Robustness of Model Predictions under Extension

Abstract

Mathematical models of the real world are simplified representations of complex systems. A caveat to using mathematical models is that predicted causal effects and conditional independences may not be robust under model extensions, limiting applicability of such models. In this work, we consider conditions under which qualitative model predictions are preserved when two models are combined. Under mild assumptions, we show how to use the technique of *causal ordering* to efficiently assess the robustness of qualitative model predictions. We also characterize a large class of model extensions that preserve qualitative model predictions. For dynamical systems at equilibrium, we demonstrate how novel insights help to select appropriate model extensions and to reason about the presence of feedback loops. We apply our ideas to a viral infection model with immune responses.

1 INTRODUCTION

There are several interesting systems for which the causal relations and Markov properties cannot be modelled by Structural Causal Models (SCMs) [Pearl, 2009, Bongers et al., 2021]. The causal ordering algorithm, first introduced by Simon [1953], can be used to better understand the qualitative model predictions of these systems [Blom et al., 2021]. In this paper, we take a closer look at what happens to these predictions when two systems are combined. Particularly, we give conditions under which properties of the whole system can be understood in terms of properties of its parts. We discuss how a holistic approach towards causal modelling may result in novel insights when we derive and test the predictions of systems for which new properties emerge from the combination of its parts.

In the first part of the paper, we focus on the practical issue of assessing whether qualitative model predictions are robust under model extensions. We revisit the observations of De Boer [2012] who demonstrated that qualitative predictions of a certain viral infection model change dramatically when the model is extended with extra equations describing simple immune responses. To assess the robustness of predicted causal relations or conditional independences under such an alteration of the model, it is useful to characterize a class of model extensions that lead to unaltered qualitative model predictions. In this work, we propose the technique of causal ordering [Simon, 1953] as an efficient method to assess the robustness of qualitative causal predictions. Under mild conditions, this allows us to characterize a large class of model extensions that preserve qualitative causal predictions. We also consider the class of models that are obtained from the equilibrium equations of dynamical models where each variable is *self-regulating*. For this class, we show that the predicted presence of causal relations and absence of conditional independences is robust when the model is extended with new equations.

Key aspects of the scientific method include generating a model or hypothesis that explains a phenomenon, deriving testable predictions from this model or hypothesis, and designing an experiment to test these predictions in the real world. The promise of causal discovery algorithms is that they are able to learn causal relations from a combination of background knowledge and data. The general idea of many constraint-based approaches (e.g. PC or FCI and variants thereof [Spirtes et al., 2000, Zhang, 2008, Colombo et al., 2012]) is to exploit information about conditional independences in a probability distribution to construct an equivalence class of graphs that encode certain aspects of the probability distribution, and then draw conclusions about the causal relations from the graphs. There is a large amount of literature concerning particular algorithms for which the learned structure expresses causal relations under certain conditions (e.g. linearity, causal sufficiency, absence of feedback loops), see for example [Richardson and Spirtes, 1999, Spirtes et al., 2000, Lacerda et al., 2008, Zhang, 2008,

Submitted to the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022). **To be used for reviewing only**.

Colombo et al., 2012, Hyttinen et al., 2012, Forré and Mooij, 2018, Strobl, 2019, Mooij and Claassen, 2020]. In the last part of this paper, our main interest is in dynamical models with the property that graphs representing relations between variables by encoding the conditional independences of their equilibrium distribution should not be interpreted causally at all. For the case that a model for a subsystem is given, we present novel insights that enable us to reject model extensions based on conditional independences in equilibrium data of the subsystem. We demonstrate how this approach allows us to reason about the presence of variables that are not self-regulating and feedback mechanisms that involve unobserved variables from the equilibrium distribution of certain dynamical models.

1.1 CAUSAL ORDERING GRAPH

Here, we give a concise introduction to the technique of causal ordering, introduced by Simon [1952].[1] In short, the causal ordering algorithm takes a set of equations as input and returns a *causal ordering graph* that encodes the effects of interventions and a *Markov ordering graph* that implies conditional independences between variables in the model [Theorem 17, Blom et al., 2021]. Compared with the popular framework of structural causal models [Pearl, 2009], the distinction between the causal ordering and Markov ordering graphs does not provide new insights for acyclic models, but it results in non-trivial conclusions for models with feedback, as suggested in the discussion in Section 2.4 and explained in detail in [Blom et al., 2021].

We consider models consisting of equations F that contain endogenous variables V, independent exogenous random variables W, and (constant, exogenous) parameters P. The structure of equations and the endogenous variables that appear in them can be represented by the *associated bipartite graph* $\mathcal{B} = \langle V, F, E\rangle$, where each endogenous variable is associated with a distinct vertex in V, and each equation is associated with a distinct vertex in F. There is an edge $(v-f) \in E$ if and only if variable $v \in V$ appears in equation $f \in F$. The causal ordering algorithm constructs a *directed cluster graph* $\langle \mathcal{V}, \mathcal{E}\rangle$, where $\mathcal{V}$ is a partition of vertices V into clusters and $\mathcal{E}$ is a set of directed edges from vertices in V to clusters in $\mathcal{V}$. Given a bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ with a perfect matching[2] M, the causal ordering algorithm proceeds with the following three steps [Nayak, 1995, Blom et al., 2021]:

1. For $v \in V$, $f \in F$ orient edges $(v - f)$ as $(v \leftarrow f)$ when $(v - f) \in M$ and as $(v \rightarrow f)$ otherwise; this yields a directed graph $\mathcal{G}(\mathcal{B}, M)$.
2. Find all strongly connected components $S_1, S_2, \ldots, S_n$ of $\mathcal{G}(\mathcal{B}, M)$. Let $\mathcal{V}$ be the set of clusters $S_i \cup M(S_i)$ for $i \in \{1, \ldots, n\}$, where $M(S_i)$ denotes the set of vertices that are matched to vertices in S_i in matching M.
3. Let $\mathrm{cl}(f)$ denote the cluster in $\mathcal{V}$ containing f. For each $(v \rightarrow f)$ such that $v \notin \mathrm{cl}(f)$ add an edge $(v \rightarrow \mathrm{cl}(f))$ to $\mathcal{E}$.

Independent exogenous random variables and parameters are then added as singleton clusters with edges towards the clusters of the equations in which they appear. It has been shown that the resulting directed cluster graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E}\rangle$, which we refer to as the *causal ordering graph*, is independent of the choice of perfect matching [Theorem 4, Blom et al., 2021]. Example 1 shows how the algorithm works and a graphical illustration of the algorithm for a more elaborate cyclic model can be found in Appendix A.2.

