



## UvA-DARE (Digital Academic Repository)

### Causality and independence in perfectly adapted dynamical systems

Blom, T.; Mooij, J.M.

**DOI**

[10.1515/jci-2021-0005](https://doi.org/10.1515/jci-2021-0005)

**Publication date**

2023

**Document Version**

Final published version

**Published in**

Journal of Causal Inference

**License**

CC BY

[Link to publication](#)

**Citation for published version (APA):**

Blom, T., & Mooij, J. M. (2023). Causality and independence in perfectly adapted dynamical systems. *Journal of Causal Inference*, 11(1), [20210005]. <https://doi.org/10.1515/jci-2021-0005>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## Research Article

Tineke Blom and Joris M. Mooij\*

# Causality and independence in perfectly adapted dynamical systems

<https://doi.org/10.1515/jci-2021-0005>

received February 09, 2021; accepted November 29, 2022

**Abstract:** Perfect adaptation in a dynamical system is the phenomenon that one or more variables have an initial transient response to a persistent change in an external stimulus but revert to their original value as the system converges to equilibrium. With the help of the causal ordering algorithm, one can construct graphical representations of dynamical systems that represent the causal relations between the variables and the conditional independences in the equilibrium distribution. We apply these tools to formulate sufficient graphical conditions for identifying perfect adaptation from a set of first-order differential equations. Furthermore, we give sufficient conditions to test for the presence of perfect adaptation in experimental equilibrium data. We apply this method to a simple model for a protein signalling pathway and test its predictions in both simulations and using real-world protein expression data. We demonstrate that perfect adaptation can lead to misleading orientation of edges in the output of causal discovery algorithms.

**Keywords:** causal ordering, causal discovery, feedback, equilibrium, perfect adaptation

**MSC 2020:** 37C20, 60G99, 62A01, 62A09, 62D20, 62H22, 62M99, 92C42, 92C45, 93C73

## 1 Introduction

Understanding causal relations is an objective that is central to many scientific endeavours. It is often said that the gold standard for causal discovery is a randomized controlled trial, but practical experiments can be too expensive, unethical, or otherwise infeasible. The promise of causal discovery is that we can, under certain assumptions, learn about causal relations by using a combination of data and background knowledge [1–10]. Roughly speaking, causal discovery algorithms construct a graphical representation that encodes certain aspects of the data, such as conditional independences, given some constraints that are imposed by background knowledge. Under additional assumptions on the underlying causal mechanisms, these graphical representations have a causal interpretation as well. In this work, we specifically consider the equilibrium distribution of *perfectly adapted dynamical systems*. We will show that such systems may have the property that the corresponding graphs that encode the conditional independences in the distribution do not have a straightforward causal interpretation in terms of the changes in distribution induced by interventions.

*Perfect adaptation* in a dynamical system is the phenomenon that one or more variables in the system temporarily respond to a persistent change of an external stimulus, but ultimately revert to their original values as the system reaches equilibrium again. We study the differences between the causal structure

---

\* **Corresponding author: Joris M. Mooij**, Korteweg-de Vries Institute, University of Amsterdam, Amsterdam, The Netherlands, e-mail: [j.m.mooij@uva.nl](mailto:j.m.mooij@uva.nl)

**Tineke Blom:** Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands, e-mail: [t.blom2@uva.nl](mailto:t.blom2@uva.nl)

implied by the dynamic equations and the conditional dependence structure of the equilibrium distribution. To do so, we make use of the technique of *causal ordering*, introduced by Simon [11], which can be used to construct a *causal ordering graph* that represents causal relations and a *Markov ordering graph* that encodes conditional independences between variables [12]. We introduce the notion of a *dynamic causal ordering graph* to represent transient causal effects in a dynamical model. We use these graphs to provide a sufficient graphical condition for a dynamical system to achieve perfect adaptation, which does not require simulations or cumbersome calculations. Furthermore, we provide sufficient conditions to test for the presence of perfect adaptation in real-world data with the help of the equilibrium Markov ordering graph, and we explain why the usual interpretation of causal discovery algorithms, when applied to perfectly adapted dynamical systems at equilibrium, can be misleading.

We illustrate our ideas on three simple dynamical systems with feedback: a filling bathtub model [13,14], a viral infection model [15,16], and a chemical reaction network [17]. We point out how perfect adaptation may also manifest itself in protein signalling networks, and take a closer look at the consequences for causal discovery from a popular protein expression data set [18]. We adapt a model for the RAS-RAF-MEK-ERK signalling pathway from [19] and study its properties under certain saturation conditions. We test its predictions both in simulations and on real-world data. We propose that the phenomenon of perfect adaptation might explain why the presence and orientation of edges in the output of causal discovery algorithms does not always agree with the direction of edges in biological consensus networks that are based on a partial representation of the underlying dynamical mechanisms.

## 2 Background

In this section, we provide an overview of the relevant background material on which this work is based. We first consider the assumptions underpinning popular constraint-based causal discovery algorithms and give a brief description of a simple causal discovery algorithm in Section 2.1. In Section 2.2, we proceed with a concise introduction to the causal ordering algorithm of Simon [11] and demonstrate how it can be applied to a set of equations together with a pre-specified set of exogenous variables to deduce the implied causal relations and conditional independences. Finally, in Section 2.3, we discuss the relationship with existing work.

### 2.1 Causal discovery

The main objective of causal discovery is to infer causal relations from experimental and observational data. The most common causal discovery algorithms can be roughly divided into score-based and constraint-based approaches. In this work, we will focus on the latter approach (examples include PC, FCI, and variants thereof, and see [1,2,4,5,7,9,10]), which exploits conditional independences in data to infer causal relations. We will first discuss assumptions for constraint-based causal discovery. We then consider the application of causal discovery algorithms to models with feedback. We proceed with a brief but concrete introduction to a simple causal discovery algorithm. Finally, we discuss how the present work relates to the existing literature.

