CAUSAL MODELING IN MULTILEVEL SETTINGS: A NEW PROPOSAL

Thomas Blanchard & Andreas Hüttemann

University of Cologne

An important question for the causal modeling approach is how to integrate non-causal dependence relations such as asymmetric supervenience into the approach. The most prominent proposal to that effect (due to Gebharter) is to treat those dependence relationships as formally analogous to causal relationships. We argue that this proposal neglects some crucial differences between causal and non-causal dependencies, and that in the context of causal modeling non-causal dependence relationships should be represented as *mutual* dependence relationships. We develop a new kind of model – "hybrid models" - based on this suggestion, and formulate a set of axioms for those models. Our formalism has important implications for Kim's exclusion problem: whereas Gebharter's framework vindicates Kim's causal exclusion objection against nonreductive physicalism, our framework has no such implication, and can help non-reductive physicalists vindicate the efficacy of high-level properties. A further benefit of our formalism is that it yields a natural and plausible way of thinking about interventions in multi-level contexts.

An important question for the causal modeling approach concerns how to model settings that involve variables representing states of affairs at different levels of reality and which are therefore related by non-causal dependence relationships such as supervenience. The most prominent proposal is due to Gebharter (2017a, 2017b, 2022), who proposes that non-causal modeling dependencies should be treated as formally analogous to asymmetric causal relationships. We argue, however, that Gebharter's account is ill-motivated and leads to serious problems, especially when applied to contexts involving part-whole relationships. We offer an alternative extension of the causal modeling framework which treats supervenience and part-whole relationships as *mutual* dependence relationships, and relies on a new type of models which we call *hybrid models*. This framework, we argue, better captures certain key differences between causal and non-causal dependence relationships, avoids a number of worries that arise within Gebharter's framework, and makes better sense of how causal and non-causal dependencies relate to each other in multilevel settings.

We start with a summary of the causal modeling framework (§1) and Gebharter's extension of it to non-causal dependencies (§2). In §3 we argue that his framework is ill-motivated, and then go on to develop an alternative extension based on a new type of model which we call *hybrid models*, in which supervenience and part-whole relationships are treated as symmetric dependence relationships (§§4-5). In the second part of the paper, we compare the implications of the two frameworks for the exclusion problem (§§6-7). We show that in contrast with Gebharter's approach, our framework does not support Kim's exclusion argument, and can help non-reductive physicalists vindicate the causal efficacy of multiply realizable properties. We also argue that it makes better sense of causation in contexts involving part-whole relationships than Gebharter's framework, as the latter leads to the disastrous result that all macroscopic phenomena are always causally excluded by their parts. Finally, we argue that our framework can help make progress on the vexed questions of how to model interventions in multi-level settings (§8).

1. Causal Modeling and Causal Bayes Nets

Causal models are mathematical devices that represent the causal structure of a set of random variables. In the causal modeling literature, the type of model most used and discussed is the causal Bayes net, or CBN (see especially Spirtes *et al*., 2000; Pearl, 2009). CBNs are composed of three elements. The first is the set **V** of random variables we are interested in studying.[1] Those random variables may represent possible values of a quantity, properties of an individual or a system, etc. **V** is assumed to be causally sufficient, i.e. every common cause of two variables in **V** is also in **V**. The second element is a set **E** of *directed edges* (arrows). On the standard interpretation of those arrows, $X \rightarrow Y$ indicates that X is a direct cause (or "parent") of Y: that is, X causally influences Y, and the influence is not mediated by any other variable in **V**. A *path* is a sequence of variables such that an arrow exists between each variable and the next in the sequence (for example, $X \rightarrow Y \leftarrow Z$). If all arrows point in the same direction the path is *directed*. If there is a directed path from X to Y, Y is a descendant of X. (By convention every variable is a descendant of itself.) Directed paths from a variable to itself are prohibited, i.e. there are no causal cycles. Together, **V** and **E** form a *directed acyclic graph* (DAG) **G**. For example, the DAG in Figure 1 represents a structure where X directly causes Y and Z, and W directly causes Z.

[1] Formally, random variables are functions defined on the sample space of a probability space that are measurable with respect to that probability space.

[Figure 1 here]

The third component of a CBN is a probability distribution P over $\mathbf{V}$. A key idea of causal modeling is that the causal structure of $\mathbf{V}$ puts substantial constraints on P that can be captured in axioms relating $\mathbf{G}$ and P. Chief among them is the *causal Markov condition* (CMC). In its most familiar version it says that every variable is independent of its non-descendants given its parents. Pearl (2009) offers a different – though logically equivalent (see Geiger & Pearl, 1989) – formulation in terms of *d-separation*, which relies on the notion of *collider*:

> **collider**: A non-endpoint variable Y on a path in a DAG is a *collider* on that path iff the two directed edges preceding and succeeding Y have arrowheads pointing at Y (i.e., $X \rightarrow Y \leftarrow Z$ for some X and Z on the path).

Intuitively, a variable is a collider when it cannot transmit influence on the path, so that a path between X and Z that contains a collider does not induce a correlation between X and Z. A variable thus counts as a collider on a path when it is caused by two variables on that path, as two variables that have a common effect are not thereby correlated. *d-separation* is then defined as follows:

> **d-separation**: Two variables X and Y in a DAG are *d-separated* by a (possibly empty) set of variables $\mathbf{Z}$ iff, for every path between X and Y, (a) there exists a non-collider on the path that is in $\mathbf{Z}$ or (b) the path contains a collider, and neither that collider nor any of its descendants is in $\mathbf{Z}$.

$\mathbf{G}$ and P then satisfy CMC just in case

> **CMC**: For every X, Y, $\mathbf{Z}$ in $\mathbf{G}$, if $\mathbf{Z}$ d-separates X and Y, then $\mathrm{P}(Y/X\&\mathbf{Z})=\mathrm{P}(Y/\mathbf{Z})$.[2]

For instance, in Figure 1 CMC entails *inter alia* that Y is independent of Z and W given X, and that X and W are unconditionally independent. It does *not* entail that X and W are independent given Z, as Z is a collider on the path between them. CMC by itself puts few constraints on the

[2] Probabilistic statements that include only variables (or sets of variables) should be read as universally quantifying over values of the variables. That is, $(\mathrm{Y}/\mathrm{X}\&\mathbf{Z})=\mathrm{P}(\mathrm{Y}/\mathbf{Z})$ means that for all values x of X, all values y of Y, and all combinations of values of $\mathbf{Z}$, $\mathrm{P}(Y=y/X=x\&\mathbf{Z}=\mathbf{z})=\mathrm{P}(Y=y/\mathbf{Z}=\mathbf{z})$.

relations between causal and probabilistic structure, as it never forbids the inclusion of an edge in a graph. (If a DAG satisfies CMC then any DAG obtained by adding arrows to it does as well.) Thus further axioms are needed to prevent the inclusion of superfluous edges. The least demanding such axiom is causal minimality (CMIN). **G** and *P* satisfy minimality when

> **CMIN**: **N**o proper subgraph of **G** (i.e. no graph obtained by removing edges from **G**) satisfies the CMC with respect to *P*.

CMIN implies that an edge $X \rightarrow Y$ is warranted iff *Y* probabilistically depends on *X* given *Y*'s other parents. (In that sense, it embodies a "difference-making" approach to causation.) For instance, for the graph of Figure 1 to satisfy CMIN, *W* and *Z* must be correlated given *X*, since otherwise the graph obtained by removing the arrow $W \rightarrow Z$ would still satisfy CMC. (In addition to CMIN, it is customary to also posit a logically stronger axiom of *causal faithfulness* on CBNs. We do not assume faithfulness here, for reasons that will appear below.)

CBNs, as we will see later, are not the only tool in the causal modeler's toolkit, but they are an enormously useful one. The formalism of CBNs provides the foundation for sophisticated and powerful causal inference techniques (Spirtes *et al*., 2000), and has been used to clarify the content of counterfactuals about causal systems (2012), to model the effects of interventions into those systems (Woodward, 2003), and to address various questions in decision theory (e.g. Stern, 2019) and the psychology of causal reasoning (Sloman, 2005). But it has an important limitation. On the standard interpretation of arrows as representing direct causes, the CBN axioms fail in multi-level settings - settings that involve variables which represent states of affairs at different levels of reality, and which thereby stand in relationships of non-causal dependence to one another. To see this, note that if *X* and *Y* are causally unrelated (i.e. one does not cause the other), CMC entails that they must be uncorrelated conditional on the set of their common causes. But if *X* and *Y* stand in a relationship of non-causal dependence (e.g. supervenience) they may and typically will be correlated, even holding their common causes fixed. So if we want to apply the causal modeling framework to multi-level contexts, we need to extend the CBN formalism to those contexts, or identify some other type of causal model suited to those contexts.

But why should we want to apply the causal modeling framework to multi-level settings in the first place? One reason has to do with mechanisms. Mechanistic explanation involves understanding the causal effects of a composite entity in terms of the behavior of its parts (see e.g. Craver, 2007). In the sciences mechanistic explanation is ubiquitous, and scientists routinely rely on mechanistic knowledge to predict the behavior of a system and the effects of potential interventions into it. Since causal modeling is arguably our best framework for understanding explanation and prediction in the sciences, it would be theoretically beneficial if we could bring it to bear on mechanistic reasoning (Casini et al., 2011). But this requires finding a way to integrate non-causal dependencies between parts and wholes in causal models.

