A Modeling Approach for Mechanisms Featuring Causal Cycles

Alexander Gebharter and Gerhard Schurz*†

Mechanisms play an important role in many sciences when it comes to questions concerning explanation, prediction, and control. Answering such questions in a quantitative way requires a formal representation of mechanisms. Gebharter's "A Formal Framework for Representing Mechanisms?" suggests to represent mechanisms by means of arrows in an acyclic causal net. In this article we show how this approach can be extended in such a way that it can also be fruitfully applied to mechanisms featuring causal feedback.

1. Introduction. Questions concerning explanation, prediction, and control in the sciences are oftentimes answered by pointing at the system of interest's underlying mechanism and showing how causal interactions of this mechanism's parts bring about the phenomenon of interest. Mechanisms are typically characterized qualitatively. Glennan (1996), for example, defines a mechanism underlying a behavior as a "complex system which produces that behavior by of the interaction of a number of parts according to direct causal laws" (52). (For other prominent characterizations see, e.g., Machamer, Darden, and Craver [2000, 3] or Bechtel and Abrahamsen [2005, 423]).

For providing quantitatively precise mechanistic explanation/prediction and answering questions concerning the results of manipulations, however, a formal representation of mechanisms is required. Casini et al. (2011) suggest to model mechanisms by means of recursive Bayesian networks (RBNs). Gebharter (2014) highlights two problems with Casini et al.'s approach and

*To contact the authors, please write to: Alexander Gebharter, Düsseldorf Center for Logic and Philosophy of Science (DCLPS), Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany; e-mail: alexander.gebharter@gmail.com.

†This work was supported by Deutsche Forschungsgemeinschaft (DFG), research unit Causation | Laws | Dispositions | Explanation (FOR 1063). We thank Lorenzo Casini, David Danks, Christian J. Feldbacher, Clark Glymour, Marie I. Kaiser, Daniel Koch, Marcel Weber, and Naftali Weinberger for helpful remarks and discussions.

Philosophy of Science, 83 (December 2016) pp. 934–945. 0031-8248/2016/8305-0025$10.00
Copyright 2016 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/687876 Published online by Cambridge University Press

suggests the multilevel causal model (MLCM) approach as an alternative.[1] While Casini et al. represent mechanisms by a special kind of node of a Bayesian network (BN), Gebharter represents them by directed or bidirected causal arrows. The latter seems promising; it suggests, for example, to develop new methods for discovering submechanisms, that is, the causal structure inside the causal arrows (see, e.g., Murray-Watters and Glymour 2015). In addition, many results from the statistics and machine learning literature can be directly applied to models of mechanisms. Zhang (2008), for example, shows how the effects of interventions can be computed in models featuring bidirected arrows, and Richardson (2009) develops a factorization criterion equivalent to the *d*-connection condition for such models.

One of the shortcomings the RBN and the MLCM approach share is that they presuppose acyclicity, and thus, do not allow for a representation of mechanisms featuring feedback.[2] Clarke, Leuridan, and Williamson (2014) further develop the RBN approach in such a way that it can be applied to mechanisms featuring causal cycles. They distinguish between static and dynamic problems. Static problems are "situations in which a specific cycle reaches equilibrium . . . and where the equilibrium itself is of interest, rather than the process of reaching equilibrium." A dynamic problem is a "situation in which it is the change in the values of variables over time that is of interest" (sec. 6). Clarke et al. suggest to solve static problems on the basis of the notion of *d*-separation (Pearl 2000, sec. 1.2.3) and dynamic problems by means of dynamic Bayesian networks (DBNs). In this article, we follow their example and demonstrate how the MLCM approach for representing mechanisms can be modified and extended in a similar way.

The article is structured as follows: in section 2 we introduce the causal modeling framework used in the article. In section 3 we give an overview of the MLCM approach. In sections 4.1 and 4.2 we demonstrate by means of a simple toy mechanism how the MLCM approach can be modified in such a way that it can be applied to static and dynamic problems, respectively. Both modifications mirror Clarke et al.'s (2014) suggestions for solving static and dynamic problems without sharing certain shortcomings.

2. Causal Nets. A causal net (or model) is a triple $\langle V, E, P\rangle$, where $\langle V, E\rangle$ is a directed graph providing causal information about the elements of V, V is a set of random variables, and E is a binary relation on V that is interpreted as direct causal connection relative to V. Set V's elements are called the graph's

1. For an attempt to defend the RBN approach against the objections made in Gebharter (2014), see Casini (2016). For another problem with the RBN approach, see Gebharter (2016).