Learning a graphical structure from conditional independence constraints typically relies on Markov and faithfulness assumptions relating conditional independences to properties of a graph. In particular, a *d-separation* is a relation between three disjoint sets of vertices in a graph that indicates whether all paths between two sets of vertices are *d*-blocked by the vertices in a third [3,4]. If every *d*-separation in a graph implies a conditional independence in the probability distribution, then we say that the distribution satisfies the *directed global Markov property* (or the *d-separation criterion*) with respect to that graph [20]. Conversely, if every conditional independence in the probability distribution corresponds to a

$d$ -separation in a graph, then we say that the distribution is  $d$ -faithful to that graph.<sup>1</sup> When a probability distribution satisfies the  $d$ -separation Markov property with respect to a graph and is also  $d$ -faithful to the graph, then this graph is a compact representation of the conditional independences in the probability distribution, and we say that it *encodes* its independence relations. A lot of work has been done to understand the (combinations of) various assumptions (e.g., linearity, Gaussianity, discreteness, causal sufficiency, acyclicity, no selection bias) under which a graph that encodes all conditional independences and dependences in a probability distribution has a certain causal interpretation [see, e.g., 2,4–10,21,22]. Perhaps the simplest assumption is that the data were generated by a causal directed acyclic graph (DAG) [3,4].<sup>2</sup> This graphical object represents both the conditional independence structure (the observational probability distribution is assumed to satisfy the directed global Markov property with respect to the graph) and the causal structure (every directed edge that is absent in the DAG corresponds with the absence of a direct causal relation between the two variables, relative to the set of variables in the DAG). For this acyclic setting, sophisticated constraint-based causal discovery algorithms (such as the PC and FCI algorithms [4,5]) have been developed. The key assumption that these algorithms make is that the same DAG expresses both the conditional independences in the observational distribution and the causal relations between the variables.

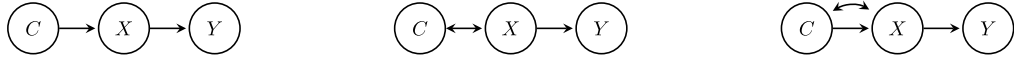
However, many systems of interest in various scientific disciplines (e.g., biology, econometrics, physics) include feedback mechanisms. Cyclic structural causal models (SCMs) can be used to model causal features and conditional independence relations of systems that contain cyclic causal relationships [23]. For linear SCMs with causal cycles, several causal discovery algorithms have been developed that are based on  $d$ -separations [2,6,21,22]. The  $d$ -separation criterion is applicable to acyclic settings and to cyclic SCMs with either discrete variables or linear relations between continuous variables, but it is too strong in general [24]. A weaker Markov property, based on the notion of  $\sigma$ -separation, has been derived for graphs that may contain cycles [23–25]. If every  $\sigma$ -separation in a graph implies a conditional independence in the probability distribution then we say that it satisfies the *generalized directed global Markov property* (or the  $\sigma$ -separation criterion) with respect to that graph [25]. Conversely, if every conditional independence in the probability distribution corresponds to a  $\sigma$ -separation in a graph then we say that it is  $\sigma$ -faithful to that graph. Richardson [26] already proposed a causal discovery algorithm that is sound under the generalized directed Markov property and the  $d$ -faithfulness assumption, assuming causal sufficiency. Recently, a causal discovery algorithm was proposed based on the  $\sigma$ -separation criterion and the assumption of  $\sigma$ -faithfulness that is sound and complete [8]. It was also shown that, perhaps surprisingly, the PC and FCI algorithms are still sound and complete in that setting, although their output has to be interpreted differently [10].

For the sake of simplicity, we will limit our attention in this article to one of the simplest causal discovery algorithms, local causal discovery (LCD) [1]. This algorithm is a straightforward and efficient search method to detect a specific causal structure from background knowledge and observational or experimental data. The algorithm looks for triples of variables  $(C, X, Y)$  for which (a)  $C$  is a variable that is not caused by  $X$ , and (b) the following (in)dependencies hold:  $C \perp\!\!\!\perp X$ ,  $X \perp\!\!\!\perp Y$ , and  $C \perp\!\!\!\perp Y \mid X$ . Figure 1 shows all acyclic directed mixed graphs (ADMGs)<sup>3</sup> that correspond to the LCD triple  $(C, X, Y)$ , assuming  $d$ -faithfulness and no selection bias. They all have in common that there is no bidirected edge between  $X$  and  $Y$ , while there is a directed edge from  $X$  to  $Y$ . Hence, we can conclude that  $X$  causes  $Y$  and the two variables are not confounded. The algorithm was proven to be sound in both the  $\sigma$ - and  $d$ -separation settings even when cycles may be present [9]. Even though this algorithm is not as sophisticated as many other causal

<sup>1</sup> While most authors refer to the notion of “ $d$ -faithful” simply as “faithful,” we distinguish between two different notions, “ $d$ -faithful” and “ $\sigma$ -faithful.”

<sup>2</sup> A DAG is a pair  $(V, E)$ , where  $V$  is a set of vertices and  $E \subseteq \{i \rightarrow j : i, j \in V\}$  is a set of directed edges between vertices such that there are no directed cycles.

<sup>3</sup> An ADMG is a triple  $(V, E, F)$ , where  $V$  is a set of vertices,  $E \subseteq \{i \rightarrow j : i, j \in V\}$  is a set of directed edges between vertices such that there are no directed cycles, and  $F \subseteq \{i \leftrightarrow j : i, j \in V : i \neq j\}$  is a set of bidirected edges between vertices.



**Figure 1:** All ADMGs that form an LCD triple  $(C, X, Y)$ , assuming  $d$ -faithfulness. In the absence of latent confounders and selection bias, the structure can only be that of the DAG on the left.

discovery algorithms, it suffices for our exposition of the pitfalls of attempting causal discovery on data generated by a perfectly adaptive dynamical system.

In this article, we consider equilibrium distributions that are generated by dynamical models. The causal relations in an equilibrium model are defined through the effects of persistent interventions (i.e., interventions that are constant over time) on the equilibrium distribution, assuming that the system again converges to equilibrium. It has been shown that directed graphs encoding the conditional independences between endogenous variables in the equilibrium distribution of dynamical systems with feedback do not always have a straightforward and intuitive causal interpretation [12–14] (see also Appendix F). As a consequence, the output of causal discovery algorithms applied to equilibrium data of dynamical systems with feedback cannot always be interpreted causally in a naive way. In this work, we will show that this not only happens in isolated cases, but that this is actually a common phenomenon in a large class of models with perfectly adapting feedback mechanisms. In our opinion, a better understanding of how perfectly adapting feedback loops affect the causal interpretation of the conditional independence structure is a prerequisite for successful applications of causal discovery in fields like systems biology, where one often encounters perfect adaptivity. One way to arrive at such understanding is through the application of the causal ordering algorithm, the topic of the next section.

## 2.2 Causal ordering

The causal ordering algorithm of Simon [11] returns an ordering of endogenous variables that occur in a set of equations, given a specification of which variables are exogenous. The algorithm was recently reformulated so that the output is a causal ordering graph that encodes the generic effects of certain interventions and a Markov ordering graph that represents conditional independences (both under certain assumptions regarding the solvability of the equations) [12]. We refer readers who are not yet familiar with the causal ordering algorithm to ref. [12] for a more extensive introduction to this approach. Here, we will provide only a brief description of the algorithm and discuss how its output should be interpreted.