Another reason is that the status of high-level causal relationships in multi-level settings raises important and difficult questions. In particular, the implications of the standard non-reductive physicalist picture of inter-level relationships for high-level causation remain actively debated. Suppose that mental and other high-level properties supervene on but are nevertheless distinct from their physical realizers, as non-reductive physicalists hold. Kim's (1998, 2005) famous exclusion argument seeks to show that in this picture high-level properties must be causally idle, as all the causal work is done by their realizers. In response, many philosophers (e.g. Bennett, 2003) maintain that high-level and low-level properties can both be causes (and must then address the worry that this implies a problematic kind of ubiquitous overdetermination). Other non-reductive physicalists such as List and Menzies (2009) turn the tables on Kim's argument and claim that it is typically high-level properties that causally exclude their realizers, not the other way around. As noted by e.g. Loewer (2002) and Hitchcock (2012), what stance one takes on the exclusion problem hinges critically on what view of causation one endorses. Given the power and fruitfulness of the causal modeling approach, it would be particularly valuable to figure out its implications for those questions. Potential impact here is not limited to philosophy: cognitive psychologists, for example, are also interested in the question how people reason about high-level causation in multilevel settings, as this may shed light on general folk practices of causal representation (see e.g. Johnson & Keil, 2014; Blanchard et al., 2022).

Thirdly, multi-level contexts raise important issues for the interventionist account of causation (Woodward, 2003), which is intimately associated with the causal modeling framework. In particular, the question of how to think about interventions in multi-level settings, and whether interventions are possible at all in such contexts, is a topic of intense disagreement: witness the debate between Baumgartner (2013, 2018) and Woodward (2015, 2022) concerning whether interventions on multiply realizable properties are possible. This issue is tightly connected with the topic of mechanisms, as the question of modeling and figuring out the macro-effects of interventions into mechanistic constituents is an important and open question (see e.g. Casini et al., 2011). In addition, Craver (2007) has proposed an influential account of constitutive relevance in mechanisms based on the notion of intervention. But this account has been roundly criticized on the ground that the Woodwardian notion of intervention on which it relies is hard to apply in multi-level settings (Baumgartner & Gebharter, 2016; Romero, 2015), and whether it can be patched remains an open issue. Because the causal modeling framework provides the foundations for the interventionist account, a clear understanding of how the framework applies in multi-level contexts may help us make progress on those debates.[3]

As these remarks make clear, multi-level settings typically involve one of two types of non-causal dependence relationships: asymmetric supervenience of high-level properties on their lower-level realizers, and dependence between wholes and their parts. Accordingly, our question in what follows is how to represent those two types of dependence in causal models. (Thus when we speak of non-causal dependence in what follows it is always these two types of dependence we have in mind.)

2. Gebharter's Extension

To motivate our account, we will start by examining the simplest and most fleshed-out extension proposal one finds in the literature, which is due to Gebharter (2017a, 2017b, 2022). (See also Stern and Eva (2023), who endorse Gebharter's framework, though they disagree with him on its implications.) According to Gebharter, non-causal dependencies such as asymmetric supervenience and part-whole relationships can and should be treated as formally analogous to causation. Thus, CBNs are apt to model multilevel contexts after all, provided one makes one small modification to their standard interpretation, and allows the direct edges in **E** to represent not only direct causation but also direct non-causal dependence. (That is, $X \rightarrow Y$ may indicate

[3] Gebharter (2017b) suggests a further reason for extending the causal modeling framework to multi-level settings: this may allow us to apply the powerful causal discovery methods developed within the framework to the identification of non-causal relationships on the basis of statistical data. However, we are skeptical of this third motivation, for reasons identified by Casini & Baumgartner (2023) and to which we will return later: see fn. 25 for further discussion.

either that *X* is a direct cause of *Y*, or that *Y* directly depends on *X* in a non-causal manner.) Everything else in the formalism, including the CMC and CMIN axioms, stays the same.

To illustrate, consider the setup of Kim's causal exclusion argument. Suppose that *M* represents whether a certain mental property is instantiated, and *P* the possible physical realizers of that property. That property is a putative cause of another mental property itself represented by variable *M**, and whose possible physical realizers are represented by *P**.[4] (Kim argued that in this situation *M* cannot be a cause of *M**: because *P* fixes the value of *P** and hence of *M**, there is no room left for *M* to do any causal work in producing *M*.) On Gebharter's proposal, the correct model for this situation includes an arrow from *P* to *M* and one from *P** to *M**, representing asymmetric supervenience, and an arrow from *P* to *P**, representing physical causation. (That *P* directly causes *P** is common ground in the debate on causal exclusion and arguably follows from the causal closure of the physical.) See Figure 2. Note that Figure 2 does not posit a causal relationship between *M* and *M** and thus may or may not turn out to be incomplete, depending on whether Kim was right about exclusion. As it turns out, Gebharter's framework implies that Figure 2 is an adequate representation of the situation, and hence that *M* is causally excluded. We will return to this point later on. For now, we want to examine the motivations for endorsing Gebharter's proposal. Among its proponents, the key motivation seems to be that asymmetric supervenience and part-whole relationships are structurally similar to causal relations in crucial respects (see Gebharter 2017a: 358-364; 2017b: 2652-3; Stern and Eva, 2023). In particular, a key consideration is that asymmetric supervenience and part-whole relations generate the same screening-off relationships as causal relationships (see e.g. Gebharter, 2017b: 2653) observes. For instance, the supervenience of the mental variable *M* on its realizer variable *P* implies that *P* screens *M* off from other variables including *P** and *M**, because fixing the value of *P* fixes the value of *M*, and hence decorrelates it from other variables. In that respect, *P* behaves like a causal parent of *M*. Likewise, fixing the behavior of the parts fixes the behavior of the whole, thereby decorrelating it from other variables. In that respect, parts behave like causal parents of wholes. Moreover, asymmetric supervenience and part-whole relationships explain in much the same way that causal relationships do. Thus as Stern and Eva (2023) observe, we can explain a psychological state in terms of its causal history (and not vice

[4] As is standard, when we say that a property causes another property we mean that an instance of the former causes (on a given occasion) an instance of the latter.

versa); but we can also explain it in terms of the subvening neurological state (and not vice versa).[5] Though Stern and Eva do not discuss part-whole relationships, the same point applies there, since it is part of scientific practice to explain the behavior of wholes in terms of the behavior of their parts, but not vice versa.

[Figure 2 here]

3. Motivations for Gebharter's Framework: A Critical Examination

Gebharter's extension proposal has advantages. In particular, it implies that causal modelers don't need to refashion their tools to handle multilevel settings, and can rely on the very well-understood formalism of CBNs. But we think that ultimately his framework should be rejected. A key reason is that the motivations for treating non-causal dependence relationships on a par with causal arrows are not compelling.

First, the argument that asymmetric supervenience and part-whole relationships give rise to the same screening-off phenomena as causal relationships is problematic – a point already made by Kinney (2023) in a critical discussion of Gebharter. As Kinney notes, supervenience and causal relationships actually differ in statistically relevant respects. Causes and effects behave differently under interventions: while intervening on a cause preserves the correlation with an effect, intervening on an effect destroys its probabilistic dependence on the cause. Not so in the case of supervenience relationships: in Figure 2, for instance, wiggling *P* leads to a change of *M* but wiggling *M* also leads to a change in *P*. Though Kinney doesn't discuss the case of part-whole relationships, similar considerations apply there. An idea that has loomed large in the literature on mechanisms since the seminal work of Craver (2007) is that parts and wholes are mutually manipulable: as Craver puts it, part-whole relationships are characterized by the fact "one can wiggle the behavior of the whole by wiggling the behavior of the component [i.e. part] and one can wiggle the behavior of the component by wiggling the behavior of the whole" (2007: 153). For example, changing the momentum of one or more of the particles that compose

[5] See (Schaffer, 2016).

a billiard ball would change the momentum of the ball; but likewise changing the momentum of the ball as a whole would also necessarily change the momentum of at least some of its parts. This is one aspect in which the statistical signature of supervenience and part-whole relationships differs from that of causal relationships, and a particularly important one in the context of causal modeling, since one of the main functions of causal models is to encode information about manipulability relationships (Woodward, 2003).

Some clarifications are in order here. First, Craver not only argues that parts and wholes are mutually manipulable, but also offers an account of constitutive relevance in terms of mutual manipulability. On this account, a part is constitutively relevant to the behavior of a whole just in case manipulating the behavior of the part changes the behavior of the whole and vice versa. Craver furthermore proposes to understand the relevant manipulations as interventions in the sense of (Woodward, 2003). As noted above, this proposal is problematic, as several reasons suggest that in the context of part-whole relationships Woodwardian interventions are impossible.[6] (Similar troubles loom if one tries to make sense of the mutual manipulability of high-level and realizer properties in terms of Woodwardian interventions, as it is highly controversial whether this type of intervention is possible in contexts involving supervenience relationships between distinct properties: see Baumgartner (2009, 2013, 2018) and Woodward (2015) for discussion.) In claiming that parts and wholes are mutually manipulable, we do not mean to endorse Craver's mutual manipulability account of constitutive relevance, nor do we mean to endorse the view that the claims about mutual manipulability made above can be understood in terms of Woodwardian interventions.[7] Rather, those claims are meant with the ordinary, non-technical sense of manipulation in mind. On that sense, it is uncontroversial that we can manipulate realizers by manipulating high-level properties and *vice versa* (I can change your mental state by manipulating the state of your brain, but I can equally well change the state of your brain by manipulating your mental state – e.g. by saying something that changes your mind). We also do not mean to claim that the phenomenon of mutual manipulability cannot possibly be accommodated within Gebharter's framework. If one already accepts the assumption

[6] See especially Romero (2015) and Baumgartner & Gebharter (2016).