Example 1. Let $V = \{v_1, v_2\}$, $W = \{w_1, w_2\}$, and $P = \{p_1, p_2\}$ be index sets. Consider model equations f_1 and f_2 with endogenous variables $(X_v)_{v\in V}$, exogenous random variables $(U_w)_{w\in W}$ and parameters C_p with $p \in P$ below.

$$f_1: \quad C_{p_1} X_{v_1} - U_{w_1} = 0, \tag{1}$$

$$f_2: \quad C_{p_2} X_{v_2} + X_{v_1} + U_{w_2} = 0. \tag{2}$$

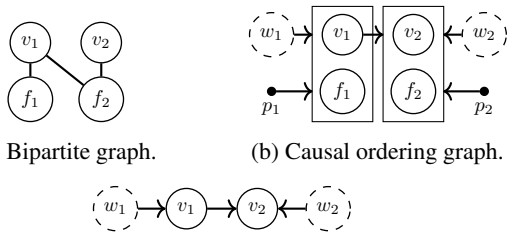
The bipartite graph $\mathcal{B} = \langle V, F, E\rangle$ in Figure 1a, with $E = \{(v_1 - f_1), (v_1 - f_2), (v_2 - f_2)\}$ is a compact representation of the model structure. This graph has a perfect matching $M = \{(v_1 - f_1), (v_2 - f_2)\}$. By orienting edges in $\mathcal{B}$ according to the rules in step 1 of the causal ordering algorithm we obtain the directed graph $\langle V \cup F, E_{\mathrm{dir}}\rangle$ with $E_{\mathrm{dir}} = \{(f_1 \rightarrow v_1), (f_2 \rightarrow v_2), (v_1 \rightarrow f_2)\}$. The clusters $C_1 = \{v_1, f_1\}$ and $C_2 = \{v_2, f_2\}$ are added to $\mathcal{V}$ in step 2 of the algorithm, and the edge $(v_1 \rightarrow C_2)$ is added to $\mathcal{E}$ in step 3. Finally, we add the parameters P and independent exogenous random variables W as singleton clusters to $\mathcal{V}$, and the edges $(p_1 \rightarrow C_1)$, $(w_1 \rightarrow C_1)$, $(p_2 \rightarrow C_2)$, and $(w_2 \rightarrow C_2)$ to $\mathcal{E}$. The resulting causal ordering graph is given in Figure 1b. △

Throughout this work, we will assume that models are uniquely solvable with respect to the causal ordering graph, which roughly means that for each cluster, the equations in that cluster can be solved uniquely for the endogenous variables in that cluster (see [Definition 14, Blom et al., 2021] for details). A perfect intervention on a cluster that contains equation vertices represents a model change where the equations in the targeted cluster are replaced by equations that set the endogenous variables in that cluster equal to constant values. A soft intervention targets an equation, parameter, or exogenous variable, but does not affect which

[1] Actually, we consider an equivalent algorithm for causal ordering that was shown to be more computationally efficient by [Nayak, 1995, Gonçalves and Porto, 2016]. For more details, see [Blom et al., 2021].

[2] A perfect matching M is a subset of edges in a bipartite graph so that every vertex is adjacent to exactly one edge in M. Note that not every bipartite graph has a perfect matching.

variables appear in the equations. We say that there is a directed path from a vertex x to a vertex y in a causal ordering graph $\langle \mathcal{V}, \mathcal{E} \rangle$ if either $\mathrm{cl}(x) = \mathrm{cl}(y)$ or there is a sequence of clusters $C_1 = \mathrm{cl}(x), C_2, \ldots, C_{k-1}, C_k = \mathrm{cl}(y)$ so that for all $i \in \{1, \ldots, k-1\}$ there is a vertex $z_i \in C_i$ such that $(z_i \to C_{i+1}) \in \mathcal{E}$. It can be shown that a) the presence of a directed path from a cluster, equation, parameter, or exogenous variable that is targeted by a soft intervention towards a certain variable in the causal ordering graph implies that the intervention has a generic effect on that variable and b) if no such path exists there is no causal effect of the intervention on that variable [Theorem 20, Blom et al., 2021].



(a) Bipartite graph. (b) Causal ordering graph.

(c) Markov ordering graph.

Figure 1: The bipartite graph in Figure 1a is a compact representation of the model in Example 1. The corresponding causal ordering graph and Markov ordering graph are given in Figures 1b and 1c respectively. Exogenous variables are denoted by dashed circles and parameters by black dots.

1.2 MARKOV ORDERING GRAPH

The causal ordering graph $\mathrm{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ of model equations F with endogenous variables V, exogenous random variables W, parameters P, and bipartite graph $\mathcal{B}$ can be used to construct the *Markov ordering graph*, which is a DAG $\mathrm{MO}(\mathcal{B}) = \langle V \cup W, E \rangle$, with $(x \to y) \in E$ if and only if $(x \to \mathrm{cl}(y)) \in \mathcal{E}$. The Markov ordering graph for the model equations in Example 1 is given in Figure 1c. It has been shown that, under the assumption of unique solvability w.r.t. the causal ordering graph, d-separations in the Markov ordering graph imply conditional independences between the corresponding variables [Blom et al., 2021]. Henceforth, we will assume that the probability distribution of the solution $(X_v)_{v \in V}$ to a set of model equations is faithful to the Markov ordering graph. In other words, each conditional independence in the distribution implies a d-separation in the Markov ordering graph. Under the assumption that data is generated from such a model, some causal discovery algorithms, such as the PC algorithm [Spirtes et al., 2000], aim to construct the Markov equivalence class of the Markov ordering graph. In this work, we will specifically focus on feedback models for which the Markov ordering graph of the equilibrium distribution, and consequently the output of many causal discovery algorithms, does not have a straightforward causal interpretation.

2 CAUSAL ORDERING FOR A VIRAL INFECTION MODEL

This work was inspired by a viral infection model discussed by De Boer [2012], who showed through explicit calculations that the predictions of the model are not robust under addition of an immune response. This sheds doubt on the correct interpretation of variables and parameters in the model. For many systems it is intrinsically difficult to study their behaviour in detail. The use of simplified mathematical models that capture key characteristics aids in the analysis of certain properties of the system. The hope is that the explanations inferred from model equations are legitimate accounts of the true underlying system [De Boer, 2012]. In reality, a modeller must take into account that the outcome of these studies may be contingent on the specifics of the model design. Here, we demonstrate how causal ordering can be used as a scalable tool to assess the robustness of model predictions without requiring explicit calculations.

2.1 VIRAL INFECTION WITHOUT IMMUNE RESPONSE

Let U_σ be a production term for target cells, d_T the death rate for target cells, U_f the fraction of successful infections, and U_δ the death rate of productively infected cells. Define $\beta = \frac{bp}{c}$, where b is the infection rate, p the amount of virus produced per infected cell, and c the clearance rate of viral particles. The following first-order differential equations describe how the amount of target cells $X_T(t)$ and the amount of infected cells $X_I(t)$ evolve over time [De Boer, 2012]:

$$\dot{X}_T(t) = U_\sigma - d_T X_T(t) - \beta X_T(t) X_I(t), \tag{3}$$

$$\dot{X}_I(t) = (U_f \beta X_T(t) - U_\delta) X_I(t), \tag{4}$$

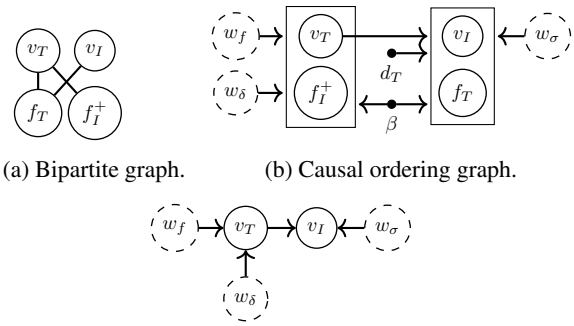
Suppose that we want to use this simple viral infection model to explain why the *set-point viral load* (i.e. the total amount of virus circulating in the bloodstream) of chronically infected HIV-patients differs by several orders of magnitude, as De Boer [2012] does. To analyse this problem we look at the equilibrium equations that are implied by equations (3) and (4):[3]

$$f_T: \qquad U_\sigma - d_T X_T - \beta X_T X_I = 0, \tag{5}$$

$$f_I^+: \qquad U_f \beta X_T - U_\delta = 0. \tag{6}$$

Throughout the remainder of this work we will use this *natural labelling* of equilibrium equations, where the equation derived from the derivative $\dot{X}_i(t)$ is labelled f_i. For first-order differential equations that are written in canonical form, $\dot{X}_i(t) = g_i(X(t))$, the natural labelling always exists.