First, note that the structure of a set of equations and the variables that appear in them can be represented by a bipartite graph  $\mathcal{B} = \langle V, F, E \rangle$ , where vertices  $F$  correspond to the equations and vertices  $V$  correspond to the endogenous variables that appear in these equations. For each endogenous variable  $v \in V$  that appears in an equation  $f \in F$ , there is an undirected edge  $(v - f) \in E$ . The output of the causal ordering algorithm is a directed cluster graph  $\langle \mathcal{V}, \mathcal{E} \rangle$ , consisting of a partition  $\mathcal{V}$  of the vertices  $V \cup F$  into clusters, and edges  $(v \rightarrow S) \in \mathcal{E}$  that go from vertices  $v \in V$  to clusters  $S \in \mathcal{V}$ . Application of the causal ordering algorithm to a bipartite graph  $\mathcal{B} = \langle V, F, E \rangle$  results in the directed cluster graph  $\text{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ , which we will call the *causal ordering graph* [12]. For simplicity, we assume here that the bipartite graph has a *perfect matching* (i.e., there exists a subset  $M \subseteq E$  of the edges in the bipartite graph  $\mathcal{B} = \langle V, F, E \rangle$  so that every vertex in  $V \cup F$  is adjacent to exactly one edge in  $M$ ).<sup>4</sup> The causal ordering graph is constructed by the following steps:<sup>5</sup>

<sup>4</sup> Although the causal ordering algorithm has been extended to general bipartite graphs that may not have a perfect matching [12], we will always assume in this work that a perfect matching exists to keep the exposition simple.

<sup>5</sup> Here, we actually give the reformulation of the causal ordering algorithm by Nayak [27] based on the block triangular form of matrices in ref. [28]. It was shown that the output of this algorithm is equivalent to that of the original causal ordering algorithm [12], but computationally more efficient than the original causal ordering algorithm [29].

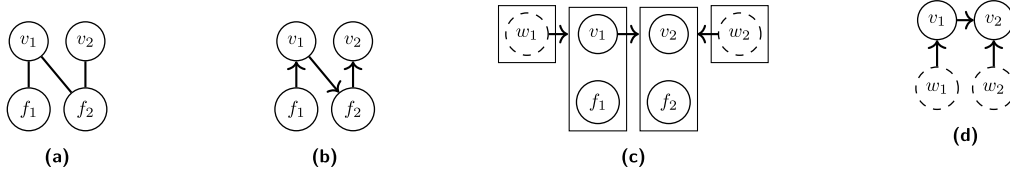
- (1) Find a perfect matching  $M \subseteq E$  and let  $M(S)$  denote the vertices in  $V \cup F$  that are joined to vertices in  $S \subseteq V \cup F$  by an edge in  $M$ .
- (2) For each  $(v - f) \in E$  with  $v \in V$  and  $f \in F$ : if  $(v - f) \in M$  orient the edge as  $(v \leftarrow f)$  and if  $(v - f) \notin M$  orient the edge as  $(v \rightarrow f)$ . Let  $\mathcal{G}(\mathcal{B}, M)$  denote the resulting directed bipartite graph.
- (3) Partition vertices  $V \cup F$  into strongly connected components  $\mathcal{V}'$  of  $\mathcal{G}(\mathcal{B}, M)$ . Create the cluster set  $\mathcal{V}$  consisting of clusters  $S \cup M(S)$  for each  $S \in \mathcal{V}'$ . For each edge  $(v \rightarrow f)$  in  $\mathcal{G}(\mathcal{B}, M)$ , add an edge  $(v \rightarrow \text{cl}(f))$  to  $\mathcal{E}$  when  $v \notin \text{cl}(f)$ , where  $\text{cl}(f)$  denotes the cluster in  $\mathcal{V}$  that contains  $f$ .
- (4) Exogenous variables appearing in the equations are added as singleton clusters to  $\mathcal{V}$ , with directed edges towards the clusters of the equations in which they appear in  $\mathcal{E}$ .
- (5) Output the directed cluster graph  $\text{CO}(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$ .

**Example 1.** Consider the following set of equations with index set  $F = \{f_1, f_2\}$  that contain endogenous variables with index set  $V = \{v_1, v_2\}$ :

$$f_1 : X_{v_1} - U_{w_1} = 0, \quad (1)$$

$$f_2 : X_{v_2} + X_{v_1} - U_{w_2} = 0, \quad (2)$$

where  $U_{w_1}$  and  $U_{w_2}$  are exogenous (random) variables indexed by  $W = \{w_1, w_2\}$ . Figure 2(a) shows the associated bipartite graph  $\mathcal{B} = \langle V, F, E \rangle$ . This graph has exactly one perfect matching  $M = \{(v_1 - f_1), (v_2 - f_2)\}$ , which is used in step 2 of the causal ordering algorithm to construct the directed graph  $\mathcal{G}(\mathcal{B}, M)$  in Figure 2(b). The causal ordering graph that is obtained after applying steps 3 and 4 of the causal ordering algorithm is given in Figure 2(c).



**Figure 2:** Several graphs occurring in Example 1. The bipartite graph  $\mathcal{B}$  associated with equations (1) and (2) is shown in (a). The oriented graph  $\mathcal{G}(\mathcal{B}, M)$  obtained in step 2 of the causal ordering algorithm (with perfect matching  $M$ ) is shown in (b). The causal ordering graph  $\text{CO}(\mathcal{B})$  is shown in (c). The corresponding equilibrium Markov ordering graph  $\text{MO}(\mathcal{B})$  is displayed in (d).

Throughout this work, we will assume that sets of equations are *uniquely solvable with respect to the causal ordering graph*, which means that for each cluster, the equations in that cluster can be solved uniquely for the endogenous variables in that cluster (see Definition 14 in ref. [12] for details). This implies amongst others that the endogenous variables in the model can be solved uniquely along a topological ordering of the causal ordering graph. Under this assumption, the causal ordering graph represents the effects of soft and certain perfect interventions [12]. *Soft interventions* target equations; they do not change which variables appear in the targeted equation and may only alter the parameters or functional form of the equation. *Perfect interventions* target clusters in the causal ordering graph and replace the equations in the targeted cluster with equations that set the variables in the cluster equal to constant values. A soft intervention on an equation or a perfect intervention on a cluster has no effect on a variable in the causal ordering graph whenever there is no directed path to that variable from the intervention target (i.e., the targeted equation or an arbitrary vertex in the targeted cluster, respectively), see Theorems 20 and 23 in ref. [12].<sup>6</sup> Since the equations in Example 1 are uniquely solvable with respect to the causal ordering graph in

<sup>6</sup> We say that there is a *directed path* from vertex  $x$  to vertex  $y$  in a directed cluster graph  $\langle \mathcal{V}, \mathcal{E} \rangle$  if either  $\text{cl}(x) = \text{cl}(y)$  or there is a sequence of clusters  $V_1 = \text{cl}(x), V_2, \dots, V_{k-1}, V_k = \text{cl}(y)$ , so that for all  $i \in \{1, \dots, k-1\}$ , there is a vertex  $z_i \in V_i$  such that  $(z_i \rightarrow V_{i+1}) \in \mathcal{E}$ .