[7] Though we think that ultimately our account does help make sense of interventions in contexts involving non-causal dependence relationships (see §9). It may therefore help articulate a more plausible version of the mutual manipulability account of constitutive relevance, though for lack of space we will not explore this point in the manuscript.

that supervenience and part-whole relationships are on a par with causal relations, one may explain why manipulations of a supervening property (or a whole) change the behavior of the subvening property (or the whole's parts) in several ways, e.g. by claiming that every such manipulation also directly affects the subvening property/parts (see Gebharter, 2022). Our point, rather, is that the phenomenon of mutual manipulability weakens the motivations for accepting Gebharter's framework in the first place, by pointing to an important *prima facie* difference between non-causal and causal dependencies. The phenomenon thus suggests that other proposals may be worth exploring.

Consider next the argument that supervenience and part-whole relationships play the same asymmetric explanatory role as causal relations: science explains wholes in terms of their parts, and high-level properties in terms of their realizers, but not *vice versa*. This supports treating those relationships like causal arrows only if this fact about explanatory practices reflects a relationship of asymmetric metaphysical dependence between parts and wholes, or high-level properties and their realizers. But this may not be the case. As Hüttemann (2021, ch. 6) notes in a discussion of part-whole relationships, an asymmetric explanatory relation does not generally warrant postulating an underlying asymmetric dependence relation. For example, to explain why a change in voltage led to a change in current in an electronic network, we will presumably appeal to Ohm's law, which describes the dependence of current on voltage. This explanation is asymmetric in the sense that there is a direction of explanation from change of voltage to change of current. But Ohm's law is symmetric, and can just as well be used to explain changes in voltage based on changes in current.[8] The explanatory asymmetry reflects our interest in explaining the change in current, not the change in voltage. As Hüttemann further notes, similar remarks apply to part-whole relationships. Part of scientific practice is that we explain the mass of (e.g.) a billiard ball in terms of the masses of its parts, and not the mass of a part in terms of the mass of the compound and that of the other parts' masses. Yet the law of composition that underlies this explanation is perfectly symmetric. Let A be a variable representing the mass of a billiard ball, and $a_1,\dots, a_n$ variables that represent the masses of its n parts, with a_i representing the mass of the ith part. (Note that in using lower-case letters to denote variables representing parts of wholes, we depart from the usual practice of reserving lower-case

[8] See Kistler (2013) for a discussion of such association laws.

letters for values of variables.) Then we have $A=a_1+a_2+\ldots+a_n$, and this law of composition is symmetric in the sense that the value of any of these variables is determined by the values for all the others. The fact that we explain wholes in terms of parts but not vice versa may be traced back to facts about our interests. Dependence relations between parts and wholes allow us to manipulate how systems behave, and this presumably explains our interests in such dependence relations. The asymmetry of explanation may therefore mirror the fact that we are often interested in manipulating the behavior of a whole by manipulating its parts, while we are rarely if ever interested in changing the behavior of a part by manipulating the behavior of the whole and holding other parts fixed. (Hüttemann, 2021: 171). Likewise, an asymmetric explanatory relationship between the physical realizer variable *P* and the mental variable *M* need not reflect a relationship of asymmetric metaphysical dependence of *M* on *P*. True, one may wonder whether the *asymmetric* supervenience of *M* on *P* is really compatible with there being a *symmetric* dependence between them. But here it is important to keep apart two different relations of metaphysical dependence. First there is the relation between mental properties and their possible physical realizers. This is the supervenience relation. Supervenience is asymmetric if mental properties can be multiply realized, e.g., if *M* need not be carbon-based as it is in our world but could also be silicon-based (in some other world). Second, there is the relation between M and its *actual*, say carbon-based, realizer P. Even if M asymmetrically supervenes on its various possible realizers the relation between M and its actual realizer might be symmetric or mutual in the following sense: Manipulating certain features of P while holding fixed others yields changes in M and *vice versa*. In the context of the causal exclusion argument, when we consider what happens to M when we manipulate P, we typically (i.e. barring certain science-fictional scenarios) are holding the realization fixed. We are not envisaging switching from a carbon-based realization to a silicon-based realization. Rather we are considering changes within the carbon-based realization and how those affect M. The upshot is that even when we are dealing with an asymmetrical supervenience relation, the relation between M and its realizer P may be symmetrical or mutual – as in the case of part-whole relationships. This observation is not undermined by the fact that we typically explain M in terms of the characteristics of its actual realizer, because – as before – the asymmetry of explanation may simply mirror the fact that we are often interested in manipulating mental states by manipulating its underlying neurological state, while we are rarely if ever interested in manipulating certain features of a subject's neurological state by manipulating her mental state while holding other features of that neurological state fixed.

4. Mutual Dependence and Hybrid Models

If the reasons for treating asymmetric supervenience and part-whole relationships as formally analogous to causal arrows are not compelling, how else might causal modelers represent multi-level structures? Note that if one wants to hold on to the traditional formalism of causal Bayes nets to model these structures, the only other option besides Gebharter's is to treat asymmetric supervenience and part-whole relationships as formally analogous to causation, but with arrows going *downward*, from supervening to subvening properties, and from wholes to their parts. But the proposal has little intrinsic plausibility, and faces crippling objections anyway. For one thing, it has the wrong statistical implications. To see this, suppose again that *A* represents the mass of a billiard ball, which for simplicity is assumed to be made of only two parts, whose masses are represented by a_1 and a_2 respectively. On the current proposal, the situation would be represented with the graph $a_1 \leftarrow A \rightarrow a_2$. CMC then implies that a_1 and a_2 should be statistically independent conditional on *A*. But this is the wrong result: actually, a_1 and a_2 *are* correlated given *A* (holding *A*'s value fixed, a_1's value fixes a_2's value and *vice versa*). The proposal appears to be a non-starter. This suggests that the formalism of causal Bayes nets is simply not apt to represent structures involving asymmetric supervenience or part-whole relationships – a conclusion also endorsed by Kinney (2023). This is not to say that multi-level settings cannot be represented by causal models at all: CBNs are one (very popular) type of causal model, but not the only one, as we will see. Given that there are compelling reasons to seek an application of the causal modeling framework to multi-level contexts, it is worth exploring whether some other type of causal models might be the right tool for the job. This is the task to which we now turn.

Our discussion so far has suggested that asymmetric supervenience and part-whole relationships are best treated as *mutual* rather than asymmetric, especially if we focus on what this dependence implies for manipulability. More generally, the picture of the world suggested by our discussion is one on which there are (at least) two basic types of dependence relations in the universe: *causal dependence* relations, which are asymmetric and hold between metaphysically unrelated entities, and *mutual dependence* relations, which are symmetric and hold primarily between variables representing states of affairs at different levels of reality (e.g. wholes and their parts, or higher-level properties and their realizers). This in turn suggests that to represent structures involving both types of relationships, we need to use models that can represent both asymmetric and symmetric dependence relationships.

As it happens, in the causal modeling tradition various tools have been proposed to represent structures of that sort, including *chain graph models* (Lauritzen & Wermuth, 1989), which can contain both directed and undirected (—) edges, *acyclic directed mixed graphs* or ADMGs (Richardson, 2003), which contain both directed and bidirected (↔) edges, and *directed cyclic graph models* (Spirtes, 1995), which include only directed edges but allow edges running in both directions between variables (e.g. the set of edges may contain both $X \rightarrow Y$ and $Y \rightarrow X$). These devices are each associated with a distinct Markov condition, and have been used to represent a variety of structures that include variables standing in relationships of mutual dependence. For instance, chain graphs are used in econometrics to represent causal structures that include causal feedback between variables[9], while ADMGs are generally used to represent structures that contain dependencies due to latent common causes. However, as we will see shortly, none of these devices can be used to properly represent structures involving part-whole and asymmetric supervenience relationships (see fn. 12 below).

Accordingly, in what follows we introduce a new type of graphs designed for this purpose, which we call *hybrid models*. Just like CBNs, they consist of a graph and a probability distribution, as well as Markov and minimality axioms connecting the two. But the graph in question is not a DAG but a *hybrid graph*. Hybrid graphs differ from DAGs in several ways, some of which will be discussed only later. One key difference is that in hybrid graphs, variables can be connected either by directed edges ($X \rightarrow Y$, indicating direct causal dependence of Y on X) or by dashed bidirected edges. (We used dashed edges to distinguish them from the plain bidirected edges of ADMGs.) $X \dashleftarrow\!\dashrightarrow Y$ indicates direct *mutual* dependence between X and Y. At most one edge can exist between variables, so that variables related by directed edges cannot be related by bidirected edges nor vice versa. (As we will see shortly, hybrid graphs also differ from DAGs in that they incorporate information about the *levels* at which variables are located.) As an

[9] See Lauritzen & Richardson (2002). In the philosophy of science, Steel (2010) has used chain graphs as part of an account of extrapolation.

illustration, consider the mental causation example discussed above (cf. Figure 2). On our view, in addition to the arrow between P and P^*, the correct representation of the case should include bidirected edges between M and P and between M^* and P^*, as in Figure 3. (Figure 3 is just like Figure 2 except that arrows from realizers to mental properties have been replaced by bidirected edges.) Whether this graph is complete and what it implies about the causal relationships in this situation are questions to which we will return later. First, we need to lay down axioms for hybrid models, and in particular specify what Markov condition they obey.

[Figure 3 here]

5. Markov and Minimality Conditions for Hybrid Models

To formulate our Markov condition for hybrid models, we will focus on the notion of a collider, in terms of which the Markov condition for causal Bayes nets CMC is formulated. Remember that informally, a collider is a variable on a path that does not transmit influence along the path, so that if a path between two variables contains a collider the path will not induce a correlation between them. In DAGs, this arises if – and only if – a variable on the path has two directed edges incoming into it (i.e. if we have $X \rightarrow Y \leftarrow Z$, Y is a collider). As we will see, one can formulate a plausible Markov condition by examining further ways in which a variable may fail to transmit influence along a path in hybrid graphs, and extending the notion of a collider accordingly.

