_i's direct constituents as its constitutional parents. Likewise, I will refer to the variables X_i connected to a variable X_j via a path of the form $X_i \rightarrow \ldots \rightarrow X_j$ as X_j's causal ancestors, and to the variables X_j connected to a variable X_i over such a path as X_i's causal descendants. If X_i and X_j are connected over a path of the form X_i ---› ... ---› X_j, I will refer to X_i as one of X_j's constitutional ancestors and to X_j as one of X_i's constitutional descendants.

[6] Probabilistic independence between X_i and X_j conditional on X_k is defined as $P(x_i|x_j, x_k) = P(x_i|x_k) \vee P(x_j, x_k) = 0$ for all X_i-, X_j-, and X_k- values x_i, x_j, and x_k, respectively. Probabilistic dependence is defined as the negation of probabilistic independence.

[7] Note that representing these different kinds of relations by different arrows is just a cosmetic procedure. Both kinds of arrows are assumed to technically work in exactly the same way.

Given the terminology introduced we can state the Markov condition for systems possibly featuring both causal and constitutional relations as follows:

> Supervenience or Constitutional Markov Condition: $\langle \mathbf{V}, \mathbf{E}, P \rangle$ satisfies the causal or constitutional Markov condition if and only if every $X_i \in \mathbf{V}$ is probabilistically independent of its non-descendants conditional on its causal and constitutional parents.

The CBN framework comes with a simple test for whether a particular arrow $X_i \rightarrow X_j$ (or $X_i \dashrightarrow X_j$) can mediate probabilistic influence between X_i and X_j. To determine whether X_j-variations are correlated with X_i-variations in some circumstances due to the arrow $X_i \rightarrow X_j$ (or $X_i \dashrightarrow X_j$), one has to check whether X_j probabilistically depends on X_i conditional on X_j's other parents (Schurz and Gebharter [2016], p. 1087), meaning that one has to check whether $P(x_j|x_i, \mathbf{r}) \neq P(x_j|\mathbf{r}) \wedge P(x_i, \mathbf{r}) > 0$ holds for some combination of X_i-values x_i, X_j-values x_j, and values $\mathbf{r}$ of $\mathbf{Par}(X_j)\backslash\{X_i\}$. If the answer to this question is positive, we say that the arrow $X_i \rightarrow X_j$ (or $X_i \dashrightarrow X_j$) is productive. Arrows that do not pass the productivity test are redundant and can be eliminated without making the model less informative.[8] This fact will play a crucial role for evaluating Glennan's ([1996]) mechanistic analysis of higher-level causation in section 4.

Since the productivity test introduced will do a lot of work in subsequent sections, further elaboration and illustration seem appropriate. I will use the test as a kind of Occam's razor: If two (or more) parents of a variable compete for relevance, the one (or ones) not making any difference must go. Assume, for example, we are interested in whether Suzy throws a stone ($S = 1/0$) and in whether Billy throws a feather ($B = 1/0$) at a vase that might, as a result, shatter ($V = 1/0$). Suppose we are interested in the following two competing causal hypotheses:

> h_1: Whether Suzy throws ($S = 1/0$) is causally relevant for whether the vase breaks ($V = 1/0$).
>
> h_2: Whether Billy throws ($B = 1/0$) is causally relevant for whether the vase breaks ($V = 1/0$).

S passes the productivity test. If Billy throws the feather, Suzy throwing the stone still makes a difference for whether the vase breaks ($P(V = 1|S = 1, B = 1) \neq P(V = 1|B = 1)$), and we can conclude that h_1 is true. But since whether Billy throws does not make a difference either if Suzy throws ($P(V = 1|B = 1, S = 1) = P(V = 1|B = 0, S = 1) = P(V = 1|S = 1)$) or if she does not throw ($P(V = 1|B = 1, S = 0) = P(V = 1|B = 0, S = 0) = P(V = 1|S = 0)$), B fails

[8] Assuming productivity for all of a model's arrows amounts to assuming minimality (Schurz and Gebharter [2016], theorem 2). Note that applying the productivity test to single causal or constitutional arrows only requires that the model satisfies the Markov condition; neither minimality (Spirtes et al. [1993], p. 34) nor the stronger faithfulness assumption (Spirtes et al. [1993], p. 56) is required. Consequently, I will assume neither minimality nor faithfulness in this article.

the test, and we conclude that h_2 is false. Note that this does not yet mean that h_2 can be reduced to (or explained in terms of) h_1. However, depending on the context, one can use the productivity test for answering questions like these as well. While h_1 and h_2 describe same-level phenomena, we can construct a similar case involving phenomena located at different levels. Assume, for example, I expect it to rain later on and we are interested in the causal impact of whether I decide to take an umbrella with me. Suppose we are interested in the following two competing causal hypotheses:

> h_3: My decision ($D = 1/0$) is causally relevant for whether I take an umbrella with me ($U = 1/0$).
>
> h_4: The corresponding brain processes ($B = 1/\ldots/n$) are causally relevant for whether I take an umbrella with me ($U = 1/0$).

Depending on how exactly my decisions are related to what is going on in my brain, one will get different results from applying the productivity test to this scenario. If my decisions are multiply realisable and supervene on brain processes, then h_3 will fail and h_4 will pass the test. In the context of different levels, this could be interpreted as h_3 being reducible to h_4. While it makes perfect sense to say that my decisions are causally relevant for whether I take an umbrella with me in everyday live situations and higher-level causal claims might also be useful for scientific explanation, they can, in the end, be reduced: Higher-level phenomena are only causally relevant because they relate to lower-level phenomena in a certain way, but the lower-level phenomena do all the actual causal work. In subsequent sections I will use the productivity test in this latter sense: as an indicator for when certain causal relations can be reduced in terms of more fundamental causal relations (and inter-level relationships of constitution).

4. Reconstructing Glennan's Analysis

In this section, I use the CBN framework as a basis for developing a reconstruction of Glennan's ([1996]) mechanistic account of higher-level causation.[9] For illustrative purposes, I introduce an abstract toy example that shall stand proxy for all possible cases involving higher-level causation. Now, for Glennan, the analysis of higher-level causation starts with two higher-level phenomena that stand in a causal relationship. Let us model the possible behaviours of the systems exhibiting these phenomena with the two random variables H_1 and H_2. In addition, we assume that H_1 is causally relevant for H_2. Hence, $H_1 \rightarrow H_2$ has to be a part of our model. Further details about how the possible behaviours of the systems described by H_1 and H_2 are related to each other is modelled by a probability distribution P over $\mathbf{V}$, where $\mathbf{V}$ is a set

[9] The basic idea for modelling causal and inter-level relations in one and the same model is the same as in (Gebharter [2017a]). While the productivity test introduced in section 3 is used to answer questions about the causal efficacy of mental properties in (Gebharter [2017a]), I use it as a device for exploring issues concerning reduction in terms of mechanisms in this article.

of variables containing H_1, H_2, and several other variables to be introduced soon. Since H_1 is assumed to be a cause of H_2, we assume that H_2 depends on H_1, meaning that $P(h_2|h_1) \neq P(h_2) \wedge P(h_1) > 0$ holds for some H_1- and H_2-values h_1 and h_2, respectively.