2. For other problems that may arise in general for attempts to model mechanisms by means of BNs, see Kaiser (2016) and Weber (2016).

https://doi.org/10.1086/687876 Published online by Cambridge University Press

vertices, E's elements its edges. And P is a probability distribution over V representing regularities produced by the causal structure underlying V.

Causal connections between variables are represented by directed and bidirected arrows: "$X \rightarrow Y$" means that X is a direct cause of Y, and "$X \leftrightarrow Y$" means that X and Y are effects of a common cause not included in V. Causal models are assumed to not feature self-edges $X \rightarrow X$ or $X \leftrightarrow X$. The set of Y's parents is $\mathrm{Par}(Y) = \{X \in V : X \rightarrow Y\}$. A chain of $n \geq 1$ edges (of any kind) connecting two variables X and Y is called a path between X and Y if it does not go through any variable more often than once. (Note that π's being a path between X and Y allows that $X = Y$.) A path $X \rightarrow \ldots \rightarrow Y$ is called a directed path from X to Y; X is called a cause of Y and Y an effect of X. A variable Z lying on a path $X \rightarrow \ldots \rightarrow Z \rightarrow \ldots \rightarrow Y$ is called an intermediate cause lying on this path. A path $X \leftarrow \ldots \leftarrow Z \rightarrow \ldots \rightarrow Y$ is called a common cause path with Z as a common cause of X and Y lying on this path. A path connecting X and Y containing a subpath $Z_i \,@\rightarrow Z_j \leftarrow @\, Z_k$ is called a collider path connecting X and Y, and Z_j is called a collider lying on this path.[3] A path between X and Y indicates a common cause path if it either is a common cause path or a collider-free path that contains a bidirected edge. A path $X \rightarrow \ldots \rightarrow X$ is called a causal cycle. A graph is called cyclic if it features causal cycles; it is called acyclic otherwise. Likewise for causal models.

For now we only require the causal net approach's most central axiom, the causal Markov condition (CMC). A model $\langle V, E, P \rangle$ satisfies CMC if and only if (iff) every $X \in V$ is probabilistically independent of its noneffects conditional on its direct causes (Spirtes, Glymour, and Scheines 1993, 54). If an acyclic causal model satisfies CMC, then its graph determines the following Markov factorization (54):

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Par}(X_i)). \tag{1}$$

3. The Multilevel Causal Model Approach. The MLCM approach is based on the simple idea that mechanisms are devices bringing about certain input-output behaviors (cf. Bechtel 2007, sec. 3; Craver 2007, 145). This suggests a representation of mechanisms by a causal model's arrows. The variables at the arrows' tails stand for the mechanism's input, the variables at the arrows' heads stand for the mechanism's output, and the arrows represent the not-further-specified mechanism. A graph describing such a mechanism can be supplemented by a probability distribution P that quantitatively describes the system's behavior.

Mechanistic explanation requires investigating how a mechanism produces the phenomenon of interest; it requires a more detailed description of

3. The metasymbol "@" stands for an arrowtail or an arrowhead.

https://doi.org/10.1086/687876 Published online by Cambridge University Press

the underlying causal structure producing that phenomenon. One can give such an explanation by supplementing a causal model M, whose graph's arrows represent a mechanism, by another causal model M' that contains new variables describing the behaviors of some parts of the mechanism. So, metaphorically speaking, we are "zooming" into the device represented by the arrows. However, it must be assured that the more detailed causal model M' fits to the original model M with respect to its causal structure and its probability distribution. The following notion of a restriction states conditions for such a fit (Gebharter 2014, 147):[4] $M = \langle V, E, P\rangle$ is a restriction of $M' = \langle V', E', P'\rangle$ iff $V \subset V'$, $P'\!\uparrow V = P$,[5] the following two conditions hold for all $X, Y \in V$, and no path not implied by these conditions is in $\langle V, E\rangle$:

1. If there is a path from X to Y in $\langle V', E'\rangle$ and no vertex on this path different from X and Y is in V, then $X \rightarrow Y$ in $\langle V, E\rangle$.
2. If X and Y are connected by a path π in $\langle V', E'\rangle$ indicating a common cause path and no vertex on π different from X and Y is in V, then $X \leftrightarrow Y$ in $\langle V, E\rangle$.

This notion tells us which causal models M' are candidates for mechanistically explaining phenomena described by a less detailed model M. It also tells us how we can marginalize out variables from M' while preserving information about the causal and probabilistic relationships among variables in V provided by M'. For a detailed motivation of this notion of a restriction, see Gebharter (2014, 147–48).