There are two situations to consider: the one in which the two incoming edges are bidirected ($X \dashleftarrow\dashrightarrow Y \dashleftarrow\dashrightarrow Z$), and the one in which one edge is bidirected and the other directed (e.g. $X \rightarrow Y \dashleftarrow\dashrightarrow Z$ or $X \dashleftarrow\dashrightarrow Y \rightarrow Z$). Let’s consider the second type of case first. Here we think the right thing to say is that influence is transmitted from X to Z via Y, and hence that Y is not a collider. Indeed, we think it is plausible what is widely assumed in the debate about exclusion: to say that when X and Z are related in this way, there is a *causal* influence of X on Z. That is, if X and Z are linked by directed chains including both non-causal and causal dependence relationships, then there will ordinarily be a further relationship of causal dependence and hence also a correlation between X and Z. Conversely, if X is causally relevant to Z, it may be that some of the links in the path by which X influences Z are links of non-causal dependence.[10] For example, in the setup of Kim's argument (see Figure 2), it is commonly accepted that the non-causal dependence of M^* on P^* and the causal dependence of P^* on P compose to make it so that P causes M^*[11] . The key upshot for our purposes is that causal and non-causal dependence relationships compose to form further causal relationships. Thus in a case where we have $X \rightarrow Y \leftarrow\cdots\rightarrow Z$ or $X \leftarrow\cdots\rightarrow Y \rightarrow Z$, Y behaves formally like it would in an ordinary causal chain ($X \rightarrow Y \rightarrow Z$) and hence should be regarded as a non-collider.

Consider now the first type of case, where the two incoming edges are bidirected ($X \leftarrow\cdots\rightarrow Y \leftarrow\cdots\rightarrow Z$). Here we think the right thing to say that whether Y is a collider depends on the *levels* of reality to which the relevant variables belong, so that information about levels must be integrated into the formalism of hybrid models. To understand why, suppose again that A represents the mass of a billiard ball, which for simplicity is assumed to be made of only two parts, whose masses are represented by a_1 and a_2 respectively. And contrast this case with a second one, in which B represent a certain biological state, C a certain chemical state, and P a certain physical state, such that B asymmetrically supervenes on C, and C asymmetrically supervenes on B. On our proposal, these two scenarios would be represented by the graphs in Figures 4 and 5 respectively. If one focuses only on variables and edges, these two graphs appear isomorphous to one another. But crucially, the two situations have different statistical profiles. In the second situation, because P and B are each mutually dependent (and hence correlated) with C, they are also correlated with one another. Thus in this case mutual dependence appears to travel across the intermediary variable C, inducing a correlation between the endpoint variables P and B. (Accordingly, that correlation disappears once we hold C fixed, i.e. C screens off P and B from each other.) But as we noted earlier, in Figure 5 a_1 and a_2 are *not* correlated with one another, despite the fact that both are mutually dependent with A. Thus in this case mutual dependence does not travel across the intermediary variable. (Accordingly A does not screen off a_1 from a_2, indeed holding A fixed induces a dependence between these two variables: holding

[10] This doesn't mean that non-causal dependence itself is a form of causation. Mutual dependence remains a *sui generis* relationship of non-causal dependence, irreducible to causation.
[11] As we will see there is another important reason for proponents of Gebharter's framework to endorse that assumption: see fn. 16 below.

the value of *A* fixed, setting a_1's value fixes a_2's value and vice versa.) Since our proposal represents these two situations with isomorphic graphs, it is hard to see how it could possibly account for this statistical difference: it would seem that any Markov condition one might impose on those graphs would yield the same independencies in both cases.[12] But once we take into account information about the *levels* at which the relevant variables are located, a clear dissymmetry appears between the two graphs. Note that in Figure 4, a_1 and a_2 are situated at the same level of reality – they both belong to the microphysical level. In Figure 5, on the other hand, *P* and *B* are situated at different levels of reality – the physical one and the biological one, respectively. Once we take this into account, the statistical asymmetry between these two situations can be explained if we posit the hypothesis that *mutual dependence does not propagate in zig-zags*. That is, mutual dependence can travel upwards (from lower to higher levels of reality) or downwards (from higher to lower levels), but cannot travel up and then down again, or down and then up again. This means in particular, that if *X* is mutually dependent with a higher-level variable *Y*, and *Y* is mutually dependent with a lower-level variable *Z*, *X* and *Z* will not thereby be mutually dependent, and hence also not thereby correlated. In that case, *Y* is a collider on the path between *X* and *Z*. By contrast, if *X* is mutually dependent with a higher-level variable *Y*, and *Y* is mutually dependent with an even higher-level variable *Z*, then *X* and *Z* will themselves be mutually dependent and hence correlated. In this case *Y* is not a collider. (Since here *Y* mediates the dependence between *X* and *Z*, we should also expect the correlation between *X* and *Z* to be screened off by *Y*.) This hypothesis thus explains why in Figure 5 the mutual dependence between *P* and *C* and between *C* and *B* induces a correlation between *P* and *C*,

[12] This problem is the reason why none of the tools for representing symmetric dependence that already exist in the causal modeling literature fit our purposes. Whichever of those devices we use, our two situations would be represented by isomorphic graphs: $a_1 – A – a_2$ and $P – C – B$ if we use chain graphs, $a_1 \leftrightarrow A \leftrightarrow a_2$ and $P\leftrightarrow C\leftrightarrow B$ if we use ADMGs, and $a_1 \leftrightarrows A \leftrightarrows a_2$ and $P\leftrightarrows C\leftrightarrows B$ if we go directed cyclic graphs. And because in each case the two resulting graphs are isomorphic to one another, the Markov condition for those graphs is bound to treat them on a par, and hence to get at least one of them wrong. For instance, the Markov condition for acyclic directed mixed graphs (Richardson, 2003) entails that if we have a path $X\leftrightarrow Y\leftrightarrow Z$ (and *X* and *Z* are otherwise unrelated) *X* and *Z* are statistically independent. Hence that condition correctly entails that a_1 and a_2 are independent in the first situation, but also incorrectly entails that *P* and *B* are independent in the second. On the other hand, the Markov conditions for chain and directed cyclic graphs entail that in chains of mutual dependence intermediary variables screen off the endpoint variables from each other (Lauritzen and Wermuth, 1989; Spirtes, 1995). Thus those Markov conditions correctly entail that *C* screens off *P* and *B* from each other in the second situation, but incorrectly entail that *A* screens off a_1 from a_2 in the first.

whereas in Figure 4 a_1 and a_2 are statistically uncorrelated despite both being mutually dependent on *A*.

[Figure 4 here]

[Figure 5 here]

To sum up, on our view there are two situations in which a variable *Y* is a collider in hybrid graphs: if it has two directed edges incoming into it (as in causal Bayes nets), and if we have a chain of mutual dependence $X \leftarrow\!\cdots\!\rightarrow Y \leftarrow\!\cdots\!\rightarrow Z$ that does a zigzag (i.e. *X* and *Z* are both located either at higher levels or at lower levels than *Y*.) We therefore extend the notion of collider introduced in section 1 to include this latter scenario. To do so, information about variable levels must first be incorporated into our formalism. Accordingly, we define hybrid graphs as containing not only a set of variables **V** and a set of edges **E**, but also an *ordered partition* $\{L_i\}_{1 \leq i \leq n}$ over **V**. The cells of the partition contain all variables located at the same level, and the order of the partition represents the hierarchy of levels, so that for any two variables *X* and *Y* in **V** such that *X* is in L_i and *Y* is in L_j, *X* is located at a (strictly) lower level than *Y* just in case $i < j$. If *X* belongs to a lower level than *Y*, we say that *Y* is a *superior* of *X*. We will not say more about the notion of level in this paper, as in all the cases discussed in this paper it is clear and uncontroversial which levels variables belong to.[13] Visually, levels can be identified by the spatial position of variables in graphs: variables situated on the same horizontal dimension belong to the same level, while variables situated higher on the vertical dimension belong to higher levels. The spatial arrangement of variables in Figure 4, for instance, indicates that the relevant ordered partition is $\{\{a_1, a_2\}, \{A\}\}$, reflecting the fact that a_1 and a_2 are located at a

[13] More generally, however, there is a debate about how to clarify the notion of level (see Eronen & Brooks (2018) for a recent overview). Our hypothesis suggests a criterion for distinguishing different levels and might thus contribute to this debate.

lower (microphysical) level than macrophysical variable A. In Figure 5, on the other hand, the ordered partition is $\{\{P\}, \{C\}, \{B\}\}$.[14]

With this added structure in place, we define the notion of h-collider in the following way:

> **h-collider**: A non-endpoint variable Y on a path in a hybrid graph is a *h-collider* iff
>
> (a) both edges preceding and succeeding Y are directed edges pointing at Y (i.e., $X \rightarrow Y \leftarrow Z$ for some X and Z on the path) or
>
> (b) both edges preceding and succeeding Y are bidirected, so that $X \leftarrow\cdots\rightarrow Y \leftarrow\cdots\rightarrow Z$ for some X and Z on the path, and either X and Z are both superiors of Y, or Y is a superior of both X and Z.