The next step in Glennan's ([1996]) analysis would be to decompose the two systems whose higher-level behaviours are described by H_1 and H_2 and to identify the component parts of the mechanisms responsible for these behaviours. There are currently several proposals for how to identify a mechanism's component parts on the market (see, for example, Craver [2007]; Harbecke [2015]; Baumgartner and Gebharter [2016]; Baumgartner and Casini [2017]; Gebharter [2017c]; Krickel [2018]). However, since we are first and foremost interested in reconstructing Glennan's analysis rather than in the question of how to distinguish the parts constitutively relevant for certain behaviours from those that are not, we can ignore the latter question for the moment. For now it is only important to stress that some parts of our systems stand in constitutive relevance relations to the higher-level behaviours of interest, while others do not. Let us assume that the mechanism underlying H_1 consists of three constitutively relevant parts, while the mechanism responsible for H_2 consists of two. We model H_1's component parts with the variables L_1, L_2, and L_3 and H_2's components with L_5 and L_6. Let us assume that we also found a non-component part that we will model with L_4. (There might be additional non-component parts that we will ignore for now.) Next, let us assume that the lower-level variables $L_1, \ldots, L_6$ are causally connected to each other like in figure 1.[10] By adding a dashed arrow for every component part, we finally arrive at the graph depicted in figure 1.[11]

Now the proxy model in figure 1 and the CBN framework can be used as a tool for investigating the commitments of Glennan's ([1996]) mechanistic account of higher-level causation. If higher-level causation can actually be reduced in terms of mechanisms, this would amount to the claim that certain constraints that allow for eliminating higher-level causal arrows such as $H_1 \rightarrow H_2$ have to hold for all models like the one in figure 1. Now the crucial question is whether we can formulate such constraints. They would have to come in the form of general assumptions one has to make in order to render higher-level arrows such as $H_1 \rightarrow H_2$ unproductive in the sense explained in section 3. The reduction of the higher-level causal relationship between H_1 and H_2 would, according to this idea, consist in demonstrating that the arrow $H_1 \rightarrow H_2$ is redundant, meaning that it does no causal work that is not already done by some lower-level causal structure.

Here comes the first assumption required to get the reduction of higher-level causation in terms of the model: All kinds of higher-level phenomena occur because

[10] Note that nothing hinges on the particular causal structure among lower-level variables as long as the assumptions to be introduced are satisfied.

[11] Constitutive relevance arrows are drawn from lower- to higher-level variables because higher-level phenomena behave with respect to their component parts similarly to how effects behave with respect to their causes. For a more detailed argumentation, see (Gebharter [2017c]).

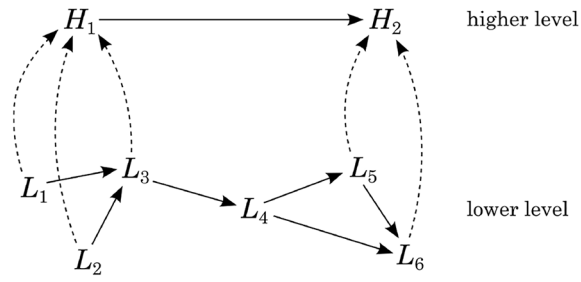

Figure 1. Toy model for exploring the commitments of Glennan's ([1996]) mechanistic account of higher-level causation.

of underlying mechanisms. The assumption guarantees that for every higher-level phenomenon described by some variable H there will be lower-level constitutively relevant parts that can be described by variables $L_1, \ldots, L_n$. It is typically also assumed that the overall behaviour of a mechanism H supervenes on the behaviours of the mechanism's component parts $L_1, \ldots, L_n$ (see, for example, Glennan [1996]; Craver [2007]; Baumgartner and Gebharter [2016]).[12] We can summarize and formulate these requirements in terms of CBN models as follows:

> Assumption 1 (Supervenience/Constitution): For each higher-level phenomenon described by a variable H there are component parts described by variables $L_1, \ldots, L_n$ such that $L_1 \dashrightarrow H, \ldots, L_n \dashrightarrow H$, and P is constrained by
>
> $$\forall h \forall h' \exists \mathbf{l} : \text{If } h \neq h', \text{ then } P(\mathbf{l}|h) \neq P(\mathbf{l}|h'), \text{ and} \tag{2}$$
>
> $$\forall \mathbf{l} \exists h : P(h|\mathbf{l}) = 1, \tag{3}$$
>
> where $\mathbf{L}$ (with values $\mathbf{l}$) is an individual variable ranging over all possible combinations of instantiations of lower-level variables $L_1, \ldots, L_n$.

Equations 2 and 3 are constraints on the probability distributions that come with models of mechanisms. They reflect probabilistic implications of supervenience and constitutive relevance: If and only if H supervenes on $\{L_1, \ldots, L_n\}$, then every change of H's value will be accompanied by a change in the probability distribution over $\{L_1, \ldots, L_n\}$ (equation 2). And vice versa: whenever the values of all variables $L_1, \ldots, L_n$ representing constitutively relevant parts are fixed, then also H's value will be determined (equation 3).

[12] Note that in mechanistic models actually used for explanation, the supervenience assumption will not always hold. However, the project of reducing higher-level causation in terms of mechanisms is not epistemic but rather metaphysical in nature, and almost everyone seems to agree that given the full metaphysical picture, a mechanism's macro behaviour supervenes on the behaviours of its component parts. Throughout this article I assume that lower-level variables $L_1, \ldots, L_n$ describing the component parts of a mechanism can in principle be defined, even if actually unknown.

Assumption 1 renders higher-level causal relations redundant. This can be illustrated on the basis of our toy model as follows: Let $\mathbf{l}$ be an arbitrarily chosen combination of instantiations of H_2's constitutively relevant variables L_5 and L_6.[13] It follows with equation 3 that there is exactly one H_2-value h_2 for every $\mathbf{l}$ such that $P(h_2|\mathbf{l}) = 1$, while $P(h_2'|\mathbf{l}) = 0$ for all H_2-values h_2' different from h_2. There are two possible cases for every H_1-value h_1: Either (i) h_1 and $\mathbf{l}$ are compatible, meaning that $P(h_1, \mathbf{l}) > 0$, or (ii) they are not, meaning that $P(h_1, \mathbf{l}) = 0$. If (i) is the case, then $P(h_2|h_1, \mathbf{l}) = P(h_2|\mathbf{l})$ and $P(h_2'|h_1, \mathbf{l}) = P(h_2'|\mathbf{l})$ hold. Thus, no H_2-value depends on h_1 conditional on $\mathbf{l}$. If (ii) is the case, then no H_2-value depends on h_1 conditional on $\mathbf{l}$ by definition. So conditionalizing on $\mathbf{l}$ renders H_2 independent from H_1 in both cases (i) and (ii). Because $\mathbf{l}$ was arbitrarily chosen, this result generalises: H_2 and H_1 are independent conditional on any possible combination of instantiations of the variables L_5 and L_6, meaning that H_2 and H_1 are independent conditional on $\mathbf{Par}(H_2)\backslash\{H_1\} = \{L_5, L_6\}$. Thus, the arrow $H_1 \rightarrow H_2$ is unproductive. Note that this reasoning does not hinge on the particular structure of our exemplary model. It can be generalized for all kinds of CBN models involving mechanisms that satisfy assumption 1.