We can now define an MLCM as a structure $\langle M_1, \ldots, M_n\rangle$ such that every causal model M_i with $i > 1$ is a restriction of M_1, while M_1 satisfies CMC (Gebharter 2014, 148). The latter condition reflects a basic assumption of the causal nets approach, that is, that for explaining a probability distribution P, reference to an underlying causal structure satisfying CMC is required (cf. Spirtes et al. 1993, sec. 6.1).[6]

Let us briefly illustrate by means of figure 1 how MLCMs can be used for modeling mechanisms. Model M_2 describes the mechanism's top level. The mechanism has two input variables (X_1, X_2) and three output variables (Y_1, Y_2, Y_3). The arrows stand for the not-further-specified mechanism. Mechanistic explanation of a certain phenomenon, for example, of an input-output behavior $P(y_1, y_2, y_3 \mid x_1, x_2)$, requires a more detailed story about what is

4. This definition is inspired by Steel (2005, 12). We thank Clark Glymour for pointing out that the marginalization method this definition provides is essentially a "slim" version of the mixed ancestral graph representation developed by Richardson and Spirtes (2002) for latent variable models.

5. Here, $P' \uparrow V$ is the restriction of P' to V.

6. Note that CMC will typically be violated by models featuring bidirected arrows.

https://doi.org/10.1086/687876 Published online by Cambridge University Press

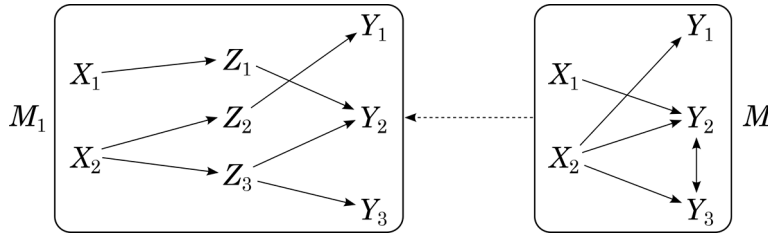


Figure 1.

happening within the mechanism, that is, within the system represented by the arrows. This story is told by M_1. Model M_1 features three new variables (Z_1, Z_2, Z_3) describing parts of the mechanism. Model M_1's causal structure tells us over which causal paths through the mechanism X_1 and X_2 cause Y_1, Y_2, and Y_3. Model M_2 is a restriction of M_1. And M_1 is assumed to satisfy CMC.

4. Modeling Mechanisms with Causal Cycles. We introduce the following toy mechanism for investigating the question of how to model mechanisms featuring causal cycles within the MLCM approach: a simple temperature regulation system, where *OT* stands for the outside temperature, *IT* for the inside temperature, and *CK* for a control knob. The behavior of interest is that *IT* is relatively insensitive to *OT* when $CK = on$, that is, that $P(it \mid ot, CK = on) \approx P(it \mid CK = on)$ for arbitrary *OT* and *IT* values.

A simple input-output representation of this mechanism would be a causal model M_2 with the graphical structure $OT \rightarrow IT \leftarrow CK$. A mechanistic explanation of $P(it \mid ot, CK = on) \approx P(it \mid CK = on)$ by means of an MLCM would require connecting M_2 to a more detailed model M_1 satisfying CMC. Since the system represented is a self-regulatory system, M_1 is expected to feature a cycle $IT \rightarrow \ldots \rightarrow IT$. But cyclic causal models do have some problems with CMC. While CMC can, in principle, be applied to cyclic causal models, it turns out to be inadequate. Let us illustrate this by means of the following example borrowed from Spirtes et al. (1993, 359): Suppose a causal model with the structure $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_1$ satisfies CMC. Then CMC implies no probabilistic independence. But since $\{X_2, X_4\}$ blocks all causal paths connecting X_1 and X_3 and correlations are assumed to arise only because of causal connections, no probabilistic influence from X_1 should reach X_3 when X_2's and X_4's values are fixed. So conditionalizing on $\{X_2, X_4\}$ should render X_1 and X_3 probabilistically independent.