Figure 2(c),<sup>7</sup> we can, for example, read off from the causal ordering graph that a soft intervention targeting  $f_1$  may have an effect on  $X_{v_2}$  (since there is a directed path from  $f_1$  to  $v_2$ ), and that a perfect intervention targeting the cluster  $\{v_2, f_2\}$  has no effect on  $X_{v_1}$  (as there is no directed path from the cluster  $\{v_2, f_2\}$  to  $v_1$ ).

Given the probability distribution of the exogenous random variables, one obtains a unique probability distribution on all the variables under the assumption of unique solvability with respect to the causal ordering graph. The *Markov ordering graph* is a DAG  $MO(\mathcal{B})$  such that  $d$ -separations in this graph imply corresponding conditional independences between variables in this joint distribution, provided that all exogenous random variables are independent [12]. The Markov ordering graph is obtained from a causal ordering graph  $CO(\mathcal{B}) = \langle \mathcal{V}, \mathcal{E} \rangle$  by constructing a graph  $\langle V', E' \rangle$  with vertices  $V' = V$  and edges  $(v \rightarrow w) \in E'$  if and only if  $(v \rightarrow \text{cl}(w)) \in \mathcal{E}$ . The Markov ordering graph for the set of equations in Example 1 is given in Figure 2(d). The  $d$ -separations in this graph imply conditional independences between the corresponding variables under the assumption that  $U_{w_1}$  and  $U_{w_2}$  are independent. For instance, since  $v_1$  and  $w_2$  are  $d$ -separated, we can read off from the Markov ordering graph that  $X_{v_1}$  and  $U_{w_2}$  must be independent.

Assuming that the probability distribution is  $d$ -faithful to the Markov ordering graph and that we have a conditional independence oracle, we know that the output of the PC algorithm represents the Markov equivalence class of the Markov ordering graph [10].<sup>8</sup> However, while for acyclic systems, the directed edges in the Markov ordering graph can also be interpreted as direct causal relations, this is not the case for all systems [12,16]. Several examples of this phenomenon are provided in Appendix F. In this work, we will show that such discrepancy between the causal and the Markov structure is actually a common feature of perfectly adapted dynamical systems at equilibrium.

## 2.3 Related work

The causal ordering algorithm can be applied, for instance, to the equilibrium equations of a dynamical system to uncover its causal properties and conditional independence relations at equilibrium. The relationship between dynamical models and causal models has received much attention over the years. For instance, the works of [13,30–35] considered causal relations in dynamical systems that are not at equilibrium, while [6,13,14,16,21,34,36–39] considered graphical and causal models that arise from studying the stationary behaviour of dynamical models. In particular, extensions of the causal ordering algorithm for dynamical systems were proposed and discussed in ref. [13]. Also, the causal ordering algorithm was applied in ref. [16] to study the robustness of model predictions when combining two systems. The relationship between the causal semantics of a dynamical system before it reaches equilibrium and at equilibrium has also been studied [14,34,36].

In previous work, researchers have noted various problems when attempting to use a single graphical model to represent both conditional independence properties and causal properties of certain dynamical systems at equilibrium [14,21,37,39,40]. Often, restrictive assumptions on the underlying dynamical models are made to avoid these subtleties – the most common one being to exclude the possibility of cycles altogether. In this work, we will not make such restrictive assumptions and instead show that such problems are pervasive in the important class of perfectly adapted dynamical systems. We follow [12,16] in addressing the issues by using the causal ordering algorithm to construct separate graphical representations for the causal properties and conditional independence relations implied by these systems.

It has been shown that the popular SCM framework [3,23] is not flexible enough to fully capture the causal semantics (in terms of perfect interventions targeting variables) of the dynamics of a basic enzyme

<sup>7</sup> Indeed, we can solve equation  $f_1$  uniquely for  $X_{v_1}$  in terms of  $U_{w_1}$  resulting in  $X_{v_1} = U_{w_1}$ , and we can solve equation  $f_2$  uniquely for  $X_{v_2}$  in terms of  $X_{v_1}$  and  $U_{w_2}$  resulting in  $X_{v_2} = U_{w_2} - X_{v_1}$ .

<sup>8</sup> More generally, if only a subset of the variables is observed, the output of the FCI algorithm represents the Markov equivalence class of the marginalization (latent projection) of the Markov ordering graph.

reaction at equilibrium, and for that purpose, [39] proposed to use causal constraints models (CCMs) instead. However, CCMs lack a graphical representation (and consequently, all the benefits that usually come with it, like a Markov property and a graphical approach to causal reasoning). The techniques in ref. [12] can also be used to construct graphical representations of causal relations and conditional independences of these models. In Appendix D, we demonstrate that the basic enzyme reaction is perfectly adapting and we show how the causal ordering technique can be used to obtain graphical presentations and a Markov property for this model. This approach offers several advantages over the CCMs approach to model this system.

An application area where obtaining a causal understanding of complex systems is often non-trivial due to feedback and perfect adaptation is that of systems biology. A research question that has seen considerable interest in that field is which reaction networks can achieve perfect adaptation [17,41–44]. The present work provides a method that facilitates the analysis of perfectly adapted dynamical systems by providing a principled and computationally efficient procedure to identify perfect adaptation either from model equations or from experimental data and background knowledge.

In Section 5, we apply our methodology in an attempt to arrive at a better understanding of the causal mechanisms present in protein signalling networks. For protein signalling networks, apparent “causal reversals” have been reported, that is, cases where causal discovery algorithms find the opposite causal relation of what is expected [9,45–48]. One explanation for these reversed edges in the output of causal discovery algorithm could be the unknown presence of measurement error [49]. As we demonstrate in this work, unknown feedback loops that result in perfect adaptation can be another reason why one might find reversed causal relations when applying causal discovery algorithms to biological data.

## 3 Perfect adaptation

In this section, we introduce the notion of perfect adaptation by taking a close look at several examples of dynamical systems that can achieve perfect adaptation. We then consider graphical representations of these systems both before and after they have reached equilibrium. This will set the stage for our main theoretical results regarding the identification of perfect adaptation in models or data, which will be presented in Section 4. The goals of this section are as follows: (i) building intuition about mechanisms that result in perfect adaptation, (ii) outlining the relevance of this phenomenon in application domains, and (iii) explaining how the causal ordering algorithm helps to understand perfect adaptation.

The ability of (part of) a system to converge to its original state when a constant and persistent external stimulus is added or changed is referred to as *perfect adaptation*. If the adaptive behaviour does not depend on the precise setting of the system’s parameters, then we say that the adaptation is *robust*. For our purposes, the most interesting of the two is robust perfect adaptation. As this is also often simply referred to as “perfect adaptation” in the literature, we will do so here as well.