Note, however, that assumption 1 is still not sufficient for a general reduction of higher-level causation to mechanisms. One needs to assume, in addition, that at least some of the components of the cause mechanism are causally relevant for at least some of the components of the effect mechanism. In other words, one needs to assume that for every higher-level causal relation there is a corresponding lower-level causal connection. In figure 1, this assumption is already satisfied due to each one of the directed causal chains going through L_4. We can state this second assumption as follows:

> Assumption 2 (Lower-Level Causal Relevance 1): If a higher-level phenomenon described by a variable H_i is causally relevant for another higher-level phenomenon described by a variable H_j, then at least some of H_i's component parts are causally relevant for H_j's component parts as well.

But why exactly do we need assumption 2 in addition to assumption 1? The need for assumption 2 can be illustrated by means of our exemplary model in figure 1 as follows: Assume that the component variables of the two mechanisms represented would not be causally connected. This could, for example, be the case if there were no arrow between L_3 and L_4. However, we still assume that H_1 is a higher-level cause of H_2, which suggests that some changes in the cause mechanism (consisting of H_1, L_1, L_2, and L_3) can bring about changes in the target mechanism (consisting of H_2, L_5, and L_6). If there is no causal connection between the supervenience bases of H_1 and H_2, then the only causal relation that can account for this fact is $H_1 \rightarrow H_2$. In

[13] An analogous argument can be formulated against the efficacy of mental causes in the context of the causal exclusion problem (Gebharter [2017a]).

order to establish such a connection between H_1 and H_2, however, $H_1 \rightarrow H_2$ must pass the productivity test introduced in section 3. Hence, the arrow $H_1 \rightarrow H_2$ can be rendered redundant only if H_1's supervenience base and H_2's supervenience base are causally connected. There are two possibilities for such a causal connection: H_1's and H_2's supervenience bases have a common cause, or H_1's supervenience base is a cause of H_2's supervenience base. In order to account for the fact that H_2 changes can be brought about by producing changes in H_1, there must be a directed causal chain from H_1's supervenience base to H_2's supervenience base.

There is one final assumption one needs to make for a successful reduction of higher-level causation to interactions among component parts of the mechanisms involved. Let me briefly illustrate why one needs to make it. Assume that H_1 has four possible values (0, 1, 2, 3), H_2 has three possible values (0, 1, 2), and L_5 and L_6 each have two possible values (0, 1). (Since details about the other variables of the model are not required for the point I want to make here, I will ignore those variables for the moment.) Now assume that H_1, L_5, and L_6 are connected in such a way that the following conditionals hold:

$$\text{If } H_1 = 0, \text{ then } L_5 = 0 \text{ and } L_6 = 0.$$
$$\text{If } H_1 = 1, \text{ then } L_5 = 0 \text{ and } L_6 = 1.$$
$$\text{If } H_1 = 2, \text{ then } L_5 = 1 \text{ and } L_6 = 0.$$
$$\text{If } H_1 = 3, \text{ then } L_5 = 1 \text{ and } L_6 = 1.$$

In addition, we assume that H_2's parameters are defined as follows, where $i \in \{0, 1, 2, 3\}$ and $h_{2,0}$ stand short for $H_2 = 0$, $h_{2,1}$ stands short for $H_2 = 1$, and so on:

$$P(h_{2,0}|\mathbf{r}) = 1 \quad P(h_{2,1}|\mathbf{r}) = 0 \quad P(h_{2,2}|\mathbf{r}) = 0 \quad \text{if } \mathbf{r} = h_{1,i}, l_{5,0}, l_{6,0}$$
$$P(h_{2,0}|\mathbf{r}) = 0 \quad P(h_{2,1}|\mathbf{r}) = 1 \quad P(h_{2,2}|\mathbf{r}) = 0 \quad \text{if } \mathbf{r} = h_{1,i}, l_{5,1}, l_{6,0}$$
$$P(h_{2,0}|\mathbf{r}) = 0 \quad P(h_{2,1}|\mathbf{r}) = 1 \quad P(h_{2,2}|\mathbf{r}) = 0 \quad \text{if } \mathbf{r} = h_{1,i}, l_{5,0}, l_{6,1}$$
$$P(h_{2,0}|\mathbf{r}) = 0 \quad P(h_{2,1}|\mathbf{r}) = 0 \quad P(h_{2,2}|\mathbf{r}) = 1 \quad \text{if } \mathbf{r} = h_{1,i}, l_{5,1}, l_{6,1}$$

These parameters are chosen in such a way that whenever two of the three parents of H_2 take any of their values, the third parent's value is also determined. This means that no change in H_2's value can lead to a change in the probability distribution over one of H_2's parents conditional on H_2's remaining parents. Given these rather extreme constraints, the productivity test implies that the higher-level causal arrow $H_1 \rightarrow H_2$ is not productive. However, it also follows that the constitutional arrows $L_5 \dashrightarrow H_2$ and $L_6 \dashrightarrow H_2$ are not productive. In addition, deleting $H_1 \rightarrow H_2$ renders $L_5 \dashrightarrow H_2$ and $L_6 \dashrightarrow H_2$, and deleting $L_5 \dashrightarrow H_2$ or $L_6 \dashrightarrow H_2$ renders $H_1 \rightarrow H_2$ productive again. In cases such as the one constructed, the productivity test proposed in section 3 cannot identify the culprit; it only tells us that at least one of the competing arrows cannot represent an efficacious relation.

How should we interpret this result? First of all, note that such cases require a very specific fine-tuning of the model's parameters and can, thus, be expected to be extremely rare.[14] However, the logical possibility of such scenarios demands for an answer to the question raised. Since a supporter of a mechanistic analysis of higher-level causation aims at reducing higher-level causation in terms of mechanisms, it is advisable for her to favour the constitutional arrows over the causal ones.[15] In other words, she needs to assume that the higher-level behaviour's component parts rather than the higher-level causes are efficacious:

> Assumption 3 (Constitutional Priority): If constitutional as well as causal arrows pointing at a variable H describing some higher-level phenomenon do not pass the productivity test (described in section 3), then H's causal parents rather than H's constitutional parents are inefficacious with respect to H.