The remainder of this section shows by means of the exemplary mechanism introduced above how the MLCM approach can be modified in such a way that it can be used to model mechanisms featuring causal cycles. To this end, as already mentioned, we have to distinguish between static and

https://doi.org/10.1086/687876 Published online by Cambridge University Press

dynamic problems. Solving the static problem requires a model capable of explaining $P(it \mid ot, CK = on) \approx P(it \mid CK = on)$ when the underlying cycle $IT \rightarrow \ldots \rightarrow IT$ has reached equilibrium. Solving the dynamic problem requires a model that allows for an explanation of how $IT \rightarrow \ldots \rightarrow IT$ produces $P(it \mid ot, CK = on) \approx P(it \mid CK = on)$ over a period of time.

4.1. Solving the Static Problem. To solve the static problem, we have to modify the definition of an MLCM: instead of requiring that the most detailed causal model M_1 of the MLCM satisfies CMC, we rather require M_1 to satisfy the *d*-connection condition. A model $\langle V, E, P \rangle$ satisfies the *d*-connection condition iff for every dependence of variables X and Y given some $Z \subseteq \{X, Y\}$ there is a *d*-connection between X and Y given Z (Schurz and Gebharter 2016). Variables X and Y are *d*-connected given Z iff there is a path π connecting X and Y such that no intermediate or common cause on π is in Z, while every collider on π is in Z or has an effect in Z. Variables X and Y are *d*-separated by Z otherwise.

The *d*-connection condition is equivalent to CMC for acyclic causal models (Lauritzen et al. 1990). This equivalence reveals the full content of CMC: whenever a causal model satisfies CMC, then every dependence can be explained by some causal connection in the model, and every independence can be explained by missing causal connections in the model. The *d*-connection condition's clear advantage over CMC is that it implies the independencies to be expected when applied to causal cycles (Spirtes 1995; Pearl and Dechter 1996). To demonstrate this, assume that the causal model $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow X_4 \rightarrow X_1$ discussed earlier in section 4 satisfies the *d*-connection condition. As we saw in section 4, CMC implies no independencies for this causal model. But since X_1 and X_3 are *d*-separated by $\{X_2, X_4\}$, the *d*-connection condition implies the expected independence of X_1 and X_3 given $\{X_2, X_4\}$.

Let us now see how the static problem can be solved for our exemplary mechanism within the modified MLCM approach. The static problem concerns our exemplary mechanism when it has reached equilibrium. The system can be represented by the two-stage MLCM depicted in figure 2. Model M_2 represents the system at the top level. Variables *OT* and *CK* are directly causally relevant to *IT*. Model M_1 provides more detailed information about what is happening within the mechanism: the inside temperature is

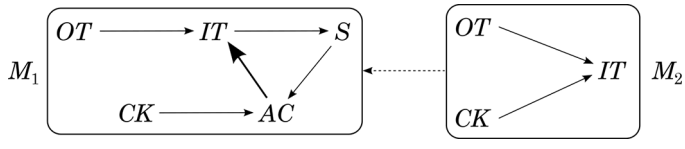


Figure 2.

https://doi.org/10.1086/687876 Published online by Cambridge University Press

measured by a temperature sensor (S), which is directly causally relevant to an air conditioner (AC), which, in turn, is under direct causal influence of CK.

Model M_1 is assumed to satisfy the d-connection condition, and M_2 is a restriction of M_1. The MLCM mechanistically explains why OT is relatively insensitive to IT when the cycle $IT \rightarrow S \rightarrow AC \rightarrow IT$ has reached equilibrium and $CK = on$, that is, why $P(it \mid ot, CK = on) \approx P(it \mid CK = on)$ holds. If CK is off, then AC is off, and there is no self-regulation due to the causal cycle $IT \rightarrow S \rightarrow AC \rightarrow IT$. Thus, OT will have an influence on IT. But when CK is set to one of its on values, then AC responds to S according to CK's adjustment. Since $AC \rightarrow IT$ overwrites $OT \rightarrow IT$ when $CK = on$, IT's value is robust with respect to changes of OT's value when $CK = on$. This overwriting property of $AC \rightarrow IT$ is represented by the bold arrow in figure 2.