### 3.1 Examples

We consider three different dynamical models corresponding to a filling bathtub, a viral infection with an immune response, and a chemical reaction network. We use simulations to demonstrate that all of these systems are capable of achieving perfect adaptation. The details of the simulations presented in this section are provided in Appendix B, and the code to reproduce these is provided under a free and open source license (see the Data availability statement at the end of this article).



### 3.1.1 Filling bathtub

We consider the example of a filling bathtub of Iwasaki and Simon [13] (see also [12,14]). Let  $I_K(t)$  be an input signal<sup>9</sup> that represents the size of the drain in the bathtub. The inflow rate  $X_I(t)$ , water level  $X_D(t)$ , water pressure  $X_P(t)$ , and outflow rate  $X_O(t)$  are modelled by the following static and dynamic equations:<sup>10</sup>

$$X_I(t) = U_I, \quad (3)$$

$$\dot{X}_D(t) = U_1(X_I(t) - X_O(t)), \quad (4)$$

$$\dot{X}_P(t) = U_2(gU_3X_D(t) - X_P(t)), \quad (5)$$

$$\dot{X}_O(t) = U_4(U_5I_K(t)X_P(t) - X_O(t)), \quad (6)$$

where  $g$  is the gravitational acceleration, and  $U_I$ ,  $U_1$ ,  $U_2$ ,  $U_3$ ,  $U_4$ , and  $U_5$  are independent exogenous random variables taking value in  $\mathbb{R}_{>0}$ . Let  $X_D$ ,  $X_P$ , and  $X_O$  denote the respective equilibrium solutions for the water level, water pressure, and outflow rate. The equilibrium equations associated with this model can easily be constructed by setting the time derivatives equal to zero and assuming the input signal  $I_K(t)$  to have a constant value  $I_K$ :

$$f_I : X_I - U_I = 0, \quad (7)$$

$$f_D : U_1(X_I - X_O) = 0, \quad (8)$$

$$f_P : U_2(gU_3X_D - X_P) = 0, \quad (9)$$

$$f_O : U_4(U_5I_KX_P - X_O) = 0. \quad (10)$$

We call the labelling  $f_D$ ,  $f_P$ , and  $f_O$  that we choose for the equilibrium equations that are constructed from the time derivatives the *natural labelling* for this dynamical system.<sup>11</sup> A solution  $(X_I, X_D, X_P, X_O)$  to the system of equilibrium equations satisfies  $X_I = U_I$  and  $X_O = X_I$  almost surely. From this we conclude that, at equilibrium, the outflow rate is independent of the size of the drain  $I_K$ , assuming that  $U_I$  is independent of  $I_K$ . We recorded the changes in the system after we changed the input signal  $I_K$  of the bathtub system in equilibrium. The results in Figure 3(a) show that the outflow rate  $X_O$  has a transient response to changes in the input signal  $I_K$ ,<sup>12</sup> but it eventually converges to its original value. We say that the outflow rate  $X_O$  in the bathtub model *perfectly adapts* to changes in  $I_K$ .

### 3.1.2 Viral infection model

We consider the example of a simple dynamical model for a viral infection and immune response of De Boer [15] (also discussed in ref. [16]). The model describes target cells  $X_T(t)$ , infected cells  $X_I(t)$ , and an immune response  $X_E(t)$ . We will treat  $I_\sigma(t)$  as an exogenous input signal that represents the production rate of target cells. The system is defined by the following dynamic equations:

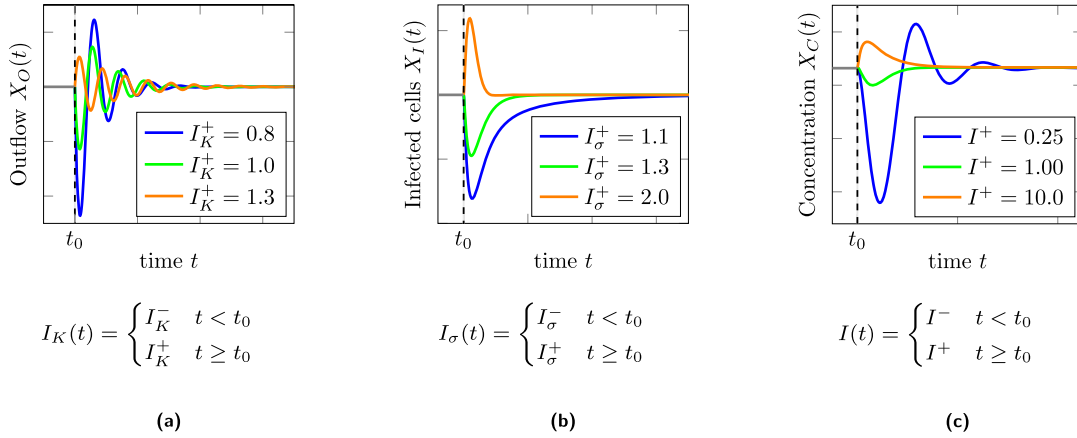
$$\dot{X}_T(t) = I_\sigma(t) - d_TX_T(t) - \beta X_T(t)X_I(t), \quad (11)$$

<sup>9</sup> In this work, “input signal” will always refer to an exogenous process, that is, a variable whose value may depend on time and is determined by some mechanism that is external to the system. In particular, it must not be causally influenced by the system.

<sup>10</sup> With the term “static equation,” we refer to equations that do not contain any time derivatives. In this article, the term “dynamic equation” always corresponds to a first-order differential equation.

<sup>11</sup> In general, the natural labelling uses the same index  $i$  to label as  $f_i$  the equilibrium equation  $0 = h_i(X(t))$  obtained from the dynamic equation  $\dot{X}_{v_i}(t) = h_i(X(t))$  that models the dynamics of variable  $X_{v_i}$  [36].

<sup>12</sup> More precisely, when we speak of a “response” to an input signal, what we mean is that if we have two identical copies of a system, and from some point in time  $t_0$  on, we change the input signal of one of the two copies, the endogenous response variables differ in distribution at some later time  $t > t_0$ .