To summarize, the three assumptions introduced imply (to the background of the CBN framework) that higher-level causation reduces to causal interactions among component parts of the mechanisms involved. Assumption 1 warrants that for every higher-level phenomenon there exist several component parts that, together, render any causal relation between that phenomenon and its higher-level causes unproductive. Assumption 3 guarantees that those higher-level causes are actually inefficacious and that the causal relations do not fail the productivity test simply because one or more constitutively relevant parts are inefficacious. Finally, assumption 2 is required in order to account for the fact that the effect mechanism can be influenced by manipulating the cause mechanism. Let me stress again that the possibility of reducing higher-level causation in terms of mechanisms does not hinge on the specific structure of the model in figure 1. The results of this section apply to all CBN models involving mechanisms as long as one subscribes to assumptions 1–3.

Before pointing at some interesting consequences of the results obtained in this section, let me say a few words about the plausibility of assumptions 1–3. Assumption 1 nicely fits the new mechanist movement and probably every mechanist would subscribe to it: Higher-level phenomena are brought about by causal interactions among their constitutively relevant parts, and the behaviour of the system as a whole is connected to the behaviours of the component parts via supervenience. Note that the supervenience assumption is not only supported by mechanists, but also by almost everyone who believes in a layered view of the world (see, for example, Kim [2007]). Assumption 2 is quite plausible as well. If my higher-level decision is causally relevant for raising my left arm, then it seems also that the processes

[14] The problem should, for example, not arise in cases of non-deterministic causal dependencies. That the causal influences among variables to be analysed are indeterministic or at least pseudo-indeterministic is a common assumption in the causal modelling literature (see, for example, Spirtes et al. [1993]).

[15] Another reason to go for the constitutional arrows is that a mechanism's components are typically assumed to be difference makers. Recall from section 3 that the productivity test is basically a test for whether a parent of a variable is a difference maker for that variable. So if L_5 and L_6 are components of H_2, the arrows $L_5 \dashrightarrow H_2$ and $L_6 \dashrightarrow H_2$ should stay, and $H_1 \rightarrow H_2$ must go.

in my brain over which this decision supervenes are causally relevant for that phenomenon. My arm going up without any neurological causal basis would be more than miraculous. Finally, assumption 3 also seems plausible. It grants priority to constitutional relationships over causal relationships. This nicely fits the idea that constitution is a metaphysical relation that is more essential for an object or phenomenon than causation. While the presence of a cause might bring about quite different effects in different possible worlds due to different causal laws, the presence and behaviour of the constituents of a higher-level object or phenomenon determine the behaviour of this object or phenomenon in all possible worlds. It is, for example, hard to see how certain spatial changes in the molecules constituting my left arm could not result in a corresponding arm movement at the higher level. To summarize, it seems that all three assumptions—at least at first sight—are quite plausible and, thus, that Glennan's ([1996]) mechanistic account of higher-level causation might be supported by a CBN analysis such as the one developed in this section. But note that for a final verdict about the three assumptions' plausibility, a much more thorough investigation would be required. However, the main aim of this article consists in spelling out which assumptions one would have to subscribe to in order to reduce higher-level causation in terms of mechanisms.

5. Consequences for the Mechanism Debate

In this section I discuss three interesting consequences and connections of the results obtained in section 4 to other issues discussed in the literature on mechanisms. I discuss (i) the worry that a mechanistic analysis of higher-level causation might just reduce higher-level causation to lower-level causation, (ii) several consequences for the debate about inter-level causation in the presence of mechanisms, and (iii) how interventions on mechanisms might work according to the analysis provided in section 4.

5.1. Do mechanisms add anything to lower-level causal structure?

If Glennan ([1996]) is right and higher-level causation can actually be reduced to interactions between component parts of mechanisms, then one might wonder whether the analysis could not also be done without mechanisms. This is basically Psillos's ([2004]) worry briefly mentioned in section 1: At some point one runs out of mechanisms when analysing higher-level causation mechanistically, and it seems that, in the end, the causal structure at the lowest level is doing all the work in reducing higher-level causation. Glennan ([2011]) countered this objection by arguing that his mechanistic account and an interventionist treatment of causation depend on each other and, thus, stand and fall together[16]: While the mechanistic approach

[16] Psillos ([2004]) and Glennan ([2011]) both presume an interventionist understanding of causation.

depends on interventionism when it comes to accounting for causal interactions among component parts, interventionism, in turn, 'relies on the mechanical approach because the truth-conditions for counterfactuals depend upon the structure of mechanisms' (Glennan [2011], p. 806). Glennan goes on by arguing that interventionism faces a regress problem similar to the one of the mechanistic approach. I agree with Casini ([2016]) that Glennan's argumentation does not go through because both accounts run into different kinds of regresses: The mechanist runs into a vicious ontological regress, while the interventionist only faces a probably rather harmless conceptual regress.

In this section I am not particularly interested in how the mechanistic account and interventionism relate to each other. I still take Psillos's ([2004]) worry as a starting point, but approach it from a different angle. Interestingly, Glennan has already thought about the possible objection that his analysis might just reduce higher-level causation to causal relations at the fundamental physical level. He addressed it by pointing out that, 'Although the mechanism responsible for connecting two events may supervene upon lower-level mechanisms, and ultimately on mechanistically inexplicable [causal] laws of fundamental physics, it is not these laws that make the causal claim true; rather, it is the structure of the higher level mechanism and the properties of its parts' (Glennan [1996], p. 66). Which role mechanisms play exactly in making the higher-level causal claim true is, however, not explained in more detail, and there are several ways to interpret this quote. Here is one such possibility: The structure of the higher-level mechanism that is required in addition to the lower-level causal structure basically consists of the particular component parts whose causal interactions give rise to the higher-level phenomenon of interest. Information about this structure consists primarily in information about which parts are constitutively relevant for that phenomenon and about their causal interactions.[17] This kind of information is required to link the higher-level phenomenon to what is causally going on at the lower level and, thus, is essential for getting the reduction done. However, just making such a claim does not yet establish its truth. The reconstruction and the CBN model developed in section 4 can be used to further support it.

The reconstruction comes with clear criteria for when a higher-level causal relation is reducible. The model shows that more is needed than just the lower-level causal structure in order to reduce higher-level causation. In particular, one also needs to connect the higher-level causal structure with the lower-level causal structure by adding several dashed arrows representing constitutive relevance relations (assumption 1). Without these additional arrows, the productivity test would yield that $H_1 \rightarrow H_2$ is productive (and, thus, the reduction would fail), and the Markov condition would be violated because there would be probabilistic dependencies between

[17] I am indebted to an anonymous referee for pointing out that the structure of a mechanism meant here might also include the spatiotemporal ordering of the component parts. Since the point I want to make in this subsection does not require this kind of structural information, I will abstract from it for now.

variables at different levels (for example, between H_2 and L_5 and between H_2 and L_6) not accounted for by any connection in the graph. This shows that a successful and consistent reduction requires mechanisms as mediators between the behaviours of higher- and lower-level variables. Note that acquiring information about the structure of mechanisms, information about which parts of a system are constitutively relevant for the overall behaviour of that system, is not a trivial task. Constitutive relevance relations cannot be read off the variables and causal relations represented in the model. Decomposing mechanisms and finding out which parts are actually components are quite demanding processes that cannot be done on a priori grounds. This is further supported by the fact that there are currently several quite sophisticated theories of constitutive relevance and approaches for constitutive relevance discovery on the market (see, for example, Craver [2007]; Harbecke [2010], [2015]; Baumgartner and Gebharter [2016]; Baumgartner and Casini [2017]; Gebharter [2017c]).