With this notion in place, we can now generalize the notion of d-separation to hybrid graphs. Because the notion of descendant appears in the definition of d-separation, this step requires a new definition of descendant tailored to hybrid graphs. Remember that the intuitive idea behind that notion is that Y is a descendant of X when there is a path from X to Y by which X influences Y. In DAGs, this happens when there is a directed path from X to Y. In hybrid graphs, there are two other situations to consider. First, the path from X to Y might be composed of arrows and bidirected edges (e.g. if we have a path $X \rightarrow W \leftarrow\cdots\rightarrow Y$, or $X \leftarrow\cdots\rightarrow W \rightarrow Y$). Second, X might influence Y entirely by way of relations of mutual dependence, as e.g. P and B influence one another via the path $P \leftarrow\cdots\rightarrow C \leftarrow\cdots\rightarrow B$ in Figure 5. (In this case, the path must not contain zig-zags, since mutual dependence does not propagate across them.) The following definitions capture those two additional situations. Say first that a *weakly directed* path is a sequence of variables such that for every consecutive pair of variables $\{X_i, X_{i+1}\}$ in it, either $X_i \leftarrow\cdots\rightarrow X_{i+1}$ or $X_i \rightarrow X_{i+1}$, and there is no h-collider on the path. Then Y is a *h-descendant* of X iff there is a weakly

[14] Note that when drawing DAGs there is already a convention of arranging variables vertically according to the levels to which they belong. However, that convention is simply a visual help and not part of the content of DAGs. In hybrid graphs, by contrast, spatial position of variables *is* part of the official and explicit content of the graph.

directed path from *X* to *Y*. We then offer the following generalization of d-separation for hybrid graphs:

> **h-separation**: Two variables *X* and *Y* in a hybrid graph are *h-separated* by a (possibly empty) set of variables **Z** just in case, for every path between *X* and *Y*, (a) there exists a non-h-collider on the path that is in **Z** or (b) the path contains a h-collider, and neither it nor any of its h-descendants is in **Z**.

This yields the following Markov condition for hybrid models:

> **HMC**: For every *X*, *Y*, **Z** in **G**, if **Z** h-separates *X* and *Y*, then P(Y/X&**Z**)=P(Y/**Z**).

Like CMC, HMC by itself puts few constraints on probabilistic structure, as it never forbids the inclusion of an edge in a graph. Inclusion of superfluous edges is prohibited by the following minimality condition, which is a straightforward extension of CMIN to hybrid models:

> **HMIN**: No proper subgraph of **G** satisfies HMC with respect to *P*.

As we will see in the next section, these axioms are empirically plausible in the sense that they yield the intuitively right conclusions in a range of cases. Another argument in their favor concerns their behavior in cases where the set of variables **V** contains no variables that are mutually dependent. Then the correct hybrid graph **H** and the correct DAG **G** over **V** are equivalent. (The only difference between them is that **H** contains information about the levels of the variables that is not contained in **G**, but in this particular case levels do not affect the probabilistic behavior of those variables.) So in such a case the axioms for hybrid graphs should yield the same results as CMC and CMIN. And indeed that is what we find: as the reader may verify, in this special case h-separation reduces to d-separation, so that HMC and HMIN have the same implications as CMC and CMIN.

This concludes the presentation of our framework. In the remainder of the paper, we will contrast it further with Gebharter's proposal, by comparing their implications for two issues concerning causation in multilevel settings: the exclusion problem, and the nature and possibility of interventions on high-level problems. (We leave exploration of the issue of mechanisms for another time.) This examination will provide further evidence for the usefulness of our framework and its superiority over Gebharter's proposal.

6. High-Level Causation and Kim's Exclusion Problem

Let's consider Kim's exclusion problem first. Strikingly, Gebharter (2017a) argues that his extension supports Kim's contention that multiply realizable properties are causally excluded by their realizers. To see why, consider again the scenario depicted in Figure 2, where the dependence of high-level properties on their realizers is represented with directed edges, in accordance with Gebharter's proposal. Positing a causal influence of M on M^* would require adding to Figure 2 either an arrow from M to M^*, or an arrow from M to P^* (thus creating a directed path from M to M^*). But as can easily be verified, the graph in Figure 2 satisfies CMC. (In particular, it correctly entails that M is independent of M^* given P or P^*.) And so adding any arrow to that graph would lead to a violation of CMIN. The argument might also be put as follows. CMIN implies that a direct cause must make a probabilistic difference to its effect given that effect's other parents. So on Gebharter's framework, M directly causes M^* only if M^* probabilistically depends on M given P^*. But this is not the case, as the value of P^* entirely fixes the value of M^*, so that conditioning on P^* necessarily decorrelates M^* from other variables, including M. And for M to cause M^* by directly causing P^*, P^* and M would have to be correlated given P^*'s other parent P. But that cannot be so, since fixing P fixes M, and hence decorrelates M from any other variable, including P^*.[15] Thus, according to Gebharter, the causal

[15] The fact that Gebharter's framework yields exclusion provides one further reason for him to endorse the assumption that causal and non-causal dependence relationships compose to yield further causal relationships. In Figure 2, this assumption entails that P causes M^* by causing P. If we were not allowed to compose causation with

modeling framework justifies the intuition at the heart of Kim’s exclusion argument: namely, that multiply realizable properties are made causally redundant by their realizers.

Our proposal, by contrast, does not lead to exclusion. Consider the hybrid graph in Figure 3, in which the dependence of high-level properties on their realizers is represented by bidirected edges. (Here the ordered partition over the variables is {{*P*, *P**}, {*M*, *M**}}, with the first level containing physical variables, and the second mental variables.) As the reader may verify, this graph satisfies the modelling axioms laid out in the previous section. When applied to this graph HCM implies in particular that (a) *M* is statistically independent of *P** and *M** given *P* (as *P* is a non-h-collider on the sole path from *M* to *P** and *M**) and (b) *M** is independent of *M* and *P* given *P** (as *P** is also a non-h-collider on the sole path from *M* and *P* to *M**). Both implications are correct. The graph also satisfies HMIN: getting rid of any of its edges would entail independencies that do not actually hold in the situation. And finally, as the reader may also verify, taking for granted that *M* and *P* are directly mutually dependent, that *M** and *P** are as well, and that *P* is a direct cause of *P** (which is a given in the situation), that graph is the only one that satisfies HMC and HMIN, and on our view it is therefore the correct way to represent the network of causal and non-causal dependencies in Kim’s scenario. Moreover, on our interpretation that graph does posit a causal influence of *M* on *M*: M* influences *P* (via mutual dependence), which causes *P**, which influences *M** (via mutual dependence again), and these links compose to create a causal influence of *M* on *M**. Note that our proposal is empirically (statistically) equivalent to Gebharter’s, as the graph in Figure 3 entails exactly the independencies that are entailed by Gebharter’s extension of the CBN framework. The difference between the two frameworks lie in the way they interpret those statistical facts. Our account, because it considers the relation between high-level properties and their realizers as one of mutual dependence, allows us to integrate this relationship in a chain of causal dependence, whereas Gebharter’s framework, which treats the relation between high-level properties and their realizers as formally analogous to causation, is unable to do so.

In addition to the motivations discussed above, this result provides further reason to favor our modelling proposal over Gebharter’s, at least for non-reductive physicalists. The advantage

influence or determination due to supervenience to yield further causation, Gebharter would have to say (implausibly) that *M** in this scenario is caused neither by *M* nor by *P*, and hence not caused at all.

of our framework is not only that it offers non-reductive physicalists a consistent way to model multi-level settings that avoids exclusion; it also offers a positive view of the way that high-level causation works in multi-level settings, and how it relates to lower-level causation. On this view, high-level properties produce their effects by way of their mutual dependence with their realizers. This view is therefore clearly in the 'compatibilist' camp which holds that high-level and low-level causes are not in competition, while also offering a way to address the worry that compatibilism implies a problematic kind of overdetermination. On our view, the fact that high-level properties and their realizers cause the same effect is no more problematic than the fact that an event can be the result of both proximal and distal causes.[16]

One wrinkle here is that it is in fact debated whether Gebharter's framework really leads to exclusion. Stern and Eva (2023), who endorse the framework, have recently argued that properly applied it does not actually make high-level properties causally impotent. Their case proceeds as follows. Consider Hesslow's (1976) famous birth control example, in which birth control pills (*BP*) reduce the risk of thrombosis (*T*) by preventing pregnancy (*PR*), but also cause thrombosis via another route (Figure 6). The two routes happen to cancel each other, so that *BP* and *T* are uncorrelated. As Stern and Eva note, the example shows that the absence of a directed path from *X* to *Y* in a DAG is no indication that *X* does not cause *Y*, even if the relevant variable set is causally sufficient. For in Hesslow's example, if our variable set includes only *BP* and *T* (which by stipulation do not have a common cause), the only graph that obeys the CBN axioms is the empty graph. On that basis, Stern and Eva propose the following principle:

> Weak Causation Principle (WCP): In order for a variable *X* to count as causally relevant to a variable *Y*, there must be *some* causally sufficient variable set **V** containing *X* and *Y* such that there exists a directed path from *X* to *Y* in some admissible graph over **V**. (Stern and Eva 2023; our emphasis)

WCP gets Hesslow's example right. For suppose we add *PR* to our variable set. The graph over that set in Figure 3 does contain a directed path from *BP* to *T*, and is admissible (i.e. satisfies the CBN axioms). So WCP entails correctly that *BP* is causally relevant to *T*.

[Figure 6 here]

[16] Which again is not to say that the relationship between a high-level property and its realizer is causal: Mutual dependence remains a *sui generis* relationship distinct from causation.