Let us finally mention some open problems. First, cyclic models possibly featuring bidirected arrows do not admit the Markov factorization. Since we assume the d-connection condition to hold, they do, however, factor according to the following equation:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid d\mathrm{Sep}(X_i)). \tag{2}$$

The set $d\mathrm{Sep}(X_i)$ is constructed as follows: Let $\mathrm{Pred}(X_i)$ be the set of X_i's predecessors in the ordering $X_1, \ldots, X_n$. Now search for sets $d\mathrm{Pred}(X_i) \subseteq \mathrm{Pred}(X_i)$ such that $U = \mathrm{Pred}(X_i) \backslash d\mathrm{Pred}(X_i)$ d-separates X_i from all elements of $d\mathrm{Pred}(X_i)$. (Note that U may be empty.) If there are no such sets $d\mathrm{Pred}(X_i)$, then identify $d\mathrm{Sep}(X_i)$ with X_i's predecessors $\mathrm{Pred}(X_i)$. If there are such sets $d\mathrm{Pred}(X_i)$, then take one of the largest of these sets and identify $d\mathrm{Sep}(X_i)$ with the corresponding separator set $U = \mathrm{Pred}(X_i) \backslash d\mathrm{Pred}(X_i)$. For the ordering $P(OT, CK, IT, AC, S)$, for example, the joint distribution of M_1 factors as $P(OT) \cdot P(CK) \cdot P(IT \mid OT, CK) \cdot P(AC \mid OT, CK, IT) \cdot P(S \mid CK, IT, AC)$.

Equation (2) has two disadvantages. First, it depends on an ordering of variables. Second, a probability distribution that factors according to equation (2) may not imply all independencies implied by the d-connection condition. For example, it does not imply an independence between OT and S conditional on $\{IT, AC\}$, although OT and S are d-separated by $\{IT, AC\}$. One open problem is to find out whether there is an order-independent factorization criterion equivalent with the d-connection condition. Another open problem is search. Causal discovery of the latent structure inside a mechanism's causal arrows in the possible presence of feedback loops can be expected to be an even harder problem than discovery without feedback (cf. Murray-Watters and Glymour 2015).

We conjecture that effects of interventions for cyclic graphs possibly featuring bidirected arrows can be computed as usual. To compute postintervention probabilities $P(z \mid \hat{x})$ for an instantiation z of a set of variables Z, one

https://doi.org/10.1086/687876 Published online by Cambridge University Press

needs, first, to delete all the arrows with an arrowhead pointing at X from the graph.[7] Second, use d-separation information provided by the manipulated graph to compute $P(z \mid \hat{x})$.

Before we take a look at how to solve the dynamic problem, let us briefly discuss the relationship of the solution to the static problem suggested above with Clarke et al.'s (2014) solution. Although both approaches use Pearl's (2000) notion of d-separation instead of CMC to account for cycles, the structures used for probabilistic reasoning differ in the two approaches. Clarke et al. use the "true" cyclic graph to construct an equilibrium network, that is, a BN that is then used "to model the probability distribution of the equilibrium solution" (2014, sec. 6.1). In our view, this move has at least two shortcomings:

i) Independencies implied by the d-connection condition and the original cyclic causal structure may not be implied by the equilibrium network. We illustrate this by means of our model M_1, whose equilibrium network could be the one depicted in figure 3. (See Clarke et al. [2014], sec. 6.1, for details on how to construct equilibrium networks.) Now note that *OT* and *CK*, for example, are not d-separated in the equilibrium network. So the equilibrium network's graph does not capture the independence between *OT* and *CK* implied by the d-connection condition and the fact that *OT* and *CK* are d-separated in M_1's graph.
ii) Since the arrows of the equilibrium network do not capture the "true" causal relations anymore, it cannot be used for predicting the effects of interventions. To illustrate this, assume we are interested in the postintervention probability $P(s|\widehat{ck})$ in our model M_1. In case we use M_1's graph for computing this probability, we arrive at $P(s|\widehat{ck}) = P(s|ck)$. If we use the equilibrium network's graph, however, we arrive at $P(s|\widehat{ck}) = P(s)$. But since the control knob is causally relevant for the sensor, $P(s|ck)$ will not equal $P(s)$ when intervening on *CK*.

4.2. Solving the Dynamic Problem. Solving the dynamic problem requires an extension of the MLCM approach that allows for representing the system's behavior over a period of time. Clarke et al. (2014) model such behavior by means of DBNs (cf. Murphy 2002). The basic idea behind this move is to roll out the causal cycles over time. We use dynamic causal models (DCMs) that also allow for bidirected arrows.

A DCM M is a quadruple $\langle V, E, P, t : V \rightarrow \mathbb{N}^+ \rangle$, where V is a set of infinitely many variables $X_{1,1}, \ldots, X_{n,1}, X_{1,2}, \ldots, X_{n,2}, \ldots$. The variables $X_{i,1}$ (with $1 \leq i \leq n$) describe the system at its initial state (stage 1), the variables

7. Here, "$\hat{x}$" is shorthand for "X is forced to take value x by intervention."

https://doi.org/10.1086/687876 Published online by Cambridge University Press

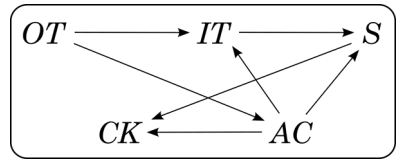


Figure 3.