5.2. Mechanisms and inter-level causation

Several authors have recently argued for or against inter-level causation in the presence of mechanisms. Leuridan ([2012]), for example, is sympathetic toward interpreting constitutive relevance as a bidirectional causal relation between variables at different levels, and Krickel ([2017], p. 543) argues for the view that mechanistic hierarchies do not exclude inter-level causation if one takes seriously that 'the relata of the mechanistic level relation are acting entities'. Authors such as Craver and Bechtel ([2007], Kistler ([2009]), and Romero ([2015]), on the other hand, claim that inter-level causation can be reduced to truths about constitutive relevance relations and lower-level causal structure. In what follows, I will explore how the analysis provided in section 4 may be used in order to support the reductionist camp.

Let me illustrate this by means of the toy example introduced in section 4. Assume, in addition to the assumptions already in place, that the model features the two inter-level causal relations $L_1 \rightarrow H_2$ and $H_1 \rightarrow L_5$ (fig. 2).[18] A reduction in the sense of Craver and Bechtel ([2007] would, again, consist in showing that these two arrows are redundant and can, in principle, be deleted from the model and that all the work these arrows seem to do is actually done by constitutive relevance relations and the lower-level causal structure. Let us start with bottom-up causation and with the arrow $L_1 \rightarrow H_2$. With assumption 1 it follows that $L_1 \rightarrow H_2$ is unproductive for exactly the same reasons the higher-level causal arrow $H_1 \rightarrow H_2$ turned out as unproductive in section 4 (L_1 and H_2 are independent conditional on $\mathbf{Par}(H_2)\backslash\{L_1\} = \{L_5, L_6\}$); and for the few cases in which the productivity test would output $L_1 \rightarrow H_2$ as well as one or both of the constitutional arrows $L_5 \dashrightarrow H_2$ and $L_6 \dashrightarrow H_2$ as unproductive, assumption 3 tells us that the bottom-up cause should be considered as inefficacious. Finally, one needs to guarantee that manipulating L_1 can bring

[18] Since the focus is on inter-level causation now, I have deleted the higher-level causal arrow $H_1 \rightarrow H_2$.

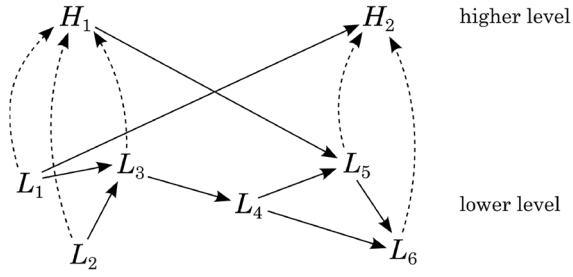

Figure 2. Toy model for exploring the reducability of inter-level causation in terms of mechanisms.

about changes in H_2. If the bottom-up arrow $L_1 \rightarrow H_2$ should turn out as unproductive, one needs to account for this fact by assuming that one or more directed chains from L_1 to H_2 go through H_2's supervenience base $\{L_5, L_6\}$. This can be guaranteed by subscribing to the following assumption:

> Assumption 4 (Lower-Level Causal Relevance 2): If a lower-level phenomenon described by a variable L_i is causally relevant for a higher-level phenomenon described by a variable H_j, then L_i is causally relevant for at least some of H_j's component parts as well.

Assumption 4 has the same function for reducing bottom-up causation as assumption 2 has for reducing higher-level causation. It seems plausible for almost the same reasons, too[19]: It would be miraculous if there were bottom-up causes without a corresponding lower-level causal story to be told. Assume, for example, that certain processes in my brain caused the raising of my left arm. Then, surely, there is a causal story to be told about how these brain processes caused the muscles in my left arm to contract in specific ways that constitute the raising of my left arm. A common cause path connecting these brain processes and the muscle contractions in my left arm would not be enough to explain the dependence of the raising of my left arm on these brain processes.

Reducing the top-down causal arrow $H_1 \rightarrow L_5$ is more cumbersome. Assumptions 1 and 3 that, together, render causes inert in the presence of constitutive relevance relations have no bearing on lower-level variables such as L_5. To render H_1 inefficacious with respect to L_5, one needs to subscribe to additional assumptions. Here is a possible candidate:

> Assumption 5 (Extended Productivity): $X_i \rightarrow X_j$ (or $X_i \dashrightarrow X_j$) is productive if and only if X_j depends on X_i conditional on $\mathbf{Par}(X_j)\backslash\{X_i\}$. If X_i is a higher-level direct cause of X_j, then X_j must also depend on X_i conditional on the set $\mathbf{V}'$ of X_i's constitutive ancestors.

[19] Like in the case of the other assumptions, a more thorough evaluation of assumption 4 would go beyond the scope of this article.

Assumption 5 is basically an extended version of the productivity test introduced in section 3. It states that for a causal arrow to be productive, it does not always suffice that the two variables the arrow connects cannot be screened off by the effect variable's other parents. If the cause variable is a higher-level variable, then the variables constitutively relevant for the cause variable have to be considered as well. Assumption 5 seems plausible on first sight. The question on which everything seems to hinge is whether the behaviour of the system at the higher level (H_1) can contribute anything to the causal work done by the mechanism's component parts (L_1, L_2, and L_3).

Unfortunately, it turns out that assumption 5 is not sufficient to reduce top-down causation in all cases. The problem is that though the set of constituents $\{L_1, L_2, L_3\}$ clearly screens H_1 off from L_5, assumption 5 does not exclude probability flow along the arrow $H_1 \rightarrow L_5$ in general. In particular, the causal arrow $H_1 \rightarrow L_5$ might be responsible for a probabilistic dependence between H_1 and L_5 conditional on L_4. As a consequence, $H_1 \rightarrow L_5$ would not be redundant since $H_1 \rightarrow L_5$ would pass the productivity test and, thus, deleting $H_1 \rightarrow L_5$ would violate the Markov condition. Since constitution is assumed to formally work exactly like causation, the situation is equivalent to a case in which a variable X_1 is a deterministic direct cause of another variable X_2 and, at the same time, a probabilistic direct cause of another variable X_3, where both X_2 and X_3 are probabilistic direct causes of a fourth variable X_4. In such a case, nothing logically excludes the possibility that X_2 is dependent on X_4 conditional on its other parent X_3, meaning that nothing excludes that $X_2 \rightarrow X_4$ passes the original productivity test and is actually required to avoid a violation of the Markov condition.