[3] Since we are only interested in strictly positive solutions we removed X_I from the equilibrium equation $f_I : (U_f \beta X_T - U_\delta) X_I = 0$ to obtain f_I^+.



(a) Bipartite graph. (b) Causal ordering graph.

(c) Markov ordering graph.

Figure 2: Graphical representations of the viral infection model in equations (5) and (6). Vertices v_i and w_j correspond to variables X_i and U_j, respectively. The causal ordering graph represents generic effects of interventions. The d-separations in Figure 2c imply conditional independences.

Suppose that U_σ, U_f and U_δ are independent exogenous random variables taking values in $\mathbb{R}_{>0}$ and d_T, β are strictly positive parameters. The associated bipartite graph, causal ordering graph, and Markov ordering graph are given in Figure 2. The causal ordering graph tells us that soft interventions targeting U_σ, U_f, U_δ, d_T, or β generically have an effect on the equilibrium distribution of the amount of infected cells X_I. From here on, we say that the causal ordering graph of a model predicts the *generic presence* or *absence* of causal effects. The Markov ordering graph shows that v_T and w_σ are d-separated. This implies that the amount of target cells X_T should be independent of the production rate U_σ when the system is at equilibrium. Henceforth, we will say that the Markov ordering graph predicts the *generic presence* or *absence* of conditional dependences.

2.2 VIRAL INFECTION WITH A SINGLE IMMUNE RESPONSE

The viral infection model in equations (3) and (4) can be extended with a simple immune response $X_E(t)$ by adding the following dynamic and static equations:

$$\dot{X}_E(t) = (U_a X_I(t) - d_E) X_E(t), \tag{7}$$

$$X_\delta(t) = d_I + U_k X_E(t), \tag{8}$$

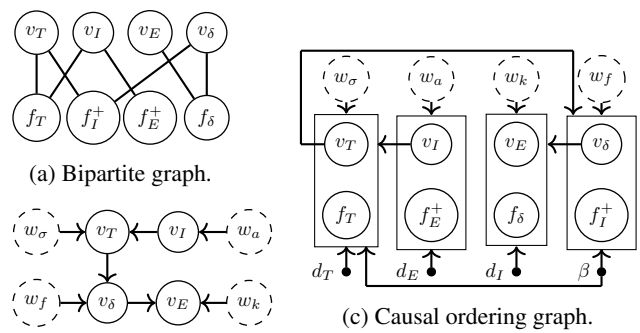
where U_a is an activation rate, d_E and d_I are turnover rates and U_k is a mass-action killing rate [De Boer, 2012]. Note that the exogenous random variable U_δ is now treated as an endogenous variable $X_\delta(t)$ instead. We derive the following equilibrium equations using the natural labelling provided by equations (7) and (8):[4]

$$f_E^+ : \qquad U_a X_I - d_E = 0, \tag{9}$$

$$f_\delta : \qquad X_\delta - d_I - U_k X_E = 0, \tag{10}$$

[4]Analogous to changing f_I to f_I^+ for strictly positive solutions, we will look at f_E^+ instead of f_E.

Henceforth, we will call the addition of equations F_+ to F a *model extension*. Notice that when two sets of equations are combined, there may exist variables that were exogenous in the submodel (i.e. the original model) but that are endogenous within the whole model (i.e. the extended model). Generally, equations F_+ may contain endogenous variables in V and exogenous variables in W but they may also contain additional endogenous variables V_+, additional exogenous variables W_+ and additional parameters P_+. Parameters and exogenous random variables that appear in equations F can appear as endogenous variables in V_+ and in the extended model $F_{\text{ext}} = F \cup F_+$. In that case, these variables are no longer considered to be parameters or exogenous variables within the extended model.



(a) Bipartite graph.

(b) Markov ordering graph.

(c) Causal ordering graph.

Figure 3: Graphical representations of the viral infection model with a single immune response. The presence or absence of causal relations and d-connections implied by the graphs in Figure 2 are not preserved if a single immune response is added.

Suppose that U_a and U_k are independent exogenous random variables taking values in $\mathbb{R}_{>0}$ and d_E, d_I are parameters taking value in $\mathbb{R}_{>0}$. The bipartite graph, causal ordering graph, and Markov ordering graph associated with equations (5), (6), (9), and (10) (with X_δ replacing U_δ) are given in Figure 3. The causal ordering graph predicts a causal effect of U_σ and d_T on X_T but not on X_I. By comparing with the predictions of the causal ordering graph in Figure 2b, we find that effects of interventions targeting U_σ and d_T are not robust under the model extension. The Markov ordering graph of the extended model shows that w_σ is d-connected to v_T, and hence U_σ and X_T are dependent. We conclude that the independence between U_σ and X_T that was implied by the Markov ordering graph of the viral infection model without immune response is not robust under the model extension.

The systematic graphical procedure followed here easily leads to the same causal conclusions as De Boer [2012] obtained by explicitly solving the equilibrium equations. In addition, it leads to predictions regarding the conditional (in)dependences in the equilibrium distribution.

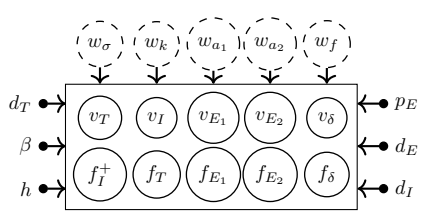
2.3 VIRAL INFECTION WITH MULTIPLE IMMUNE RESPONSES

The following static and dynamical equations describe multiple immune responses:

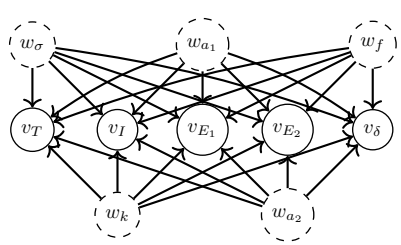
$$\dot{X}_{E_i}(t) = \frac{p_E X_{E_i}(t) U_{a_i} X_I(t)}{h + X_{E_i}(t) + U_{a_i} X_I(t)} - d_E X_{E_i}(t), \quad i = 1, 2, \ldots, n \tag{11}$$

$$X_\delta(t) = d_I + U_k \sum_i^n U_{a_i} X_{E_i}(t), \tag{12}$$

where there are n immune responses, U_{a_i} is the avidity of immune response i, p_E is the maximum division rate, and h is a saturation constant [De Boer, 2012]. For $n = 2$ we can derive equilibrium equations f_{E_1}, f_{E_2}, and f_δ using the natural labelling as we did for the equilibrium equations in the previous section. Together with the equilibrium equations (5) and (6) (with X_δ replacing U_δ) for the viral infection model this is another extended model. The bipartite graph of this extended model is given in Figure 5a, while the causal ordering graph can be found in Figure 4a. By comparing the directed paths in this causal ordering graph with that of the original viral infection model (i.e. the model without an immune response) in Figure 2b, it can be seen that the predicted presence of causal relations is preserved under extension of the model with multiple immune responses, while the predicted absence of causal relations is not. Similarly, by comparing d-separations in the Markov ordering graphs in Figure 2c with those in Figure 4b, we find that predicted conditional dependences are preserved under the extensions, while the predicted conditional independences are not.