$X_{i,t+1}$ (with $1 \leq i \leq n$), the system at later stages $t + 1$. The DCMs we consider involve some idealization: directed arrows only connect variables at different stages, and if there is a directed arrow going from a variable $X_{i,t}$ to a variable $X_{j,t+u}$ for some stage t, then for every stage t there is such a directed arrow going from $X_{i,t}$ to $X_{j,t+u}$. So the pattern of directed arrows between stages t and $t + u$ is always the same.

What one ideally wants is a DCM $\langle V, E, P, t\rangle$ with the following additional properties: (i) arrows do not skip stages, (ii) bidirected arrows occur only between variables of one and the same stage, (iii) every two stages t_i, t_j (with $i, j > 1$) share the same pattern of bidirected arrows, and (iv) $P(X_{i,t} \mid \mathrm{Par}(X_{i,t})) = P(X_{i,t+1} \mid \mathrm{Par}(X_{i,t+1}))$ holds for all $X_{i,t} \in V$ with $t > 1$. For a finite segment of such an "ideal" DCM, see figure 4. The depicted graph's first stage features more bidirected arrows than later stages. These additional bidirected arrows account for correlations between $X_{1,1}$ and $X_{2,1}$, $X_{2,1}$ and $X_{3,1}$, and $X_{1,1}$ and $X_{3,1}$ because of not-represented past common causes (of the kind described by variables in V).

Let us now come back to the question of how the dynamic problem can be solved within the MLCM approach. The phenomenon we are interested in is that IT is relatively robust to variations of OT over a period of time when $CK = on$. Our simple temperature regulation system can be modeled by a two-stage MLCM $\langle M_1, M_2\rangle$ (see fig. 5 for a finite segment). The mechanism's top level is represented by M_2, which is a restriction of M_1. Model M_1, which is assumed to satisfy the d-connection condition, provides more detailed information about the mechanism bringing about the phenomenon of interest.

When adding new intermediate causes, we will typically also add new stages. In our example, we added two new variables (S and AC) and two new stages between consecutive stages of M_2 arriving at $IT_{t^*} \rightarrow S_{t^*+1} \rightarrow$

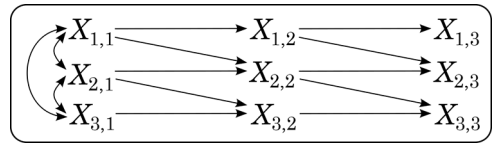


Figure 4.

https://doi.org/10.1086/687876 Published online by Cambridge University Press

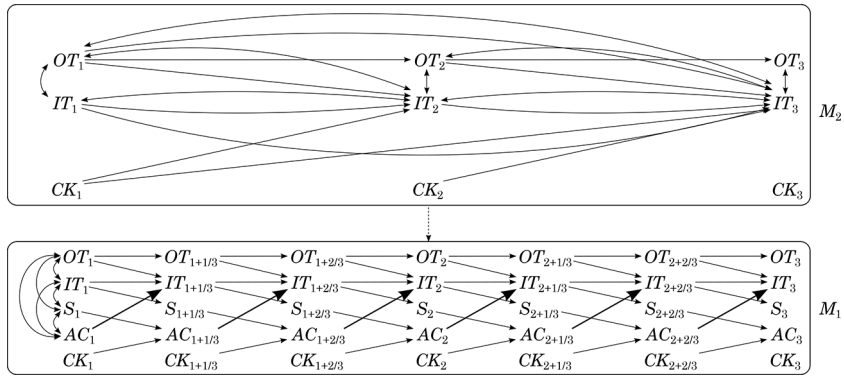


Figure 5.

AC_{t^*+2} in M_1. We assume the intervals between M_1's stages to correspond to the time the causal processes we are interested in require to bring about their effects. To guarantee that M_1's and M_2's probability distributions fit together, we require $t^* = t, t^* + 1 = t + 1/3, t^* + 2 = t + 2/3, t^* + 3 = t + 3/3$, and so on (where t stands for M_2's and t^* for M_1's stages).

Now M_1 provides information about the causal structure within the temperature regulation system. Variables *OT* and *IT* are directly causally relevant to themselves at the next stage. Variable *IT* also causally depends on *OT* and *AC*, while *S* depends only on *IT*, *AC* only on *S* and *CK*, and *CK* on no variable in the model. The bidirected arrows at stage 1 account for dependencies to be expected because of not-represented common causes.