Another—and, as far as I can see, the only—possible way to go consists in subscribing to a variant of the closure of the physical domain, which says that if there are physical causes for a physical event, then citing all of these physical causes gives us all the information we might get for whether the physical event of interest actually occurs (see, for example, Kim [2007]). In light of this assumption, higher-level causes cannot provide any additional information. This assumption can be adopted for levels of mechanisms:

> Assumption 6 (Causal Closure of Levels): If there exist lower-level direct causes $L_1, \ldots, L_n$ as well as higher-level direct causes $H_1, \ldots, H_m$ for a lower-level variable L_{n+1}, then $H_1, \ldots, H_m$ do not pass the productivity test introduced in section 3.

Assumption 6 amounts to the claim that higher-level causes always fail the original productivity test in the presence of same-level causes. So while the reduction of higher-level as well as bottom-up causation came quite naturally due to assumptions that have—at least at first glance—some plausibility, the reducibility of top-down causation needs to be assumed by brute force. In any case, it seems clear that to

reduce top-down causation in terms of mechanisms, a variant of the causal closure of levels is required. If this is so, and subscribing to assumption 6 amounts to nothing over and above begging the question, then the question of whether there is top-down causation in the presence of mechanisms becomes, in the end, an empirical question; it entirely depends on whether there are overall behaviours of mechanistic systems that pass the productivity test.

Note that even if one is ready to subscribe to the causal closure of levels or is able to establish it on empirical grounds, one still has to guarantee that lower-level causal arrows have priority over competing top-down causal arrows if both kinds of arrows turn out to be unproductive and the deletion of one of these arrows would render the others productive again. One could do this analogously to how constitutional arrows were prioritized by assumption 3:

> Assumption 7 (Lower-Level Causal Priority): If lower-level as well as higher-level causal arrows pointing at a variable X describing some phenomenon do not pass the productivity test (described in section 3), then X's higher-level causal parents rather than X's lower-level causal parents are inefficacious with respect to X.

Assumption 7 has some plausibility. If one has to give up causal efficacy either for the higher- or for the lower-level cause, almost everyone would decide in favour of the lower-level cause. This is, for example, an assumption commonly accepted in the debate surrounding causal exclusion (see Kim [2007]). It can also be supported by the arm-raising example mentioned in support of assumption 2 in section 4.

To summarize, the analysis provided in section 4 reduces bottom-up causation to lower-level causation plus constitutive relevance if one is ready to accept assumption 4. In order to reduce top-down causation as well, one has to make additional assumptions such as the more problematic assumptions 6 and 7. Supporters of reductive accounts of inter-level causation, such as Craver and Bechtel ([2007], might be able to strengthen their position by coming up with plausible arguments for assumptions 4 and 7, and by supporting assumption 6 on empirical grounds.

5.3. Interventions into mechanistic hierarchies

How interventions work in the presence of mechanistic hierarchies is still not well understood. Craver ([2007]), for example, suggests that it might be possible to surgically intervene on the whole as well as on any part of a mechanism, where interventions are to be understood in the sense of Woodward ([2003], p. 98).[20] Baumgartner and Gebharter ([2016]) and Romero ([2015]) show that there is no way to surgically intervene on a mechanism's overall behaviour as long as one accepts that the higher-level behaviour of a mechanism supervenes on the behaviours of its

[20] An intervention on a variable is 'surgical' if it causes no other variable directly.

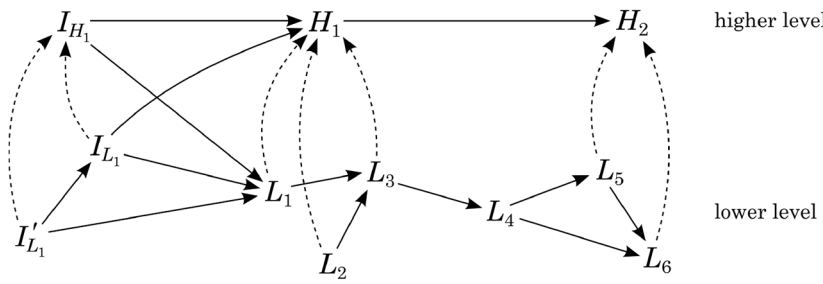

Figure 3. Toy model for exploring how interventions into mechanistic hierarchies might work theoretically.

component parts. They argue that all interventions on the mechanism as a whole are actually fat-handed, meaning that they are common causes of the higher-level behaviour as well as of the behaviours of some component parts. Let us see whether we can add something to that debate.

For illustrating how interventions might work according to the analysis provided in section 4, let us have another look at the exemplary model introduced in that section. Let us introduce a higher-level intervention variable I_{H_1} for H_1.[21] According to the mechanistic account of higher-level causation, there must be some lower-level component parts for I_{H_1}. Let us assume that I_{L_1} and I'_{L_1} model two of them, that I'_{L_1} is a direct cause of I_{L_1}, and that both lower-level variables are direct causes of L_1. In addition, we add the causal arrows $I_{H_1} \rightarrow L_1$ and $I_{L_1} \rightarrow H_1$ and arrive at the structure depicted in figure 3.[22]

As we have seen, subscribing to assumptions 1–4 renders possible causes of higher-level variables such as H_1 causally inert. (Note that the problematic assumption 6 and also assumption 7 are not required for this result.) Under these conditions, the result of Baumgartner and Gebharter ([2016]) and Romero ([2015]) that the behaviour of a mechanism at the higher level can only be manipulated while, at the same time, intervening on some of that mechanism's component parts turns out to be correct. But we can also add something to this result: Actually, all the work is done by the lower-level variables. So interventions on mechanisms are in some sense fat-handed. They are, however, not fat-handed in the sense of Baumgartner, Gebharter, and Romero. By intervening on a mechanism, one does not intervene on higher- and lower-level variables via different effective causal paths. One rather influences higher-level variables in virtue of exerting causal influence on the mechanism's constitutively relevant parts.

[21] I use a very weak notion of 'intervention variable' here. For I_X to count as an intervention variable for another variable X, I will only assume that I_X is a direct cause of X, or of one or several of X's constitutional ancestors. 'Under an intervention on a mechanism' I will understand as 'an intervention on a variable of that mechanism'.

[22] The I-variables should represent all the possible ways that a mechanism consisting of H_1, L_1, L_2, L_3 could be manipulated. They will stand proxy for all kinds of possible intra- and inter-level interventions on mechanisms and serve as a theoretical tool for investigating how such interventions on mechanistic systems might actually work.

Our analysis can be used to shed new light on the following two related questions: (i) How can post-intervention distributions be computed in CBN models capturing mechanistic hierarchies, and (ii) How can constitutive relevance relations be inferred on the basis of interventions? Let us start with question (i). The effect of setting a variable X_i to value x_i by means of an intervention in a purely causal BN can be computed as follows (see Pearl [2000]): As a first step, one deletes (or 'breaks') all the arrows pointing at X_i. Next, one applies equation 1 to the resulting truncated graph in order to compute the conditional probabilities $P(\mathbf{V}'|X_i = x_i)$ for any set $\mathbf{V}' \subseteq \mathbf{V}$ of variables of interest. These conditional probabilities are then identified with the post-intervention distribution over $\mathbf{V}'$. This procedure guarantees that intervening on a variable X_i can have an effect on X_i's descendants (its effects), but not on its ancestors (its causes) or causally independent variables.