(a) Causal ordering graph.



(b) Markov ordering graph.

Figure 4: Graphical representations of the viral infection model with multiple immune responses. The presence of causal relations and d-connections in Figure 2 is preserved.

2.4 MARKOV ORDERING GRAPHS AND CAUSAL INTERPRETATIONS

In [Blom et al., 2021], it was shown that the Markov ordering graph may not have a straightforward causal interpretation. Here, we illustrate for the viral infection models that the Markov ordering graphs neither have a straightforward causal interpretation at equilibrium in terms of soft interventions targeting parameters, exogenous variables, or equations *nor* in terms of perfect interventions on variables in the dynamical model. To see this, consider the Markov ordering graph in Figure 3b for the viral infection with a single immune response. The edge $(v_I \to v_T)$ cannot correspond to the effect of a soft intervention targeting f_I^+, because the causal ordering graph in Figure 3c shows that there is no such effect. Clearly, directed paths in the Markov ordering graph do not necessarily represent the effects of soft interventions. The natural way to model a perfect intervention targeting a variable in the Markov ordering graph is to replace the (differential) equation of that variable with an equation setting that variable equal to a certain value in the underlying dynamical model [Mooij et al., 2013]. By explicitly solving equilibrium equations it is easy to check that replacing f_δ with an equation setting X_δ equal to a constant generically changes the distribution of X_I. Since there is no directed path from v_δ to v_I in the Markov ordering graph, the effect of this perfect intervention would not have been predicted by the Markov ordering graph, if it would have been interpreted causally. Therefore, contrary to the causal ordering graph, the Markov ordering graph does not have a causal interpretation in terms of soft or perfect interventions on the true underlying dynamical model.

3 ROBUST CAUSAL PREDICTIONS UNDER MODEL EXTENSIONS

One way to gauge the robustness of model predictions is to check to what extent they depend on the model design. The example of a viral infection with different immune responses in the previous section indicates that qualitative causal predictions entailed by the causal ordering graph of a mathematical model may strongly depend on the particulars of the model. Both the implied presence or absence of causal relations at equilibrium and the implied presence or absence of conditional independences at equilibrium may change under certain model extensions. Under what conditions are these qualitative predictions preserved under model extensions? In this section, we characterize a large class of model extensions under which qualitative equilibrium predictions are preserved.

Theorem 1 gives a sufficient condition on model extensions under which the predicted generic presence of causal relations and predicted generic presence of conditional dependences at equilibrium is preserved. The proof is given in

Appendix A.2.

Theorem 1. *Consider model equations F containing endogenous variables V with bipartite graph $\mathcal{B}$. Suppose F is extended with equations F_+ containing endogenous variables in $V \cup V_+$, where V_+ contains endogenous variables that are added by the model extension.*[5] *Let $\mathcal{B}_{\text{ext}}$ be the bipartite graph associated with $F_{\text{ext}} = F \cup F_+$ and $V_{\text{ext}} = V \cup V_+$, and $\mathcal{B}_+$ the bipartite graph associated with the extension F_+ and V_+, where variables in V appearing in F^+ are treated as exogenous variables (i.e. they are not added as vertices in $\mathcal{B}_+$). If $\mathcal{B}$ and $\mathcal{B}_+$ both have a perfect matching then:*

1. *$\mathcal{B}_{\text{ext}}$ has a perfect matching,*
2. *ancestral relations in $\text{CO}(\mathcal{B})$ are also present in $\text{CO}(\mathcal{B}_{\text{ext}})$,*
3. *d-connections in $\text{MO}(\mathcal{B})$ are also present in $\text{MO}(\mathcal{B}_{\text{ext}})$.*

This result characterizes a large set of extensions under which the implied causal effects and conditional dependences of a model are preserved. Consider again the equilibrium behaviour of the viral infection models in Section 2. We already showed explicitly that the extension of the viral infection model with multiple immune responses preserved the predicted presence of causal relations and conditional dependences, but with the help of Theorem 1 we only would have needed to check whether the bipartite graph in Figure 5c has a perfect matching to arrive at the same conclusion. The bipartite graph for the extension with a single immune response in Figure 5b does not have a perfect matching and hence the conditions of Theorem 1 do not hold. Recall that this model extension did not preserve the predicted presence of causal relations.

The theorem below gives a stronger condition under which (conditional) independence relations and the absence of causal relations that are implied by a model are also predicted by the extended model. The proof is provided in the supplement.

Theorem 2. *Let F, F_+, F_{ext}, V, V_+, V_{ext}, $\mathcal{B}$, $\mathcal{B}_+$, and $\mathcal{B}_{\text{ext}}$ be as in Theorem 1. If $\mathcal{B}$ and $\mathcal{B}_+$ both have perfect matchings and no vertex in V_+ is adjacent to a vertex in F in $\mathcal{B}_{\text{ext}}$ then:*[6]

1. *ancestral relations absent in $\text{CO}(\mathcal{B})$ are also absent in $\text{CO}(\mathcal{B}_{\text{ext}})$,*
2. *d-connections absent in $\text{MO}(\mathcal{B})$ are also absent in $\text{MO}(\mathcal{B}_{\text{ext}})$.*

[5] V_+ may also contain parameters or exogenous variables that appear in F and become endogenous in the extended model.

[6] A vertex in V_+ is considered adjacent to F if it corresponds with one of the exogenous random variables or parameters in F that become endogenous in the model extension.

Together with Theorem 1, this result characterizes a large class of model extensions under which all qualitative model predictions are preserved. Consider again the equilibrium models for the viral infection in Section 2. The bipartite graph for the extension with a single immune response, which we obtain by adding equations (9) and (10), does not have a perfect matching. In the bipartite graph associated with the viral infection model with multiple immune responses the additional endogenous variable v_δ is adjacent to f_I. Neither of the model extensions satisfies the conditions of Theorem 2. We already demonstrated that neither of the model extensions preserves all qualitative model predictions. An example of a model extension that does satisfy the conditions in Theorem 1 and 2 is an acyclic structural causal model that is extended with another acyclic structural causal model such that the additional variables are non-ancestors of the original ones. Together, Theorem 1 and 2, can be used to understand when the causal and Markov properties of a system can be understood by studying the corresponding properties of its parts.