But now, given WCP, exclusion doesn't follow in Kim-style situations anymore, even accepting Gebharter's extension of the causal modeling framework. Although on that extension no admissible CBN over the set {*M*, *M**, *P*, *P**} contains a directed path from *M* to *M**, this does not show that *M* does not cause *M**. For consider the causally sufficient variable set {*M*, *M**}, and the causal graph over it does contain an arrow from *M* to *M**. That graph satisfies CMC and CMIN. By WCP, the existence of this graph establishes that *M* does cause *M** after all.[17]

However, we take issue with the lessons drawn by Stern and Eva from Hesslow's example. Importantly, WCP is not the only principle that gets the case right: another one is *monotonicity*, according to which a graph's claim about whether *X* causes *Y* is correct if that claim continues to hold if we add more variables to the set. The intuition behind monotonicity is that if by adding more variables to the set we can overturn the verdict provided by the graph, then there is good reason to think that the graph was based on too sparse a set of variables, and hence that its causal verdicts should not be trusted.[18] This seems to be exactly what is going on in Hesslow's example. Consider once again the empty graph over the set {*BP*, *T*}. Intuitively, the reason why that graph's contention that *BP* does not cause *T* should not be trusted is that the variable set {*BP*, *T*} is *impoverished*: it fails to represent a crucial aspect of the causal structure under consideration, namely whether the person gets pregnant or not.[19] In other words, the intuitive reason why the claim that *BP* causes *T* is not undermined by the absence of a *BP*→*T* path in the model over {*BP*, *T*} is not simply that there are other models in which such a path exists, but that these models contain enough information to adequately represent the causal structure of the case in question. Thus, Hesslow's example may be taken to provide better support for monotonicity than for WCP. But if we accept monotonicity, one cannot both endorse Gebharter's extension and claim that M causes *M** anymore. Even though the DAG over {*M*, *M**} does contain a directed path from *M* to *M** does not mean that *M* causes *M** anymore, that

[17] Stern and Eva also consider the possibility that {*M*, *M**} fails an extended version of the causal sufficiency condition that requires appropriate models to include *all* parents (whether direct causes or subvening variables) of any two variables in the model. That condition would make the set {*M*, *M**} inappropriate since *P* is a parent of *M* and *M**. They argue that the best way to formulate this additional condition entails that {*M*, *M**} is an appropriate set after all. We will leave this aspect of their view aside since it is not relevant to our main criticism of their view.

[18] See Woodward (2008: 208) for a statement of this idea. Hoffman-Kolss (forthcoming) offers a number of arguments for a monotonicity requirement on causal relations in the context of causal modeling. See also Halpern (2016) for a discussion of monotonicity (or `stability') in the context of actual causation.

[19] See Blanchard and Schaffer (2017) for a discussion of the idea that an apt causal model should represent the "essential structure" of the situation, and an application to various problems concerning actual causation.

directed path is bound to disappear as soon as we add *P* to the variable set, on pain of violating CMIN. From the standpoint of monotonicity, what is going on is that the causal graph over {*M*, *M**} gives the *illusion* that there is causation flowing from *M* to *M**. That illusion stems from the fact that the relevant variable set is impoverished, and has failed to include the physical variable that realizes *M*. Once that variable is included, the conclusion is that the true causation runs from *P* to *M** only.[20] Thus, we maintain that in the choice between our and Gebharter's way of modelling multi-level settings, only the former can help non-reductive physicalists avoid exclusion worries.[21]

7. Part-Whole Relationships and the Causal Exclusion of Wholes

There are further considerations pertaining to exclusion that favor our framework over Gebharter's. Gebharter's framework, if correct, yields far more than exclusion of multiply realizable properties; it also entails that all composite entities (entities with proper parts) are causally excluded by their parts. (A point noted by Gebharter (2022) himself, though we think he

[20] Both WCP and monotonicity appear to get in trouble in cases involving failures of transitivity. Consider McDermott's famous case, in which a dog bites a terrorist's right hand, forcing them to use their left hand to detonate a bomb (McDermott, 1995). Let *DB* represent whether the dog bites, *R* whether the terrorist uses their right or left hand to detonate the bomb, and *B* whether the bomb explodes. The graph over those variables contains a path $DB \rightarrow R \rightarrow B$, and yet DB is intuitively causally irrelevant to *B*. WCP also wrongly counts *DB* as causally relevant to *B* since in the causally sufficient variable set {*DB*, *R*, *B*} there is a path from one to the other. Monotonicity yields the same result, as adding further variables to the set does not make that path disappear. However, Stern (2021) has proposed another principle very close in spirit to WCP, and which says that *X* is causally relevant to *Y* just in case *X* is a *direct cause* of *Y* in some causally sufficient variable set. This principle still allows Stern and Eva to reach their main conclusions, while also giving the right results in cases of intransitivity. A similar move is available to the proponent of monotonicity. That is, we may say that *X* is causally relevant to *Y* just in case it is a direct cause of *Y* in some variable set, and there remains a directed path from *X* to *Y* when one adds further variables to the set.

[21] Besides Eva and Stern, another author who argues that we can reject Gebharter's exclusion argument without going beyond the framework of CBNs is Kinney (2023). Kinney's account also relies on WCP, and his proposal captures mental causation in the same way as Stern and Eva, i.e. on the ground that the correct CBN over the variable set {*M*, *M**} includes an arrow from *M* to *M**. But his proposal differs from Eva and Stern in that it doesn't even allow causal Bayes nets to contain variables that stand in relationships of non-causal dependence: CBNs are only applicable to sets of variables that are logically and metaphysically independent from one another. Kinney's proposal is thus arguably immune to our criticism of Eva and Stern: on Kinney's proposal the set {*M*, *M**, *P*, *P**} is simply not an admissible variable set, so the fact that there is no path from *M* to *M** is no reason to reject mental causation, even if one accepts the independently plausible principle of Monotonicity. Yet Kinney does not consider whether some other type of model can be applied to multi-level settings. Our proposal shows that it is possible to model such situations in a coherent and plausible way, and that this has important philosophical benefits: it helps understand how micro and macro-causation can fit together (i.e. how physical realizers and high-level properties can both cause the same effect), and why this does not imply any problematic overdetermination.

misrepresents its significance.) To illustrate, suppose that billiard ball A collides with billiard ball B, which starts to move. Let variable *A* represent the mass of A, and *B* the momentum of B. Let variables $a_{1,}$ …, a_n represent the masses of the particles composing billiard ball A, and $b_{1,}$…, b_m the momenta of the particles composing billiard ball B.[22] Assuming the laws of classical mechanics (collision laws and conservation laws) varying any of the a_i may lead to changes in any of the b_j and this causal relationship does not seem mediated by any other variable in the graph. Therefore, we assume that each a_i is a direct cause of each b_j (see Figure 7). On Gebharter's framework, parts are treated as parents of wholes. (See the red arrows in Figure 7.) But if so *A* cannot be a cause of *B*. To see this, note that the graph in Figure 7 satisfies CMC. In particular, it correctly implies that *B* is probabilistically independent of $a_{1\ldots}$ a_n and *A* conditional on $b_{1\ldots}b_m$, and that $b_{1\ldots}b_m$ are probabilistically independent of *A* given $a_{1\ldots}a_n$. Because every graph that implies that *A* causes *B* – by adding either a direct arrow from *A* to *B*, or from *A* to (some or all of) $b_1\ldots b_m$ – is a supergraph of the graph in Figure 7, every such graph violates CMIN. Since there is nothing special about the example, the argument generalizes to threaten all other cases of causation by composites. In short, if Gebharter's take on the causal modeling framework is correct, wholes are made causally redundant by their parts, just like high-level properties are made causally redundant by their realizers.

[Figure 7 here]

The parallels between asymmetric supervenience and part-whole relationships are not lost on Gebharter, who in his (2022) offers an argument similar to the one just sketched. However, there he frames the argument as showing not that wholes are causally *excluded* by their parts, but that causal relations between wholes are *reduced* to lower-level causal relations between their parts. But we do not think this is right. Consider why it is appropriate to describe the causal modeling version of Kim's argument as an argument for eliminativism, i.e. as showing (in Gebharter's own words) that on non-reductive physicalism high-level properties "possess no

[22] For simplicity, we assume that that there is only one spatial dimension.

causal power" (2017a: 364). The answer, presumably, is that if correct the argument establishes that the causal powers that seemingly belong to *M* in fact belong to a *distinct* property, namely *P*. But if we take for granted that wholes are distinct from their parts, Gebharter's (2022) argument must be read as establishing a similar conclusion: namely, that causal powers that seemingly belong to a certain entity (the whole) in fact belong to *distinct* entities (its parts). The causal powers of wholes are eliminated, not reduced. True, this line of reasoning presupposes that wholes are not identical with their parts. Accordingly, Gebharter might be able to support his reductionist reading by endorsing composition as identity. However, that thesis is highly controversial, and justifiably so. (For one thing, it entails that wholes have their parts essentially, and hence cease to exist as soon as they lose a part.)

In short, unless some highly controversial metaphysical assumptions are correct, Gebharter's extension of the causal modeling framework entails that composites lack causal efficacy. We think this is a bad result, and one significantly worse than the fact that it makes multiply realizable properties causally excluded. After all one may endorse the latter conclusion but preserve high-level causation by rejecting non-reductive physicalism, as Kim himself does. By contrast, eliminativism about macrocausation strikes us as more problematic, being at odds with commonsense and scientific practice.[23] (Indeed Kim himself took pain to argue that his argument properly understood does not make all macrocausation impossible (Kim, 2003).) Moreover, it may well be that there is no fundamental mereological level (i.e. every part of an object itself has proper parts), in which case Gebharter's exclusion argument leads to causal powers `draining away', just like a number of authors have argued happens with Kim's original argument (see e.g. Block, 2003).

But things get worse. As it turns out, in our example, Gebharter's framework can also be used to argue that the *parts* are causally inefficient. To see why, imagine that we replace the arrows from a_1 to $b_1 \ldots b_m$ in Figure 7 by arrows running from *A* to $b_1 \ldots b_m$. (Everything else stays the same, in particular the arrows from all the other a_is to $b_1 \ldots b_m$.) As the reader may verify, that graph (call it **G')** satisfies CMC, so that the graph obtained by now adding arrows from a_1 to $b_1 \ldots b_m$ would violate CMIN. (Similar reasoning applies for the other a_is.)[24] This problem arises

[23] Though see Merricks (2001) for a defense.