Here we assumed that model M_1 is especially nice, that is, that it satisfies (i)–(iv) discussed a few paragraphs above. Unfortunately, model M_2 is not that nice. Since we marginalized out *S* and *AC* and there were directed paths from OT_{t^*} to IT_{t^*+6} and from IT_{t^*} to IT_{t^*+6} all going through S_{t^*+3} or AC_{t^*+3} in M_1, M_2 features directed arrows $OT_t \rightarrow IT_{t+2}$ and $IT_t \rightarrow IT_{t+2}$ skipping stages. Since there were paths indicating a common cause path between OT_{t^*} and IT_{t^*+6} going through S_{t^*+3} or AC_{t^*+3} in M_1, M_2 features bidirected arrows $OT_t \leftrightarrow IT_{t+2}$. Note that there are also bidirected arrows between OT_t and IT_{t+1} and between IT_t and IT_{t+1}.

Now the MLCM mechanistically explains why *IT* is relatively robust with respect to *OT* changes when $CK = on$ over a period of time. If *CK* is off over several stages, then also *AC* is off, and there is no regulation of *IT* over paths $IT_{t^*} \rightarrow S_{t^*+1} \rightarrow \mathrm{AC}_{t^*+2} \rightarrow IT_{t^*+3}$; *IT*'s value will increase and decrease (with a slight time lag) with *OT*'s value. If, however, *CK* is fixed to one of its on values over several stages, then over several stages AC_{t^*+1} responds to S_{t^*} according to CK_{t^*}'s adjustment. Now the crucial control mechanism consists of IT_{t^*+1} and its parents OT_{t^*}, IT_{t^*}, and AC_{t^*}. The bold

https://doi.org/10.1086/687876 Published online by Cambridge University Press

arrows $AC_{t^*} \rightarrow IT_{t^*+1}$ in figure 5 overwrite $OT_{t^*} \rightarrow IT_{t^*+1}$ and $IT_{t^*} \rightarrow IT_{t^*+1}$ when $CK_{t^*} = \text{on}$; that is, $P_{\text{CK}_{t^*}=\text{on}}(\text{it}_{t^*+1} \mid \text{ac}_{t^*}, u_{t^*}) \approx P_{\text{CK}_{t^*}=\text{on}}(it_{t^*+1} \mid \text{ac}_{t^*})$ holds, where $U_{t^*} \subseteq \{OT_{t^*}, IT_{t^*}\}$. This control mechanism will, after a short period of time, cancel deviations of *IT*'s value from *CK*'s adjustment brought about by *OT*'s influence.

Here are some possible open problems: first, some of the arrows in M_2 may seem to misrepresent the "true" causal processes going on inside the temperature regulation system. There is, for example, a directed arrow going from CK_t to IT_{t+1} but no directed arrow from CK_t to AC_{t+1}, although *CK* can clearly influence *IT* only through *AC*. This is a typical problem arising for dynamic models. One can, however, learn something about M_1's structure from M_2: the (direct or indirect) cause-effect relationships among variables in M_2 will also hold for M_1. Another problem is, again, search. For solutions of several discovery problems involving time series, see, for example, Danks and Plis (2014). Finally, factorization and interventions: since our DCMs do not feature feedback loops, we conjecture that Richardson's (2009) factorization criterion and Zhang's (2008) results about how to compute the effects of interventions in models with bidirected arrows can be fruitfully applied to DCMs.

Let us finally have a look at how our solution to the dynamic problem relates to the one suggested by Clarke et al. (2014). Both modeling strategies use the same basic idea, that is, to roll out the cycles over time. While the arrows of the DCMs we use are intended to capture the "true" causal relations between variables of interest, the directed arrows in Clarke et al.'s DBNs surprisingly are not intended to represent the "true" causal relationships (cf. sec. 6.2). Thus, their models share problem (ii) discussed at the end of section 4.1 with the equilibrium network they use for solving static problems: the model cannot be used to compute the effects of interventions.