4 SELECTION OF MODEL EXTENSIONS

So far, we have considered methods to assess the robustness of qualitative model predictions. In this section we will show how this idea results in novel opportunities regarding causal discovery. In particular, if we assume that the systems that we observe are part of a larger partially observed system, then we can use the methods in this paper to reason about causal mechanisms of unobserved variables. Consider, for example, the viral infection model for which we have demonstrated that extensions with different immune responses imply different (conditional) independences between variables in the original model. The Markov ordering graphs in Figures 2c, 3b, and 4b imply the following (in)dependences:

1. Viral infection without immune response: $U_\sigma \perp\!\!\!\perp X_T$, $U_\sigma \not\perp\!\!\!\perp X_I$.
2. Viral infection with single immune response: $U_\sigma \not\perp\!\!\!\perp X_T$, $U_\sigma \perp\!\!\!\perp X_I$.
3. Viral infection with multiple immune responses: $U_\sigma \not\perp\!\!\!\perp X_T$, $U_\sigma \not\perp\!\!\!\perp X_I$.

Given a model for variables X_T and X_I only, we can reject model extensions based on the (conditional) independences for variables X_T, X_I, and U_σ. Using this holistic modelling approach, we can reason about an unknown model extension without observing the new mechanisms or variables. In the remainder of this section, we further discuss how this idea can be applied to equilibrium data of dynamical systems.

4.1 REASONING ABOUT SELF-REGULATING VARIABLES

We say that a variable in a set of first-order differential equations in canonical form is *self-regulating* if it can be solved uniquely from the equilibrium equation that is constructed from its derivative. For models in which every variable is self-regulating there exists a perfect matching where each variable v_i is matched to its associated equilibrium equation f_i according to the natural labelling, for more details see Lemma 1 in the supplement.[7] It then follows from Theorem 1 that the presence of ancestral relations and d-connections is robust under dynamical model extensions in which each variable is self-regulating, as is stated more formally in Corollary 1 below.

Corollary 1. *Consider a first-order dynamical model in canonical form for endogenous variables V and an extension consisting of canonical first-order differential equations for additional endogenous variables V_+. Let F and $F_{\text{ext}} = F \cup F_+$ be the equilibrium equations of the original and extended model respectively. If all variables in $V \cup V_+$ are self-regulating then 2 and 3 of Theorem 1 hold.*

Corollary 1 characterizes a class of models under which certain qualitative predictions for the equilibrium distribution are robust, but the result can also be interpreted from a different angle. Suppose that we have equilibrium data that is generated by an extended dynamical model with equilibrium equations F_{ext}, but we only have a *partial* model consisting of equations in F for a subset $V \subseteq V_{\text{ext}} = V \cup V_+$ of variables that appear in $F_{\text{ext}} = F \cup F_+$. If we would find conditional independences between variables in V that do not correspond to d-separations in the Markov ordering graph of the partial model, this does not necessarily mean that the model equations are wrong. It could also be the case, for example, that we are wrong to assume that the system can be studied in a reductionist manner and that the model should be extended. Furthermore, under the assumption that data is generated from the equilibrium distribution of a dynamical model, Corollary 1 tells us that conditional independences in the data that are not predicted by the equations of a partial model imply the presence of variables that are not self-regulating, if we assume faithfulness. This shows that, given a model for a subsystem, we can reason about the properties of unobserved and unknown variables in the whole system. Consider, for example, the model of the viral infection without immune response and assume that this is a submodel of a larger system. Suppose that we observe a conditional independence between U_σ and X_I and assume that the model equations of the submodel are correct. Since the Markov ordering graph in Figure 2c implies that U_σ and X_I are dependent, Corollary 1 tells us that there must be variables that are not self-regulating in the extended system. If the extended system can be described by the strictly positive solutions of the viral infection model with a single immune response, so that U_σ and X_I are independent, then we see from equations (5), (6), (9), and (10) that both $X_E(t)$ and $X_I(t)$ are not self-regulating.

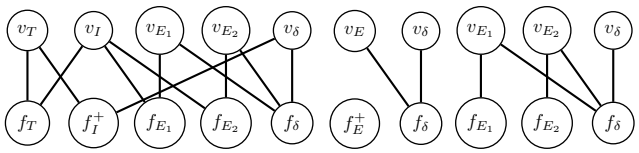
4.2 REASONING ABOUT FEEDBACK LOOPS

We say that an extension of a dynamical model *introduces a new feedback loop with the original dynamical model* when there is feedback in the extended dynamical model that involves variables in both the original model and the model extension. To make this definition more precise, consider the set E_{nat} of edges $(v_i - f_i)$ that are associated with the natural labelling of the equilibrium equations of the extended dynamical model. The feedback loops in the dynamical model coincide with cycles in the directed graph $\mathcal{G}(\mathcal{B}_{\text{nat}}, M_{\text{nat}})$ that is obtained by applying step 1 of the causal ordering algorithm to the bipartite graph $\mathcal{B}_{\text{nat}} = \langle V_{\text{ext}}, F_{\text{ext}}, E_{\text{ext}} \cup E_{\text{nat}} \rangle$ using the perfect matching $M_{\text{nat}} = E_{\text{nat}}$.[8] The following proposition can be used to reason about the presence of partially unobserved feedback loops given a model and observations for a subsystem.

Proposition 1. *Consider a first-order dynamical model in canonical form for endogenous variables V and an extension consisting of canonical first-order differential equations for additional endogenous variables V_+. Let F and $F_{\text{ext}} = F \cup F_+$ be the equilibrium equations of the original and extended model respectively. Let $\mathcal{B} = \langle V, F, E \rangle$ be the bipartite graph associated with F and $\mathcal{B}_{\text{ext}} = \langle V_{\text{ext}}, F_{\text{ext}}, E_{\text{ext}} \rangle$ the bipartite graph associated with F_{ext}. Assume that $\mathcal{B}$ and $\mathcal{B}_{\text{ext}}$ both have perfect matchings. If the model extension does not introduce a new feedback loop with the original dynamical model, then d-connections in $\text{MO}(\mathcal{B})$ are also present in $\text{MO}(\mathcal{B}_{\text{ext}})$.*

[7] Interestingly, the Markov ordering graph for the equilibrium equations of such a model always has a causal interpretation. By construction of the causal ordering graph from the bipartite graph and the perfect matching provided by the natural labelling, we know that a vertex v_i always appears in a cluster with f_i in the causal ordering graph. The presence or absence of directed paths in the Markov ordering graph can then easily be associated with the presence or absence of directed paths in the causal ordering graph. Consequently, the Markov ordering graph can be interpreted in terms of both soft interventions targeting equations and perfect interventions that set variables equal to a constant by replacement of the associated dynamical and equilibrium equations. Note that dynamical systems with only self-regulating variables were also considered in [Mooij et al., 2013], where it was shown that their equilibria can be modelled as structural causal models without self-cycles.

[8] Note that a feedback loop in the dynamical model does not imply a feedback loop in the equilibrium equations as well. For example, there is feedback in the dynamical equations (3), (4), but there is no feedback in the causal ordering graph of the equilibrium equations in Figure 2b nor in the directed graph that is constructed in step 1 of the causal ordering algorithm.