**Figure 3:** Simulations of the outflow rate  $X_O(t)$  in the bathtub model (a), the amount of infected cells  $X_I(t)$  in the viral infection model (b), and the concentration  $X_C(t)$  in the biochemical reaction network with a negative feedback loop (c). The input signal suddenly changes from a (constant) value for  $t < t_0$  to a different (constant) value for  $t \geq t_0$ . The timing of this change is indicated by a vertical dashed line. The three systems started with input signals  $I_K^- = 1.2$ ,  $I_\sigma^- = 1.6$ , and  $I^- = 1.5$ , respectively. After a transient response,  $X_O(t)$ ,  $X_I(t)$ , and  $X_C(t)$  all converge to their original equilibrium value (i.e., they perfectly adapt to the input signal).

$$\dot{X}_I(t) = (\beta X_T(t) - d_I - k X_E(t)) X_I(t), \quad (12)$$

$$\dot{X}_E(t) = (a X_I(t) - d_E) X_E(t). \quad (13)$$

Here,  $\beta = \frac{bp}{c}$ , where  $b$  is the infection rate,  $p$  is the number of virus particles produced per infected cell, and  $c$  is the clearance rate of viral particles. Furthermore,  $d_T$  is the death rate of target cells,  $a$  is an activation rate,  $d_E$  and  $d_I$  are turnover rates, and  $k$  is a mass-action killing rate. We assume that  $a$  and  $k$  are constants and that  $d_T$ ,  $d_I$ ,  $d_E$ , and  $\beta$  are independent exogenous random variables. We use the natural labelling for the equilibrium equations that are constructed from the differential equations:<sup>13</sup>

$$f_T : I_\sigma - d_T X_T - \beta X_T X_I = 0, \quad (14)$$

$$f_I : \beta X_T - d_I - k X_E = 0, \quad (15)$$

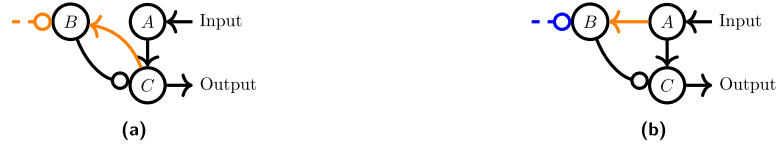
$$f_E : a X_I - d_E = 0, \quad (16)$$

assuming a constant value  $I_\sigma$  of the input signal. We initialized the model in an equilibrium state and simulated the response of the model after changing the input signal  $I_\sigma$  to three different values. Figure 3(b) shows that the amount of infected cells  $X_I(t)$  has a transient response to a change in the input signal  $I_\sigma$ , but then returns to its original value. Hence, the amount of infected cells perfectly adapts to changes in the production rate of target cells.

### 3.1.3 Reaction networks with a negative feedback loop

The phenomenon of perfect adaptation is a common feature in biochemical reaction networks, and there exist many reaction networks that can achieve (near) perfect adaptation [41,43]. For networks consisting of only three nodes, Ma et al. [17] found by an exhaustive search that there exist two major classes of reaction networks that produce (robust) adaptive behaviour. The reaction diagrams for these networks are given in Figure 4. Here, we will only analyze the “Negative Feedback with a Buffer Node” (NFBN) network. An

<sup>13</sup> Following [15], we are only interested in strictly positive solutions of this dynamical system. Therefore, we use the equilibrium equation  $f_I$  instead of  $(\beta X_T - d_I - k X_E) X_I = 0$  and  $f_E$  instead of  $(a X_I - d_E) X_E = 0$ .



**Figure 4:** The two reaction networks that can achieve perfect adaptation [17]. (a) NFBN network, and (b) IFFLP network. Orange edges represent saturated reactions, blue edges represent linear reactions, and black edges are unconstrained reactions. Arrowheads represent positive influence and edges ending with a circle represent negative influence.

analysis of the other network, the “Incoherent Feed-Forward Loop with a Propportioner Node” (IFFLP), is provided in Appendix E. The NFBN system can be described by the following first-order differential equations:

$$\dot{X}_A(t) = I(t)k_{IA}\frac{(1 - X_A(t))}{K_{IA} + (1 - X_A(t))} - F_A k_{FA} \frac{X_A(t)}{K_{FA} + X_A(t)}, \quad (17)$$

$$\dot{X}_B(t) = X_C(t)k_{CB}\frac{(1 - X_B(t))}{K_{CB} + (1 - X_B(t))} - F_B k_{FB} \frac{X_B(t)}{K_{FB} + X_B(t)}, \quad (18)$$

$$\dot{X}_C(t) = X_A(t)k_{AC}\frac{(1 - X_C(t))}{K_{AC} + (1 - X_C(t))} - X_B(t)k_{BC}\frac{X_C(t)}{K_{BC} + X_C(t)}, \quad (19)$$

where  $X_A(t)$ ,  $X_B(t)$ , and  $X_C(t)$  are concentrations of three compounds A, B, and C, respectively, while  $I(t)$  represents an external input into the system. Assume that  $k_{IA}$ ,  $k_{CB}$ , and  $k_{AC}$  are independent exogenous random variables, which we will denote as  $U_A$ ,  $U_B$ , and  $U_C$ , respectively, and that the other parameters are constants. Perfect adaptation is achieved under saturation conditions [17],  $(1 - X_B(t)) \gg K_{CB}$  and  $X_B(t) \gg K_{FB}$ , in which case the following approximation can be made:

$$\dot{X}_B(t) \approx X_C(t)k_{CB} - F_B k_{FB}. \quad (20)$$

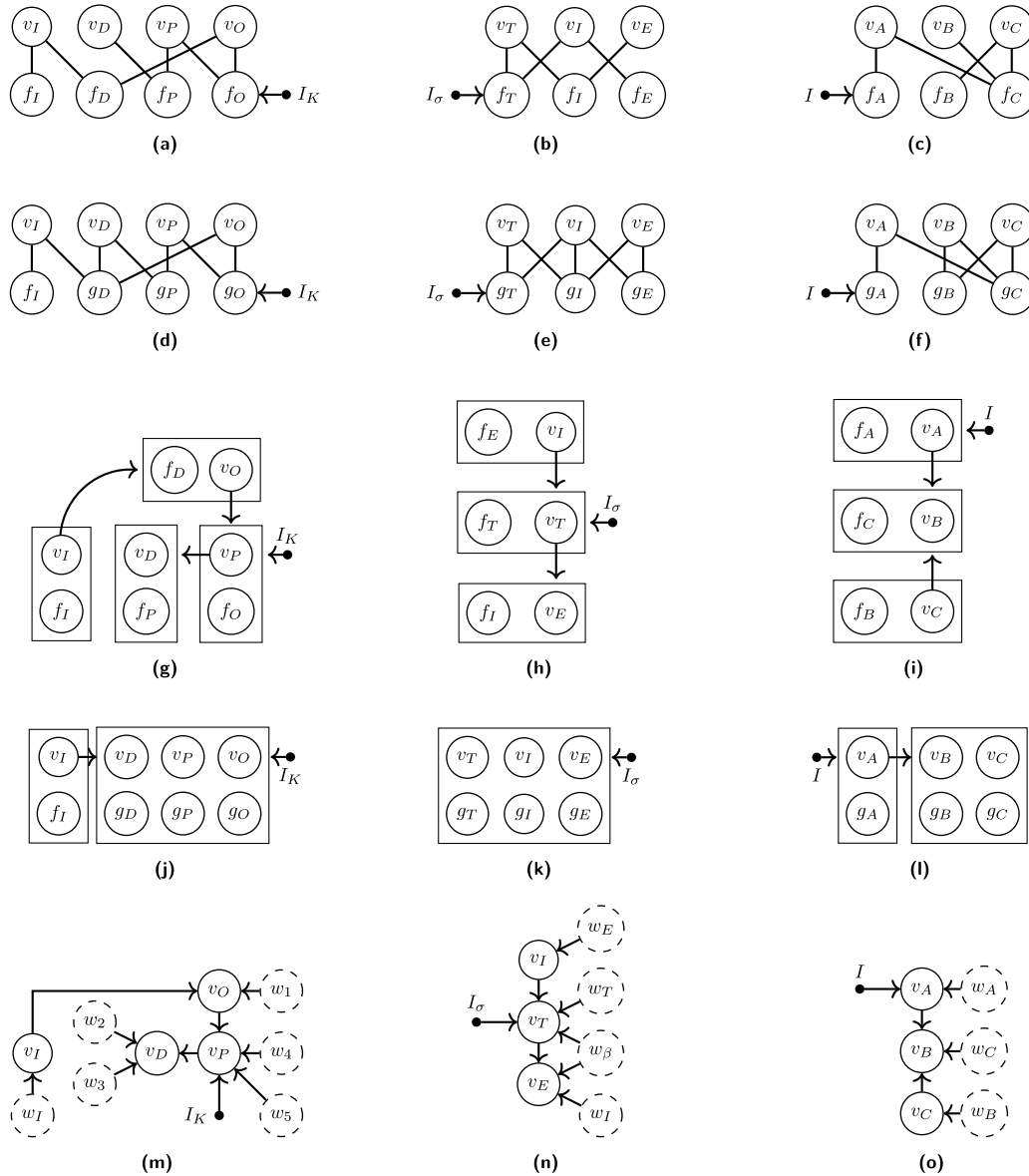
Under the assumption that  $I(t)$  has a constant value, the system converges to an equilibrium. We will denote the equilibrium equations that are associated with the time derivatives  $\dot{X}_A(t)$  and  $\dot{X}_C(t)$  using the natural labelling  $f_A$  and  $f_C$ . The equilibrium equation  $f_B$  is obtained by setting the approximation of the time derivative  $\dot{X}_B(t)$  in (20) equal to zero. We initialized this model in an equilibrium state and then simulated its response after changing the input signal  $I$  to three different values. Figure 3(c) shows that  $X_C(t)$  perfectly adapts to changes in the input signal  $I$ .

### 3.2 Graphical representations

We will now provide the different graphical representations of the perfectly adapted dynamical systems that were introduced in the previous section. These representations are based on the graphs that are used in refs [12,13] to encode the equilibrium structure of equations, causal relations, and conditional independences. The main difference with the previous work is that we also explicitly consider similar graphical representations for systems that have not yet reached equilibrium.

*Bipartite graph:* The *equilibrium bipartite graphs* associated with the equilibrium equations of the filling bathtub, the viral infection, and the reaction network with feedback are given in Figure 5(a), (b), and (c), respectively. We have added also a node representing the input signal. The *dynamic bipartite graphs* for the dynamics of these models are constructed from first-order differential equations in canonical form<sup>14</sup> in the

<sup>14</sup> A set of first-order differential equations is said to be in *canonical form* if they are of the form  $\dot{X}_i(t) = h_i(X_1(t), \dots, X_n(t))$  with  $i = 1, \dots, n$ . That is, each differential equation has a single derivative on the left-hand side and a function of the variables (without any derivatives) on the right-hand side.



**Figure 5:** Graphical representations of the bathtub model (left column), the viral infection model (center column), and the reaction network with negative feedback (right column). The input vertices  $I_K$ ,  $I_\sigma$ , and  $I$  are represented by black dots, while exogenous variables are indicated by dashed circles. For each model, the structure of the equilibrium equations can be read off from the equilibrium bipartite graphs in (a)–(c). The structure of the first-order differential equations can be represented with the dynamic bipartite graphs in (d)–(f). The equilibrium causal ordering graphs corresponding to the equilibrium bipartite graphs are given in (g)–(i). Similarly, the dynamic causal ordering graphs corresponding to the dynamic bipartite graphs can be found in (j)–(l). The equilibrium Markov ordering graphs for the equilibrium distribution of the models are given in (m)–(o).

following way. Both the derivative  $\dot{X}_i(t)$  and the corresponding variable  $X_i(t)$  are associated with the same vertex  $v_i$ . The natural labelling is used for the differential equations, so that a vertex  $g_i$  is associated with the differential equation for  $\dot{X}_i(t)$ . We then construct the dynamic bipartite graph  $\mathcal{B}_{\text{dyn}} = \langle V, F, E \rangle$  with variable vertices  $v_i \in V$  and the corresponding dynamical equation vertices  $g_i \in F$ . Additional static equation vertices  $f_i \in F$  are added as well in case the dynamical system consists of a combination of dynamic and static equations. The edge set  $E$  has an edge  $(v_i - f_j)$  whenever  $X_i(t)$  appears in the static equation  $f_j$ . In addition, there are edges  $(v_i - g_j)$  whenever  $X_i(t)$  or  $\dot{X}_i(t)$  appears in the dynamic equation  $g_j$  (which includes the cases

$i = j$  due to the natural labelling used). The dynamic bipartite graphs for the bathtub model, the viral infection, and the reaction network with feedback are given in Figure 5(d), (e), and (f), respectively.

Comparing the equilibrium bipartite graphs with the dynamic bipartite graphs we note that there is no edge  $(v_D - f_D)$  in Figure 5(a), while  $(v_D - g_D)$  is present in Figure 5(d). This is a direct consequence of the fact that the time derivative  $\dot{X}_D(t)$  in equation (4) does not depend on  $X_D(t)$  itself. Similarly, the edges  $(v_I - f_I)$  and  $(v_E - f_E)$  are not present in Figure 5(b), whilst the edges  $(v_I - g_I)$  and  $(v_E - g_E)$  are present in Figure 5(e). In this case, we see that even though the time derivatives  $\dot{X}_I(t)$  and  $\dot{X}_E(t)$  depend on  $X_I(t)$  and  $X_E(t)$  in differential equations (12) and (13), these variables do not appear in the associated equilibrium equations (15) and (16). Finally, there is no edge  $(v_B - f_B)$  in Figure 5(c), while the edge  $(v_B - g_B)$  is present in Figure 5(f). Here, the variable  $X_B(t)$  does not appear in the equilibrium equation under saturation conditions (20) that stems from the dynamic equation (18) for  $\dot{X}_B(t)$ . The equilibrium bipartite graph can be compared to the dynamic bipartite graph to read off structural differences between the equations before and after equilibrium has been reached.