Here is a first sketch of how this technique could also be used in order to compute the effects of interventions on mechanisms in systems spanning no more than two levels. According to our analysis, there are at least two different kinds of background knowledge one might have when intervening on mechanisms that nicely correspond to the different kinds of experiments biologists and neuroscientists carry out when they study how manipulating multi-level systems might influence behaviours at different levels (see, for example, Craver [2007]; Kästner [2018]). For each kind of knowledge or experiment, there is a different way to compute the corresponding post-intervention distribution.

> Bottom-up experiment: This kind of experiment corresponds to a situation in which the experimenter knows exactly which lower-level variable(s) she actively changes, but not the effect of the intervention on the system's behaviour at the higher level. In this case, one just applies the arrow-breaking technique described: One first deletes the arrows into those lower-level variables that are the targets of the intervention and then computes the post-intervention distribution on the basis of the resulting graph and equation 1. By doing this, one can learn about how changing the values of these lower-level variables probabilistically influences the overall behaviour of the mechanism or the higher- or lower-level behaviours of other mechanisms causally connected to it.

> Top-down experiment: One might, however, know that an experimental intervention produced a specific higher-level behaviour $Y = y$, but not which parts of the system the intervention influenced directly and which of these parts were actually responsible for that higher-level behaviour. According to our analysis, this corresponds to a top-down experiment. The effect of such an intervention on any variable X's value x (describing, for example, a specific behaviour of a lower-level part) can be computed as follows: First, consider all the possible ways the mechanism's lower-level variables could have been influenced by a (possibly fat-handed) intervention. For a set $\mathbf{L}$ of m lower-level

variables there will be $|\mathcal{P}(\mathbf{L})| - 1$ such possibilities. Let us label them $S_1, \ldots, S_n$. Next, produce the truncated graph $\mathbf{G}_i$ for each possible scenario S_i by deleting the arrows into the variables intervened on in that scenario. For each pair S_i and $\mathbf{G}_i$, list all those possible value combinations $\mathbf{l}$ of the variables intervened on that are compatible with $X = x$ and $Y = y$. For each of them, compute the conditional probability $P(x|y, \mathbf{l})$ on the basis of equation 1 and the truncated graph $\mathbf{G}_i$. Record all of these probabilities in a set $\mathbf{P}$. After having done this for every pair S_i and $\mathbf{G}_i$, compute the average of all the probabilities recorded in $\mathbf{P}$ and identify the post-intervention distribution with this average.[23] The suggested procedure corresponds to observing $Y = y$ and going through all the possible ways this observation could have been produced by interventions on parts of the mechanism. On its basis one can learn how an experiment might have influenced a mechanism's parts as well as how the intervention and the observation $Y = y$ might influence the probability distributions over variables of other mechanisms causally connected to the first one.

Our analysis shows that whether an experiment counts as a bottom-up or a top-down experiment depends on the experimenter's background knowledge and explanatory goals. The result that these different kinds of experiments correspond to shifts in epistemic perspectives also nicely fits Kästner's ([2018], p. 77) observation that 'By fleshing out the perspectival aspect inherent in the mechanistic view we gain a better understanding of the explanatory practices in science'. There is clearly also space for several kinds of mixed experiments that combine the strategies for bottom-up and top-down experiments sketched.

Let me now come to the related point (ii) mentioned: constitutive relevance discovery. While it is not possible to surgically intervene on a mechanism's higher-level behaviour as a whole—which is exactly what Craver's ([2007]) prominent mutual manipulability criterion for constitutive relevance relations demands—in particular, bottom-up experiments as characterized turn out to be useful for identifying constitutive relevance relations. According to the analysis developed, the following procedure would, for example, be sufficient to do the job: First, decompose the system (described by H) into several parts (described by $L_1, \ldots, L_n$). Next, pick one of the lower-level variables L_i. Now check whether setting L_i to one of its values by intervention leads to any change in the probability distribution of H while the values of all other lower-level variables L_j are fixed by interventions. If H wiggles in such a bottom-up experiment, then L_i must be a component part of H or a cause of a component part not included in $L_1, \ldots, L_n$. This procedure requires, of course, that the lower-level variables can be controlled independently. More sophisticated discovery procedures for constitutive relevance relations might be developed on the basis of

[23] This reflects the situation in which the experimenter considers all possible intervention scenarios S_i as equally likely. If one has a more concrete idea of how likely it is that $Y = y$ has been produced by intervening on specific parts of the system, one could use the weighted average instead. The weights would then correspond to the experimenter's confidence in particular intervention hypotheses.

the analysis presented in this article. The analysis might also be compatible with some of the approaches already on the market. The algorithmic procedure discussed in (Gebharter [2017c]), for example, should be compatible.

6. Conclusion

In this article I provided a reconstruction of Glennan's ([1996]) mechanistic account of higher-level causation in terms of CBNs. Reducing higher-level causation in a CBN setting means to specify general conditions under which higher-level causal arrows can be rendered redundant in all kinds of models involving mechanisms. In section 4 I specified three such assumptions: Higher-level phenomena can always be decomposed in component parts in such a way that they supervene on the behaviours of these component parts (assumption 1). If a higher-level phenomenon is causally relevant for another one, then at least some component parts of the former must be causes of some component parts of the latter (assumption 2). And, finally, constitutional arrows are always prioritized over causal arrows competing for productivity (assumption 3). Since all three assumptions seem—at least at first glance—quite plausible from a mechanist's perspective, the CBN reconstruction might be used to provide additional support for Glennan's mechanistic analysis of higher-level causation from the perspective of an empirically informed theory of causation.

In section 5 I pointed to some interesting consequences of the findings of section 4 for other issues discussed in the mechanism literature. The analysis provided (i) shows that and also why lower-level causal structure as well as mechanisms are required for reducing higher-level causation. It also (ii) shows that a reductive analysis of inter-level causation in terms of causal and constitutive relevance relations such as Craver and Bechtel's ([2007]) follows from the CBN machinery if one is ready to subscribe to three additional assumptions (assumptions 4, 6, and 7). Finally, the analysis sheds new light on (iii) how interventions on mechanisms might actually work, how post-intervention distributions could be computed for systems involving mechanisms, and how different kinds of experiments could be used to study mechanisms and to identify a mechanism's constitutively relevant parts. This is, of course, only the tip of the iceberg, and much more work needs to be done in order to develop a full-fledged account of interventions on mechanisms and reliable methods for the discovery of causal and constitutive relevance relations in mechanisms.

Acknowledgements

I would like to thank Michael Baumgartner, Lorenzo Casini, Christian J. Feldbacher-Escamilla, Stuart Glennan, Mario Günther, Stephan Hartmann, Lena Kästner, Beate Krickel, Gerhard Schurz, Reuben Stern, and Jim Woodward for their input and important discussions. Thanks also to three anonymous referees for many helpful comments.