(a) Extended model. (b) Single. (c) Multiple.

Figure 5: The bipartite graphs associated with the viral infection model with multiple immune responses, the single immune response extension, and the multiple immune response extension are given in Figures 5a, 5b, and 5c, respectively.

Proposition 1 characterizes a class of model extensions under which certain qualitative model predictions are robust, but it also shows how we can reason about the existence of unobserved feedback loops. To be more precise, it shows that, given a submodel for a subsystem, the presence of conditional independences that are not predicted by the submodel imply the existence of an unobserved feedback loop, if we assume faithfulness. If, for example, we assume that the viral infection model without an immune response is a submodel of the system that is described by the strictly positive equilibrium solutions of the viral infection model with a single immune response, then we would observe an independence between U_σ and X_I that is not predicted by the model equations of the submodel. Proposition 1 would then imply that there is an unobserved feedback loop. Indeed, it can be seen from equations (3), (4), (7), (8) that there is an unobserved feedback loop from $X_I(t)$ to $X_E(t)$ to $X_\delta(t)$ and back to $X_I(t)$, while the Markov ordering graphs in Figures 2c and 3b imply that U_σ and X_I are dependent in the original model and independent in the extended model. We consider the use of existing structure learning algorithms for the detection of feedback loops in models with variables that are not self-regulating from a combination of background knowledge and observational equilibrium data to be an interesting topic for future work.

5 DISCUSSION

In this work we revisited several models of viral infections and immune responses. In our treatment of these models we closely followed the approach in De Boer [2012] and therefore we only considered strictly positive solutions. If we would have modelled all solutions then, for example, we would have considered the equilibrium equation $f_I : (U_f \beta X_T - U_\delta)X_I = 0$ instead of f_I^+ in equation (6). In that case, we would have obtained the causal ordering graph in Figure 6 instead of that in Figure 2b. Clearly, the model predictions of the causal ordering graph for the positive solutions in Figure 2b are more informative. The choice of only modelling strictly positive solutions depends on the application.

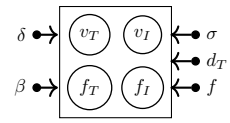


Figure 6: Causal ordering graph for positive and non-positive solutions of the viral infection model.

In many application domains mathematical models are used to predict the equilibrium behaviour of complex systems. An important issue is that (causal and Markov) predictions may strongly depend on the specifics of the model design. We revisited an example of a viral infection model [De Boer, 2012], in which implied causal relations and conditional independences change dramatically when equations, describing immune reactions, are added. Analysis of this behaviour through explicit calculations is neither insightful nor scalable. We showed how the technique of causal ordering can be used to efficiently analyse the robustness of implied causal effects and conditional independences under certain solvability assumptions. Using key insights provided by this approach we characterized large classes of model extensions under which predicted causal relations and conditional independences are robust. We hope that the results presented in this paper are a step towards bringing the world of causal modeling and reasoning closer to practical applications.

Our results for the characterization of the robustness of model extensions can also be used to reason about the properties of models that are the combination of two submodels. This way, we can study systems whose causal and Markov properties can be understood in a reductionistic manner by considering the properties of its parts. When the properties of the whole model differ from those of its parts, a holistic modelling approach is required. For models of the equilibrium distribution of dynamical systems, we proved that extensions of dynamical models where each variable is self-regulating preserve the predicted presence of causal effects and d-connections in the original model. Based on those insights, we proposed a novel approach to model selection, where information about conditional independences can be used in combination with model equations to reason about possible model extensions or the presence of feedback mechanisms. For dynamical models with feedback, the output of structure learning algorithms does not always have a causal interpretation in terms of soft or perfect interventions for the equilibrium distribution. We have shown that in dynamical systems where each variable is self-regulating the identifiable directed edges in the learned graph do express causal relations between variables.

References

Tineke Blom, Mirthe M. van Diepen, and Joris M. Mooij. Conditional independences and causal relations implied by sets of equations. *Journal of Machine Learning Research*, 22(178):1–62, 2021.

Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M. Mooij. Foundations of structural causal models with cycles and latent variables. *Annals of Statistics*, 49(5): 2885–2915, 2021.

D. Colombo, M. Maathuis, M. Kalisch, and Thomas S. Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *Annals of Statistics*, 40:294–321, 2012.

Rob J. De Boer. Which of our modeling predictions are robust? *PLOS Computational Biology*, 8(7), 2012.

Patrick Forré and Joris M. Mooij. Markov properties for graphical models with cycles and latent variables. *arXiv.org preprint*, arXiv:1710.08775 [math.ST], 2017. URL `https://arxiv.org/abs/1710.08775`.

Patrick Forré and Joris M. Mooij. Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. In *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*, 2018.

Bernardo Gonçalves and Fabio Porto. A note on the complexity of the causal ordering problem. *Artificial Intelligence*, 238:154–165, 2016.

Antti Hyttinen, Frederick Eberhard, and Patrik O. Hoyer. Learning linear cyclic causal models with latent variables. *The Journal of Machine Learning Research*, 13(1):3387–3439, 2012.

Gustavo Lacerda, Peter Spirtes, Joseph Ramsey, and Patrick O. Hoyer. Discovering cyclic causal models by independent components analysis. *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 1159–1168, 2008.

Joris M. Mooij and Tom Claassen. Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In Jonas Peters and David Sontag, editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI-20)*, volume 124, pages 1159–1168. PMLR, 8 2020.

Joris M. Mooij, Dominik Janzing, and Bernhard Schölkopf. From ordinary differential equations to structural causal models: the deterministic case. In *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*, pages 440–448, 2013.

P. Nayak. *Automated Modeling of Physical Systems*. Springer-Verlag Berlin Heidelberg, 1995.

Judea Pearl. *Causality : models, reasoning, and inference*. Cambridge University Press, 2009.

Alex Pothen and Chin-Ju Fan. Computing the block triangular form of a sparse matrix. *ACM Transactions on Mathematical Software (TOMS)*, 16(4):303–324, 1990.

Thomas Richardson and Peter Spirtes. Automated discovery of linear feedback models. *Computation, Causation, and Discovery*, pages 254–304, 1999.

Herbert A. Simon. On the definition of the causal relation. *The Journal of Philosophy*, 49(16):517–528, 1952.

Herbert A. Simon. Causal ordering and identifiability. In *Studies in Econometric Methods*, pages 49–74. John Wiley & Sons, 1953.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT press, 2000.

Erik V. Strobl. A constraint-based algorithm for causal discovery with cycles, latent variables and selection bias. *International Journal of Data Science and Analytics*, 8: 33–56, 2019.

J. Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16–17): 1873–1896, 2008.

A APPENDIX

A graphical illustration of the causal ordering algorithm applied to the equations of a cyclic model is provided in the first section. The second section contains proofs of the results in the main paper.