[24] Note a significant disanalogy between the mental-physical-case and the part-whole-case. While in the first case the closure of the physical implies that it is assumed from the outset that P causes P* and the only question is

due to the fact that in our example, the masses of the parts and whole are related by the equation $A=a_1+a_2+ \ldots +a_n$. The value of the mass a_1 is determined by the other mass variables through the equation $a_1=A-a_2-a_3- ..a_n$. The same kind of determination relation obtains as for *A*. (Note that even though the value of *A* asymmetrically supervenes on the values for the a_is, the values for *A* and the $a_i s$ mutually determine each other according to the equation mentioned.) This means that when the values of *A*, $a_2 \ldots a_n$ are fixed, there is no room left for a_1 to vary and hence make a difference to other variables in the graph, so that adding arrows originating from a_1 becomes prohibited by CMIN. Hence, we now get the result that none of the a_is causes $b_1 \ldots b_m$ nor *B*.

This result, note, does not generalize to all cases of causation that involve parts and wholes, since the sort of mutual determination at work in the billiard ball example does not always obtain. For instance, the physical behavior of the *n* atoms composing a gun taken together fix whether the gun fires. But fixing the behavior of *n*-1 atoms plus the fact that the gun fires may not necessarily fix the behavior of the *n*th atom. But even though Gebharter's extension does not entail parts are inefficacious in all cases, the mere fact that it can be used to produce such a result in some cases is (we think) a further weighty consideration for thinking that Gebharter's extension of the causal modeling framework is incorrect.

It is therefore a further significant advantage of our framework that it yields no such counterintuitive result in settings involving part-whole relationships. For consider the hybrid graph of Figure 8, in which the dependence between parts and wholes is modelled as mutual dependence relationships. That graph satisfies HCM. In particular, it correctly entails that (a) *A* is statistically independent of $b_1 \ldots b_m$ and *B* given $a_1 \ldots a_n$ (since for every path from *A* to $b_1 \ldots b_m$ and/or *B*, one of the a_is is a non-h-collider on that path) and (b) *B* is statistically independent of $a_1 \ldots a_n$ and *A* given $b_1 \ldots b_m$ (as for every path linking *B* to $a_1 \ldots a_n$ and/or *A*, one of the b_is is a non-h-collider on that path). The reader may also verify that the graph satisfies HMIN, and that it is the only graph that satisfies the two axioms if we take it for granted that the wholes *A* and *B* are directly mutually dependent on their parts, and that every a_i is a direct cause of every b_j. And on our proposal, the graph in Figure 8 *does* posit a causal influence of *A* on *B*, which works as

whether *in addition* M is causally efficacious too, in the part-whole-case there is no such asymmetry: the causal closure of the physical does not single out one of the levels, because both, the level of the parts as well as the level of the compound, belong to the physical.

follows: *A* influences the behavior of its parts via mutual dependence; the parts of *A* causally influence (causation in the strict sense) the parts of *B*; and the parts of *B* influence *B* via mutual dependence again.[25]

[Figure 8 here]

8. Interventions

In closing we will briefly examine the implications of our framework for the vexed question of how to think about interventions in multilevel settings. In traditional causal modeling, it is standard to interpret arrows in CBNs in interventionist terms. In particular, Woodward (2003) offers the following well-known interventionist definition of direct cause:

[25] This is a good place to mention an objection to Gebharter's framework raised by Casini and Baumgartner (2023). They target Gebharter's claim that his framework allows one to run the well-known PC algorithm (Spirtes et al., 2000) to discover which parts are constitutively relevant to a whole's behavior. But as they point out, Gebharter's framework is in conflict with several of the PC's algorithm background assumptions. In particular, the PC algorithm relies crucially on the condition of causal faithfulness:

> **CFC**: **G** and *P* satisfy CFC iff every independency in *P* is entailed by the CMC.

CFC in effect forbids situations such as Hesslow's birth control example (Figure 3), in which *BP* and *T*'s independence does not follow from the CMC. In contexts where we are dealing only with causal relations CFC is a reasonable assumption since violations of faithfulness require causal routes canceling each other out in highly improbable ways. But faithfulness cannot be expected to hold in contexts involving deterministic relationships, including supervenience and part-whole relationships. Thus if we treat those relationships like causation we should expect to find massive violations of faithfulness, and indeed we do. (For example, in Figure 2, *P* is independent of *P** and *M** given *M*, but that independence does not follow from CFC. Likewise in Figure 7, the set {A, a_2, …, a_n} determines the value of a_1 and hence screens it off from b_1 (or indeed every other variable in the graph), an independency which is not entailed by CFC.) We agree with Baumgartner and Casini that this severely limits the usefulness of Gebharter's framework for the task of inferring constitutive relationships. But note that our framework entails the same violations of faithfulness as Gebharter's. such violations are also common in our framework. In the framework of hybrid models, the faithfulness condition amounts to the principle that every independency follows from HCM. (Call that condition HFC.) As the reader may verify, the fact that in Figure 3 *P* is independent of *P** and *M** given *M* is a violation of HFC, and so is the fact that b_1 is independent of a_1 given A, a_2,…, a_n in Figure 8. (Indeed, we suspect that *any* extension of the causal modeling framework to multilevel settings must involve violations of faithfulness, as faithfulness in general cannot be expected to hold in contexts involving deterministic dependence relationships.) This means means that we should not expect our framework to be particularly helpful for the task of causal inference in contexts involving supervenience or part-whole relationships. But as noted above our proposal is not intended as a method for causal discovery in those contexts. Our primary motivation, instead, is to offer a consistent representation of those structures that is both independently well-motivated and can make sense of the causal facts in those contexts. For that project the faithfulness assumption is not needed.

> “(M) A necessary and sufficient condition for *X* to be a (type-level) direct cause of *Y* with respect to a variable set **V** is that there be a possible intervention on *X* that will change *Y* or the probability distribution of *Y* when one holds fixed at some value all other variables Z_i in **V**.

Woodward (2003) also characterizes intervention variables as follows:

> “(IV) *I* is an intervention variable on *X* with respect to *Y* iff
>
> I.1. *I* causes X.
>
> I.2. *I* acts as a switch for all the other variables that cause *X*. (…)
>
> I.3. Any directed path from *I* to *Y* goes through *X*.
>
> I.4. *I* is (statistically) independent of any variable *Z* that causes *Y* and that is on a directed path that does not go through *X*.” (2003: 98)

An intervention on *X* corresponds to a value of an intervention variable that sets *X* at a specific value. I.3 and I.4 in (IV) are meant to exclude confounding manipulations and thereby to ensure that an intervention on *X* cannot be followed by a change in *Y* unless *X* causes *Y*. (M) and (IV) work well for variable sets that do not include non-causal dependencies, but becomes problematic in multilevel settings. In particular, in Gebharter’s framework for modelling those settings interventions on high-level properties become impossible, as Gebharter (2015) and especially Baumgartner (2009, 2013, 2018) have argued. Consider Figure 2. Because every cause of *M* must be statistically correlated with *P*, and *P* lies on a directed path ($P \to P^* \to M^*$) that does not go through *M*, every manipulation of *M* must violate condition I.4 for being an intervention on *M* with respect to *M**. Moreover, because every cause of *M* plausibly counts as a cause of *P*, there exists a directed path from any such cause to *M** that does not go through *M*, so that I.3 is violated as well. It follows that no cause of *M* can satisfy the conditions for being an intervention on *M* with respect to *M**. (Baumgartner argues that interventionism thereby gives rise to a sui generis exclusion argument, as only properties that can be targeted by interventions can be efficacious in this framework.) In response, Woodward (2015) has proposed an updated version of (IV) which he calls (IV*), and which bypasses the problem by incorporating explicit exclusion clauses for supervenience bases in its last two conditions. Those two conditions now read as follows:

> I.3* Any directed path from *I* to *Y* goes through *X*, or some variable in the supervenience base of *X*.
>
> I.4. *I* is (statistically) independent of any variable *Z* that causes *Y* and that is on a directed path that does not go through *X*, unless *Z* is in the supervenience base of *X*.

(IV*), as is easy to see, makes interventions on *M* perfectly possible in Figure 2. However, if we hold on to Figure 2 as the correct model of the situation, Woodward's proposal seems open to a charge of adhocness. If supervenience relations are treated as formally analogous as causal relations, why should supervenience bases be granted an exception that does not apply to (ordinary) causes?

But once we adopt the view that the relationship between mental properties and their realizers should be understood as one of mutual dependence, and that multilevel settings should be represented with hybrid models, the claim that a proper intervention on (e.g.) *M* need not and should not leave *P* untouched becomes perfectly natural: obviously, an intervention on a variable can and should be allowed to vary further variables that depend on its target.[26] Formally, a simple way to implement this idea is to substitute references to directed paths in (IV) with references to weakly directed paths, so that I.3 and I.4 are replaced with the following clauses:

> I.3**. Any weakly directed path from *I* to *Y* goes through *X*.
>
> I.4**. *I* is (statistically) independent of any variable *Z* that causes *Y* and that is on a weakly directed path that does not go through *X*.

[26] Kroedel (2019) argues that on an interventionist account of causation, *M* may be regarded as a (non-causal) difference-maker for *P*, as an intervention on *M* would change *P*. (The relationship here goes both ways, since an intervention on *P* would also lead to a change in *M*.) Our proposal can be seen as providing a precise formal implementation of this idea in the framework of causal modeling.

Call the resulting definition of intervention variables (IV**). As is easy to see, (IV**) makes it entirely possible for a cause of M to count as an intervention on M (with respect to M*). In particular, the fact that any such cause must be correlated with P is no issue anymore, as in our formalism the path relating P to M* is a weakly directed one that does go through M. In our view, the fact that our formalism offers an elegant and well-motivated way to think about interventions in contexts involving non-causal dependencies is a further important reason to regard it as the most plausible extension of the traditional causal modeling framework to those contexts.