Munich Center for Mathematical Philosophy
Ludwig Maximilian University of Munich
Munich, Germany
alexander.gebharter@gmail.com

References

Baumgartner, M. and Casini, L. [2017]: 'An Abductive Theory of Constitution', *Philosophy of Science*, **84**, pp. 214–33.

Baumgartner, M. and Gebharter, A. [2016]: 'Constitutive Relevance, Mutual Manipulability, and Fat-Handedness', *British Journal for the Philosophy of Science*, **67**, pp. 731–56.

Bechtel, W. and Abrahamsen, A. [2005]: 'Explanation: A Mechanist Alternative', *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, **36**, pp. 421–41.

Casini, L. [2016]: 'Can Interventions Rescue Glennan's Mechanistic Account of Causality?', *British Journal for the Philosophy of Science*, **67**, pp. 1155–83.

Casini, L., Illari, P. M., Russo, F. and Williamson, J. [2011]: 'Models for Prediction, Explanation and Control: Recursive Bayesian Networks', *Theoria*, **26**, pp. 5–33.

Clarke, B., Leuridan, B. and Williamson, J. [2014]: 'Modelling Mechanisms with Causal Cycles', *Synthese*, **191**, pp. 1651–81.

Craver, C. [2007]: *Explaining the Brain*, Oxford: Clarendon.

Craver, C. and Bechtel, W. [2007]: 'Top-Down Causation without Top-Down Causes', *Biology and Philosophy*, **22**, pp. 547–63.

Eronen, M. I. [2015]: 'Levels of Organization: A Deflationary Account', *Biology and Philosophy*, **30**, pp. 39–58.

Gebharter, A. [2014]: 'A Formal Framework for Representing Mechanisms?', *Philosophy of Science*, **81**, pp. 138–53.

Gebharter, A. [2017a]: 'Causal Exclusion and Causal Bayes Nets', *Philosophy and Phenomenological Research*, **95**, pp. 353–75.

Gebharter, A. [2017b]: *Causal Nets, Interventionism, and Mechanisms: Philosophical Foundations and Applications*, Cham: Springer.

Gebharter, A. [2017c]: 'Uncovering Constitutive Relevance Relations in Mechanisms', *Philosophical Studies*, **174**, pp. 2645–66.

Gebharter, A. and Koch, D. [2021]: 'Combining Causal Bayes Nets and Cellular Automata: A Hybrid Modeling Approach to Mechanisms', *British Journal for the Philosophy of Science*, **72**, pp. 839–64.

Glennan, S. [1996]: 'Mechanisms and the Nature of Causation', *Erkenntnis*, **44**, pp. 49–71.

Glennan, S. [2002]: 'Rethinking Mechanistic Explanation', *Philosophy of Science*, **69**, pp. S342–53.

Glennan, S. [2011]: 'Singular and General Causal Relations: A Mechanist Perspective', in P. M. Illari, F. Russo and J. Williamson (*eds*.), *Causality in the Sciences*, Oxford: Oxford University Press, pp. 789–817.

Harbecke, J. [2010]: 'Mechanistic Constitution in Neurobiological Explanations', *International Studies in the Philosophy of Science*, **24**, pp. 267–85.

Harbecke, J. [2015]: 'The Regularity Theory of Mechanistic Constitution and a Methodology for Constitutive Inference', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **54**, pp. 10–19.

Illari, P. M. and Williamson, J. [2012]: 'What Is a Mechanism? Thinking about Mechanisms across the Sciences', *European Journal for Philosophy of Science*, **2**, pp. 119–35.

Kaiser, M. I. [2016]: 'On the Limits of Causal Modeling: Spatially-Structurally Complex Biological Phenomena', *Philosophy of Science*, **83**, pp. 921–33.

Kästner, L. [2018]: 'Integrating Mechanistic Explanations through Epistemic Perspectives', *Studies in History and Philosophy of Science Part A*, **68**, pp. 68–79.

Kim, J. [2007]: *Physicalism, or Something Near Enough*, Princeton, NJ: Princeton University Press.

Kistler, M. [2009]: 'Mechanisms and Downward Causation', *Philosophical Psychology*, **22**, pp. 595–609.

Koch, D., Eisinger, R. S. and Gebharter, A. [2017]: 'A Causal Bayesian Network Model of Disease Progression Mechanisms in Chronic Myeloid Leukemia', *Journal of Theoretical Biology*, **433**, pp. 94–105.

Krickel, B. [2017]: 'Making Sense of Interlevel Causation in Mechanisms from a Metaphysical Perspective', *Journal for General Philosophy of Science*, **48**, pp. 453–68.

Krickel, B. [2018]: 'Saving the Mutual Manipulability Account of Constitutive Relevance', *Studies in History and Philosophy of Science Part A*, **68**, pp. 58–67.

Leuridan, B. [2012]: 'Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms', *British Journal for the Philosophy of Science*, **63**, pp. 399–427.

Machamer, P., Darden, L. and Craver, C. [2000]: 'Thinking about Mechanisms', *Philosophy of Science*, **67**, pp. 1–25.

Murray-Watters, A. and Glymour, C. [2015]: 'What's Going on inside the Arrows? Discovering the Hidden Springs in Causal Models', *Philosophy of Science*, **82**, pp. 556–86.

Pearl, J. [2000]: *Causality*, Cambridge: Cambridge University Press.

Psillos, A. [2004]: 'A Glimpse of the Secret Connexion: Harmonizing Mechanisms with Counterfactuals', *Philosophy of Science*, **12**, pp. 288–319.

Reichenbach, H. [1956]: *The Direction of Time*, Berkeley: University of California Press.

Romero, F. [2015]: 'Why There Isn't Inter-level Causation in Mechanisms', *Synthese*, **192**, pp. 3731–55.

Russo, F. and Williamson, J. [2007]: 'Interpreting Causality in the Health Sciences', *International Studies in the Philosophy of Science*, **21**, pp. 157–70.

Schaffer, J. [2016]: 'Grounding in the Image of Causation', *Philosophical Studies*, **173**, pp. 49–100.

Schurz, G. and Gebharter, A. [2016]: 'Causality as a Theoretical Concept: Explanatory Warrant and Empirical Content of the Theory of Causal Nets', *Synthese*, **193**, pp. 1073–103.

Spirtes, P., Glymour, C. and Scheines, R. [1993]: *Causation, Prediction, and Search*, Dordrecht: Springer.

Weber, M. [2016]: 'On the Incompatibility of Dynamical Biological Mechanisms and Causal Graphs', *Philosophy of Science*, **83**, pp. 959–71.

Woodward, J. [2002]: 'What Is a Mechanism? A Counterfactual Account', *Philosophy of Science*, **69**, pp. S366–77.

Woodward, J. [2003]: *Making Things Happen*, Oxford: Oxford University Press.