A.1 CAUSAL ORDERING ALGORITHM APPLIED TO A CYCLIC MODEL

In this section we demonstrate how the causal ordering algorithm works on a set of equations for a cyclic model. The algorithm is also presented graphically. Consider the following equations for endogenous variables $\boldsymbol{X}$ and exogenous random variables $\boldsymbol{U}$:

$$f_1: \quad g_1(X_{v_1}, U_{w_1}) = 0, \tag{13}$$
$$f_2: \quad g_2(X_{v_2}, X_{v_1}, X_{v_4}, U_{w_2}) = 0, \tag{14}$$
$$f_3: \quad g_3(X_{v_3}, X_{v_2}, U_{w_3}) = 0, \tag{15}$$
$$f_4: \quad g_4(X_{v_4}, X_{v_3}, U_{w_4}) = 0, \tag{16}$$
$$f_5: \quad g_5(X_{v_5}, X_{v_4}, U_{w_5}) = 0. \tag{17}$$

The associated bipartite graph in Figure 7a consists of variable vertices $V = \{v_1, \dots, v_5\}$ and equation vertices $F = \{f_1, \dots, f_5\}$. There is an edge between a variable vertex and an equation vertex whenever that variable appears in the equation. The associated bipartite graph has exactly two perfect matchings:

$$M_1 = \{(v_1 - f_1), (v_2 - f_2), (v_3 - f_3), (v_4 - f_4), (v_5 - f_5)\},$$
$$M_2 = \{(v_1 - f_1), (v_2 - f_3), (v_3 - f_4), (v_4 - f_2), (v_5 - f_5)\}.$$

Application of the first step of the causal ordering algorithm results either in the directed graph in Figure 7b or that in Figure 7c, depending on the choice of the perfect matching. The segmentation of vertices into strongly connected components, which takes place in the second step of the algorithm, results in the clusters $\{v_1\}$, $\{f_1\}$, $\{v_2, v_3, v_4, f_2, f_3, f_4\}$, $\{v_5\}$, and $\{f_5\}$. To construct the clusters of the causal ordering graph we add $S_i \cup M(S_i)$ to a cluster set $\mathcal{V}$ for each S_i in the segmentation. The segmentation of vertices into strongly connected components is displayed in Figures 7d and 7e. Notice that the segmentation in Figure 7d is the same as that in Figure 7e. It is known that the segmentation into strongly connected components is unique (i.e. it does not depend on the choice of the perfect matching) [Pothen and Fan, 1990, Blom et al., 2021]. The cluster set $\mathcal{V}$ for the causal ordering graph in Figure 7f is constructed by merging clusters in the segmented graph whenever two clusters contain vertices that are matched and by adding exogenous variables as singleton clusters. The edge set $\mathcal{E}$ for the causal ordering graph is obtained by adding edges $(v \to C)$ from an endogenous vertex v to a cluster C, whenever $v \notin C$ and there is an edge from v to $f \in C$ in the directed graph. We also add edges from exogenous vertices to clusters that contain equations in which the corresponding exogenous random variables appear.

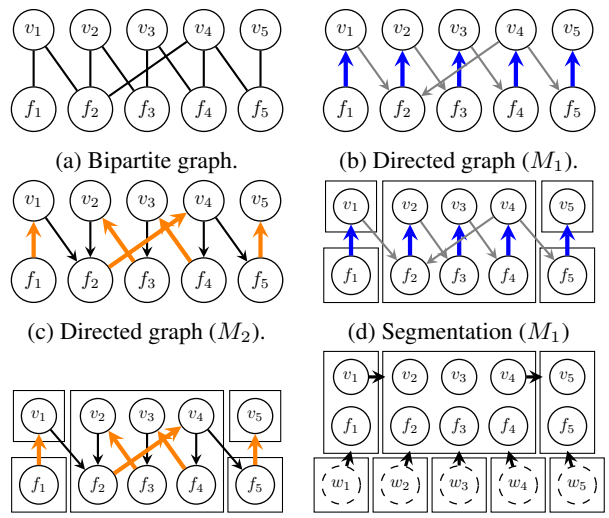


Figure 7: Graphical illustration of the causal ordering algorithm that was described in Section 1.1. Figure 7a shows the bipartite graph that is associated with equations (13) to (17). Application of the first step of the causal ordering algorithm results in the directed graph in Figure 7b for perfect matching M_1 and that in Figure 7c for perfect matching M_2. The blue and orange edges correspond to the edges in the perfect matchings M_1 and M_2, respectively. Figures 7d and 7e show that the segmentation into strongly connected components does not depend on the choice of the perfect matching. Exogenous vertices and edges from these vertices to clusters were added to the causal ordering graph in Figure 7f.

A.2 PROOFS

Theorem 1. *Consider model equations F containing endogenous variables V with bipartite graph $\mathcal{B}$. Suppose F is extended with equations F_+ containing endogenous variables in $V \cup V_+$, where V_+ contains endogenous variables that are added by the model extension.*[9] *Let $\mathcal{B}_{\text{ext}}$ be the bipartite graph associated with $F_{\text{ext}} = F \cup F_+$ and $V_{\text{ext}} = V \cup V_+$, and $\mathcal{B}_+$ the bipartite graph associated with the extension F_+ and V_+, where variables in V appearing in F^+ are treated as exogenous variables (i.e. they are not added as vertices in $\mathcal{B}_+$). If $\mathcal{B}$ and $\mathcal{B}_+$ both have a perfect matching then:*

1. *$\mathcal{B}_{\text{ext}}$ has a perfect matching,*

[9] V_+ may also contain parameters or exogenous variables that appear in F and become endogenous in the extended model.

2. *ancestral relations in* $\mathrm{CO}(\mathcal{B})$ *are also present in* $\mathrm{CO}(\mathcal{B}_{\mathrm{ext}})$*,*
3. *d-connections in* $\mathrm{MO}(\mathcal{B})$ *are also present in* $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$*.*

Proof. The causal ordering graph $\mathrm{CO}(\mathcal{B})$ is constructed from a perfect matching M for the bipartite graph $\mathcal{B} = \langle V, F, E\rangle$. Let M_+ be a perfect matching for $\mathcal{B}_+$. Note that $M_{\mathrm{ext}} = M \cup M_+$ is a perfect matching for $\mathcal{B}_{\mathrm{ext}} = \langle V \cup V_+, F \cup F_+, E_{\mathrm{ext}}\rangle$. Following the causal ordering algorithm for $\mathcal{B}, M$ and $\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}}$, we note that $\mathcal{G}(\mathcal{B}, M)$ is a subgraph of $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$ and hence clusters in $\mathrm{CO}(\mathcal{B})$ are fully contained in clusters in $\mathrm{CO}(\mathcal{B}_{\mathrm{ext}})$. Therefore ancestral relations in $\mathrm{CO}(\mathcal{B})$ are also present in $\mathrm{CO}(\mathcal{B}_{\mathrm{ext}})$.