9. Conclusion

Let us retrace our steps. On their standard interpretation, CBNs are not applicable to contexts that involve variables that are located at multiple levels and therefore non-causally depend on each other. But given the enormous fruitfulness of the causal modeling framework there are compelling reasons to bring it to bear on such settings, as this may shed light on longstanding philosophical issues concerning mechanisms, high-level causation and interventions. According to Gebharter's proposal, non-causal dependence relationships can be modelled on a par with causal arrows, so that there is no need to go beyond the well-understood formalism of CBNs to represent multi-level settings. We have argued, however that Gebharter's representational choice is poorly motivated given the structural dissimilarities between non-causal and causal dependence relationships – in particular the fact that non-causal dependence relationships are manipulable in both directions. This, we argued, suggests that causal modelers should represent those relationships as mutual dependence. But while several tools have already been developed in the causal modeling literature to represent symmetric dependence, none of them work well for multi-level settings. Thus our main contribution in this paper was to propose a new type of formalism - hybrid models – for representing symmetric dependence that is specifically suited to those settings. We proposed axioms for hybrid models that are theoretically well-motivated and empirically plausible (in the sense that they have the right statistical implications). We further explored the implications of our formalism for the status of high-level causation in non-reductive physicalism. The key difference with Gebharter's framework is that while the latter provides support for Kim's exclusion argument, our framework vindicates the causal efficacy of high-level multiply realizable properties. In addition, our framework offers a positive way to understand how high-level and low-level causation are compatible with one another. These results provide a powerful motivation for non-reductive physicalists to endorse our framework. Exclusion considerations also provide another argument in favor of our framework that should appeal to philosophers of all stripes: namely, Gebharter's framework makes all macrocausation impossible, whereas our proposal avoids this result. We closed with a brief discussion of interventions, showing how hybrid models offer a plausible and well-motivated way to think about interventions in multi-level settings. Our formalism is more complex than CBNs, in particular because it requires information about variable levels to be included in our models. But in light of the theoretical benefits of the framework we think this is a cost well worth paying, and that hybrid models have a good claim at being the best way to represent multi-level settings within the causal modeling framework.[27]

References

Baumgartner, M. (2009). Interventionist Causal Exclusion and Non-Reductive Physicalism. *International Studies in the Philosophy of Science*, *23*, 161–178. https://doi.org/10.1080/02698590903006909

Baumgartner, M. (2013). Rendering Interventionism and Non-Reductive Physicalism Compatible. *Dialectica*, *67*, 1–27.

Baumgartner, M. (2018). The Inherent Empirical Underdetermination of Mental Causation. *Australasian Journal of Philosophy*, *96*, 335–350. https://doi.org/10.1080/00048402.2017.1328451

[27] Special thanks to Vera Hoffmann-Kolss and Thomas Kroedel for extensive suggestions and comments and in particular to Alexander Gebharter who commented on several versions of the paper. We would also like to thank audiences at the University of Hamburg, the 2022 Conference of the German Society for Philosophy of Science (GWP) in Berlin and the 2022 workshop "(Non-)Reductionism in the Philosophy of Mind" at the SOPhiA conference in Salzburg.

Baumgartner, M., & Gebharter, A. (2016). Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. *The British Journal for the Philosophy of Science*, *67*, 731–756. https://doi.org/10.1093/bjps/axv003

Bennett, K. (2003). Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It. *Noûs*, *37*, 471–497.

Blanchard, T., Murray, D., & Lombrozo, T. (2022). Experiments on causal exclusion. *Mind & Language*, *37*, 1067–1089. https://doi.org/10.1111/mila.12343

Blanchard, T., & Schaffer, J. (2017). Cause without Default. In H. Beebee, C. Hitchcock, & H. Price (Eds.), *Making a Difference*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198746911.003.0010

Block, N. (2003). Do Causal Powers Drain Away? *Philosophy and Phenomenological Research*, *67*, 133–150.

Briggs, R. (2012). Interventionist Counterfactuals. *Philosophical Studies*, *160*, 139–166.

Casini, L., & Baumgartner, M. (2023). The PC Algorithm and the Inference to Constitution. *The British Journal for the Philosophy of Science*, *74*(2), 405–429. https://doi.org/10.1086/714820

Casini, L., Illari, P. M., Russo, F., & Williamson, J. (2011). Models for Prediction, Explanation and Control: Recursive Bayesian Networks. *THEORIA. An International Journal for Theory, History and Foundations of Science*, *26*, 5-33. https://doi.org/10.1387/theoria.784

Craver, C. F. (2007). *Explaining the Brain*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199299317.001.0001

Eronen, M., & Brooks, D. (2018). Levels of Organization in Biology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018). https://plato.stanford.edu/archives/spr2018/entries/levels-org-biology/.

Gebharter, A. (2017a). Causal Exclusion and Causal Bayes Nets. *Philosophy and Phenomenological Research*, *95*, 353–375. https://doi.org/10.1111/phpr.12247

Gebharter, A. (2017b). Uncovering Constitutive Relevance Relations in Mechanisms. *Philosophical Studies*, *174*, 2645–2666.

Gebharter, A. (2022). A Causal Bayes Net Analysis of Glennan's Mechanistic Account of Higher-Level Causation (and Some Consequences). *The British Journal for the Philosophy of Science*, *73*, 185–210. https://doi.org/10.1093/bjps/axz034

Geiger, D., & Pearl, J. (1989). *Logical and Algorithmic Properties of Conditional Independence and Qualitative Independence. Report CSD 870056, R-97-IIL*. University of California, Cognitive Systems Laboratory.

Hitchcock, C. (2012). Theories of Causation and the Causal Exclusion Argument. *Journal of Consciousness Studies*, *19*, 40–56.

Hoffmann-Kolss, V. (forthcoming). Interventionist Causal Exclusion and the Challenge of Mixed Models. In K. Robertson & A. Wilson (Eds.), *Levels of Explanation*. Oxford: Oxford University Press.

Hüttemann, A. (2021). *A Minimal Metaphysics for Scientific Practice*. Cambridge: Cambridge University Press.

Johnson, S. G. B., & Keil, F. C. (2014). Causal Inference and the Hierarchical Structure of Experience. *Journal of Experimental Psychology: General*, *143*, 2223–2241. https://doi.org/10.1037/a0038192

Kim, J. (1998). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Kim, J. (2003). Blocking Causal Drainage and Other Maintenance Chores with Mental Causation. *Philosophy and Phenomenological Research*, *67*, 151–176.

Kim, J. (2005). *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Kinney, D. (2023). Bayesian Networks and Causal Ecumenism. *Erkenntnis*, *88*, 147–172. https://doi.org/10.1007/s10670-020-00343-z

Kistler, M. (2013). The Interventionist Account of Causation and Non-causal Association Laws. *Erkenntnis*, *78*, 65–84. https://doi.org/10.1007/s10670-013-9437-4

Kroedel, T. (2019). *Mental Causation: A Counterfactual Theory*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108762717

Lauritzen, S. L., & Richardson, T. S. (2002). Chain Graph Models and their Causal Interpretations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 321–348. https://doi.org/10.1111/1467-9868.00340

Lauritzen, S. L., & Wermuth, N. (1989). Graphical Models for Associations between Variables, some of which are Qualitative and some Quantitative. *The Annals of Statistics*, *17*, 31–57.

List, C., & Menzies, P. (2009). Nonreductive Physicalism and the Limits of the Exclusion Principle. *Journal of Philosophy*, *106*, 475–502.

Loewer, B. (2002). Comments on Jaegwon Kim's Mind and the Physical World. *Philosophy and Phenomenological Research, 65*, 655–662. https://doi.org/10.1111/j.1933-1592.2002.tb00229.x

McDermott, M. (1995). Redundant Causation. *British Journal for the Philosophy of Science, 46*, 523–544.

Merricks, T. (2001). *Objects and Persons*. Oxford: Oxford University Press. https://doi.org/10.1093/0199245363.001.0001

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge: Cambridge University Press.

Richardson, T. (2003). Markov Properties for Acyclic Directed Mixed Graphs. *Scandinavian Journal of Statistics, 30*, 145–157.

Romero, F. (2015). Why there Isn't Inter-Level Causation in Mechanisms. *Synthese, 192*, 3731–3755.

Schaffer, J. (2016). Grounding in the Image of Causation. *Philosophical Studies*, *173,* 49–100. https://doi.org/10.1007/s11098-014-0438-1

Sloman, S. (2005). *Causal Models: How People Think About the World and Its Alternatives*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195183115.001.0001

Spirtes, P. (1995). Directed Cyclic Graphical Representations of Feedback Models. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence,* 491–498.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search* (2nd ed.). Cambridge, MA: MIT Press.

Steel, D. (2010). A New Approach to Argument by Analogy: Extrapolation and Chain Graphs. *Philosophy of Science*, *77*, 1058–1069. https://doi.org/10.1086/656543

Stern, R. (2019). Decision and Intervention. *Erkenntnis*, *84*, 783–804. https://doi.org/10.1007/s10670-018-9980-0

Stern, R. (2021). Causal concepts and temporal ordering. *Synthese*, *198*, 6505–6527. https://doi.org/10.1007/s11229-019-02235-4

Stern, R., & Eva, B. (2023). Anti-reductionist Interventionism. *The British Journal for the Philosophy of Science*, *74*, 241–267. https://doi.org/10.1086/714792

Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.

Woodward, J. (2008). Response to Strevens. *Philosophy and Phenomenological Research*, *77*, 193–212.

Woodward, J. (2015). Interventionism and Causal Exclusion. *Philosophy and Phenomenological Research*, *91*, 303–347.

Woodward, J. (2022). Modeling Interventions in Multi-Level Causal Systems: Supervenience, Exclusion and Underdetermination. *European Journal for Philosophy of Science*, *12*, 1–34. https://doi.org/10.1007/s13194-022-00486-6