5. Conclusion. Clarke et al. (2014) have extended Casini et al.'s (2011) RBN approach for modeling mechanisms in such a way that it can be applied to mechanisms featuring causal feedback. In this article we followed their example and showed how the MLCM approach can be modified in a similar way. Like Clarke et al. we distinguish between static and dynamic problems when it comes to modeling mechanisms with causal cycles. Our solutions to both problems within the MLCM approach mirror Clarke et al.'s solutions for the RBN approach while avoiding several problems. The MLCM approach can be used for modeling mechanisms whose causal cycles have reached equilibrium (i.e., static problems) by introducing the requirement that the most detailed causal model M_1 has to satisfy the *d*-connection condition instead of CMC. The dynamic problem, which concerns the development of the system over a period of time, can be solved within the MLCM approach by using DCMs. Both solutions, however, come with new challenges, whose investigation we leave to future research.

https://doi.org/10.1086/687876 Published online by Cambridge University Press

REFERENCES

Bechtel, W. 2007. "Reducing Psychology While Maintaining Its Autonomy via Mechanistic Explanation." In *The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience, and Reduction*, ed. M. Schouton and H. L. de Jong, 172–98. Oxford: Blackwell.
Bechtel, W., and A. Abrahamsen. 2005. "Explanation: A Mechanist Alternative." *Studies in History and Philosophy of Biological and Biomedical Sciences* 36:421–41.
Casini, L. 2016. "How to Model Mechanistic Hierarchies." *Philosophy of Science*, in this issue.
Casini, L., P. M. lllari, F. Russo, and J. Williamson. 2011. "Models for Prediction, Explanation and Control: Recursive Bayesian Networks." *Theoria* 26 (70): 5–33.
Clarke, B., B. Leuridan, and J. Williamson. 2014. "Modelling Mechanisms with Causal Cycles." *Synthese* 191 (8): 1651–81.
Craver, C. 2007. *Explaining the Brain*. Oxford: Clarendon.
Danks, D., and S. Plis. 2014. "Learning Causal Structure from Undersampled Time Series." Paper presented at the NIPS 2013 Workshop on Causality.
Gebharter, A. 2014. "A Formal Framework for Representing Mechanisms?" *Philosophy of Science* 81 (1): 138–53.
———. 2016. "Another Problem with RBN Models of Mechanisms." *Theoria* 31 (2): 177–88.
Glennan, S. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44 (1): 49–71.
Kaiser, M. I. 2016. "On the Limits of Causal Modeling: Spatially-Structurally Complex Biological Phenomena." *Philosophy of Science*, in this issue.
Lauritzen, S. L., A. P. Dawid, B. N. Larsen, and H. G. Leimer. 1990. "Independence Properties of Directed Markov Fields." *Networks* 20 (5): 491–505.
Machamer, P., L. Darden, and C. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25.
Murphy, K. P. 2002. *Dynamic Bayesian Networks*. Berkeley: University of California Press.
Murray-Watters, A., and C. Glymour. 2015. "What's Going on Inside the Arrows? Discovering the Hidden Springs in Causal Models." *Philosophy of Science* 82 (4): 556–86.
Pearl, J. 2000. *Causality*. 1st ed. Cambridge: Cambridge University Press.
Pearl, J., and R. Dechter. 1996. "Identifying Independencies in Causal Graphs with Feedback." In *Proceedings of the 12th International Conference on Uncertainty in Artificial Intelligence*, 420–26. San Francisco: Kaufmann.
Richardson, T. 2009. "A Factorization Criterion for Acyclic Directed Mixed Graphs." In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, ed. J. Bilmes and A. Ng, 462–70. Arlington, VA: Association for Uncertainty in Artificial Intelligence.
Richardson, T., and P. Spirtes. 2002. "Ancestral Graph Markov Models." *Annals of Statistics* 30 (4): 962–1030.
Schurz, G., and A. Gebharter. 2016. "Causality as a Theoretical Concept: Explanatory Warrant and Empirical Content of the Theory of Causal Nets." *Synthese* 193 (4): 1073–1103.
Spirtes, P. 1995. "Directed Cyclic Graphical Representations of Feedback Models." In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, 491–98. San Francisco: Kaufmann.
Spirtes, P., C. Glymour, and R. Scheines. 1993. *Causation, Prediction, and Search*. 1st ed. Dordrecht: Springer.
Steel, D. 2005. "Indeterminism and the Causal Markov Condition." *British Journal for the Philosophy of Science* 56 (1): 3–26.
Weber, M. 2016. "On the Incompatibility of Dynamical Biological Mechanisms and Causal Graphs." *Philosophy of Science*, in this issue.
Zhang, J. 2008. "Causal Reasoning with Ancestral Graphs." *Journal of Machine Learning Research* 9:1437–74.

https://doi.org/10.1086/687876 Published online by Cambridge University Press