It follows directly from the definition [Forré and Mooij, 2017] that σ-connections in a graph remain present if the graph is extended with additional vertices and edges. The directed graphs $\mathcal{G}(\mathcal{B}, M)$ and $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$ can be augmented with exogenous variables by adding exogenous vertices to these graphs with directed edges towards the equations in which they appear. The σ-connections in the augmentation of $\mathcal{G}(\mathcal{B}, M)$ must also be present in the augmentation of $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$. By [Corollary 2.8.4, Forré and Mooij, 2017] and [Lemma 43, Blom et al., 2021] we have that d-connections in $\mathrm{MO}(\mathcal{B})$ must also be present in $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$. □

Theorem 2. *Let* F*,* F_+*,* F_{ext}*,* V*,* V_+*,* V_{ext}*,* $\mathcal{B}$*,* $\mathcal{B}_+$*, and* $\mathcal{B}_{\mathrm{ext}}$ *be as in Theorem 1. If* $\mathcal{B}$ *and* $\mathcal{B}_+$ *both have perfect matchings and no vertex in* V_+ *is adjacent to a vertex in* F *in* $\mathcal{B}_{\mathrm{ext}}$ *then:*[10]

1. *ancestral relations absent in* $\mathrm{CO}(\mathcal{B})$ *are also absent in* $\mathrm{CO}(\mathcal{B}_{\mathrm{ext}})$*,*
2. *d-connections absent in* $\mathrm{MO}(\mathcal{B})$ *are also absent in* $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$*.*

Proof. Since $\mathcal{B}$ and $\mathcal{B}_+$ both have perfect matchings the results of Theorem 1 hold. Let $\mathcal{G}(\mathcal{B}, M)$, and $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$ be as in the proof of Theorem 1. Note that in M_{ext} vertices in F_+ are matched to vertices in V_+ and therefore edges between $f_+ \in F_+$ and $v \in \mathrm{adj}_{\mathcal{B}_{\mathrm{ext}}}(F_+) \setminus V_+$ are oriented as $(f_+ \leftarrow v)$ in $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$. By assumption, we therefore have that vertices in V_+ are non-ancestors of vertices in $V \cup F$ in $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$. Since $M \subseteq M_{\mathrm{ext}}$ we know that the same directed edges between vertices in V and F appear in both $\mathcal{G}(\mathcal{B}, M)$ and $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$. Notice that the subgraph of $\mathcal{G}(\mathcal{B}_{\mathrm{ext}}, M_{\mathrm{ext}})$ induced by the vertices $V \cup F$ coincides with $\mathcal{G}(\mathcal{B}, M)$. Hence $\mathrm{CO}(\mathcal{B})$ is the induced subgraph of $\mathrm{CO}(\mathcal{B}_{\mathrm{ext}})$ and $\mathrm{MO}(\mathcal{B})$ is the induced subgraph of $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$. □

[10] A vertex in V_+ is considered adjacent to F if it corresponds with one of the exogenous random variables or parameters in F that become endogenous in the model extension.

Lemma 1. *Consider a first-order dynamical model in canonical form for endogenous variables* V *and let* F *be the equilibrium equations of the model. If all variables in* V *are self-regulating then* $\mathcal{B}$ *has a perfect matching.*

Proof. Recall that the equilibrium equation constructed from the derivative of a variable i is labelled f_i according to the natural labelling. When a variable in $v_i \in V$ is self-regulating then it can be matched to its equilibrium equation f_i. If this holds for all variables in V then $\mathcal{B}$ has a perfect matching. □

Lemma 2. *Let* $\mathcal{B}$ *be a bipartite graph and let* M *and* M' *be two distinct perfect matchings. The associated directed graphs* $\mathcal{G}(\mathcal{B}, M)$ *and* $\mathcal{G}(\mathcal{B}, M')$ *that are obtained in step 1 of the causal ordering algorithm differ only in the direction of cycles.*

Proof. This follows directly from the fact that the output of the causal ordering algorithm does not depend on the choice of the perfect matching. This result is a direct consequence of Theorem 4 and Theorem 6 in Blom et al. [2021]. □

Proposition 1. *Consider a first-order dynamical model in canonical form for endogenous variables* V *and an extension consisting of canonical first-order differential equations for additional endogenous variables* V_+*. Let* F *and* $F_{\mathrm{ext}} = F \cup F_+$ *be the equilibrium equations of the original and extended model respectively. Let* $\mathcal{B} = \langle V, F, E\rangle$ *be the bipartite graph associated with* F *and* $\mathcal{B}_{\mathrm{ext}} = \langle V_{\mathrm{ext}}, F_{\mathrm{ext}}, E_{\mathrm{ext}}\rangle$ *the bipartite graph associated with* F_{ext}*. Assume that* $\mathcal{B}$ *and* $\mathcal{B}_{\mathrm{ext}}$ *both have perfect matchings. If the model extension does not introduce a new feedback loop with the original dynamical model, then d-connections in* $\mathrm{MO}(\mathcal{B})$ *are also present in* $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$*.*

Proof. Let E_{nat} be the set of edges $(v_i - f_i)$ associated with the natural labelling of the equilibrium equations of the extended dynamical model. Note that the feedback loops in the dynamical model coincide with cycles in the directed graph $\mathcal{G}(\mathcal{B}_{\mathrm{nat}}, M_{\mathrm{nat}})$ that is obtained by applying step 1 of the causal ordering algorithm to the bipartite graph $\mathcal{B}_{\mathrm{nat}} = \langle V_{\mathrm{ext}}, F_{\mathrm{ext}}, E_{\mathrm{ext}} \cup E_{\mathrm{nat}}\rangle$ using the perfect matching $M_{\mathrm{nat}} = E_{\mathrm{nat}}$.

By Theorem 1, we know that if $\mathcal{B}$ and $\mathcal{B}_+$ (the subgraph of $\mathcal{B}_{\mathrm{ext}}$ induced by $V_+ \cup F_+$) both have perfect matchings then d-connections in $\mathrm{MO}(\mathcal{B})$ must also be present in $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$. Therefore, if there exists a perfect matching M_{ext} for $\mathcal{B}_{\mathrm{ext}}$ so that each $f \in F$ is M_{ext}-matched to a vertex $v \in V$ and each $f_+ \in F_+$ is M_{ext}-matched to a vertex $v_+ \in V_+$ in $\mathcal{B}_{\mathrm{ext}}$, d-connections in $\mathrm{MO}(\mathcal{B})$ are also present in $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$.

We will prove the contrapositive of the proposition, so we start with the assumption that the d-connections in $\mathrm{MO}(\mathcal{B})$ are not preserved in $\mathrm{MO}(\mathcal{B}_{\mathrm{ext}})$. In that case, there must

exist a perfect matching M_{ext} for $\mathcal{B}_{\mathrm{ext}}$ so that there is an $f \in F$ that is M_{ext}-matched to a $v_+ \in V_+$ and a $v \in V$ that is M_{ext}-matched to a $f_+ \in F_+$. Note that since $\mathcal{B}_{\mathrm{ext}}$ is a subgraph of $\mathcal{B}_{\mathrm{nat}}$, this perfect matching M_{ext} is also a perfect matching for $\mathcal{B}_{\mathrm{nat}}$. Lemma 2 says that $\mathcal{G}(\mathcal{B}_{\mathrm{nat}}, M_{\mathrm{nat}})$ and $\mathcal{G}(\mathcal{B}_{\mathrm{nat}}, M_{\mathrm{ext}})$ only differ in the direction of cycles. We know that vertices in V are only M_{nat}-matched to vertices in F, while vertices in V_+ are only M_{nat}-matched to vertices in F_+. Therefore, the vertices v_+ and f must be on a directed cycle in both directed graphs, as well as v and f_+. Hence the model extension F_+ introduced a new feedback loop that includes variables in the original model. $\square$