*Causal ordering graph:* The application of the causal ordering algorithm to the equilibrium bipartite graphs of the filling bathtub, the viral infection, and the reaction network results in the *equilibrium causal ordering graphs* in Figure 5(g), (h), and (i), respectively. Henceforth, we will assume that the dynamic bipartite graph has a perfect matching that extends the natural labelling of the dynamic equations, i.e., such that all pairs  $(v_i, g_i)$  are matched. Application of the causal ordering algorithm to the associated dynamic bipartite graph for the model of a filling bathtub, the viral infection model, and the reaction network results in the *dynamic causal ordering graphs* in Figure 5(j), (k), and (l), respectively.<sup>15</sup>

As shown in ref. [12], the absence (presence) of a directed path from an equation vertex to a variable vertex in the equilibrium causal ordering graph indicates that a soft intervention targeting a parameter in that equation has no (a generic) effect on the value of that variable once the system has reached equilibrium again. Furthermore, the absence (presence) of a directed path from a cluster to a variable vertex in the equilibrium causal ordering graph indicates that a perfect intervention targeting the cluster has no (a generic) effect on the value of that variable once the system has reached equilibrium again. Notice that the variables  $v_i$  in the equilibrium causal ordering graph do not always end up in the same cluster with the equilibrium equation  $f_i$  of the natural labelling. For example, we see in Figure 5(g) that a soft intervention targeting the equilibrium equation  $f_O$  (e.g., a change in the value of  $U_5$ ) does *not* affect the value of the outflow rate  $X_O$  at equilibrium (since there is no directed path from  $f_O$  to  $v_O$ ), even though  $f_O$  was obtained from the dynamic equation for the time derivative of the outflow rate  $X_O(t)$ . Similarly, targeting  $f_I$  with a soft intervention in the viral infection model has no effect on  $X_I$  at equilibrium and targeting  $f_C$  in the reaction network model has no effect on the equilibrium distribution of  $X_C$ .<sup>16</sup> This suggests that the causal structure at equilibrium of perfectly adapted dynamical systems may differ from the transient causal structure. In the next section, we will use this idea to detect perfect adaptation from background knowledge and experimental data.

*Equilibrium Markov ordering graph:* As explained in Section 2.2, the Markov ordering graph is constructed from the causal ordering graph and includes exogenous variables. For the bathtub model, we let vertices  $w_I, w_1, \dots, w_5$  represent the independent exogenous random variables  $U_I, U_1, \dots, U_5$  that appear in the model. For the viral infection model, we let  $w_T, w_I, w_E,$  and  $w_\beta$  represent independent exogenous random variables  $d_T, d_I, d_E,$  and  $\beta$  in equations (11), (12), and (13). Finally, for the reaction network with negative feedback, we let  $w_A, w_B,$  and  $w_C$  represent the independent exogenous random variables  $k_{IA}, k_{CB},$  and  $k_{AC},$  respectively. The equilibrium Markov ordering graphs for the filling bathtub model, the viral infection model, and the model of a reaction network with a negative feedback loop are given in Figure 5(m), (n), and (o),

<sup>15</sup> Our approach here differs from the dynamic causal ordering algorithm proposed in ref. [13], which includes separate vertices for derivatives and variables that are linked by “definitional” integration links.

<sup>16</sup> To preserve an unambiguous causal interpretation, equations and clusters that may be targeted by interventions should be clearly distinguished from the variables that could be affected by those interventions [12].

respectively.<sup>17</sup> These equilibrium Markov ordering graphs can be used to read off conditional independences in the equilibrium distribution that are implied by the equilibrium equations of the model. For example, since  $v_I$  is  $d$ -separated from  $v_D$  given  $v_P$  in the equilibrium Markov ordering graph in Figure 5(m),  $X_I$  will be independent of  $X_D$  given  $X_P$  once the system has reached equilibrium. These independences can be tested for in equilibrium data by means of statistical conditional independence tests. These implied conditional independences can, for instance, be used in the process of model selection [16].

### 3.3 Existence and uniqueness of solutions

The causal ordering algorithm is a graphical tool that can be useful when solving a system of equations. It decomposes the question of existence and uniqueness of a “global” solution into several “local” existence and uniqueness problems corresponding to a partitioning of the equations. When a unique global solution exists for all possible joint values of the (independent) exogenous variables, this leads to both a causal semantics and a Markov property [12]. We argue here that these ideas can also be extended to include differential equations. We will illustrate this with the filling bathtub model. We start with the (conceptually simpler) equilibrium model, which solely contains static equations, before discussing what to do when dynamic equations are present.

The equilibrium equations (7)–(10) can be solved in steps by following the topological ordering of the clusters in the equilibrium causal ordering graph in Figure 5(g). First, use  $f_I$  to solve for  $X_I$ , resulting in  $X_I = U_I$ . Then, use  $f_D$  to solve for  $X_O$ , which results in  $X_O = X_I$ . Subsequently, use  $f_P$  to solve for  $X_P$ , yielding  $X_P = \frac{X_O}{U_5 I_K}$ . Finally, use  $f_D$  to solve for  $X_D$ , resulting in  $X_D = \frac{X_P}{g U_3}$ . By substitution, we obtain a global solution of the form

$$(X_I, X_O, X_P, X_D) = \left( U_I, U_I, \frac{U_I}{U_5 I_K}, \frac{U_I}{g U_3 U_5 I_K} \right).$$

Since we obtain a unique solution of each equation for the target variable in terms of the other variables appearing in the equation, this procedure shows that there exists a unique global solution of the system of equations for any value of the exogenous variables  $U_I, U_1, U_2, U_3, U_4, U_5$  and any value of the input signal  $I_K$ . Because of this, we obtain both a causal interpretation and a Markov property for the filling bathtub model at equilibrium as described in Section 3.2.

For the dynamic filling bathtub model, we can follow a similar procedure, but now the clusters may also contain differential equations. We can make use of the theory for the existence and uniqueness of solutions of ordinary differential equations. First, note that the dynamic bipartite graph reflects the structure of the static and dynamic equations once we rewrite the differential equations as integral equations. For example, for the time interval  $[t_0, t]$ :

$$X_I(t) = U_I, \tag{21}$$

$$X_D(t) = X_D(t_0) + \int_{t_0}^t U_1(X_I(\tau) - X_O(\tau))d\tau, \tag{22}$$

$$X_P(t) = X_P(t_0) + \int_{t_0}^t U_2(g U_3 X_D(\tau) - X_P(\tau))d\tau, \tag{23}$$

<sup>17</sup> The equilibrium Markov ordering graph for the bathtub model coincides with the result of [14], who simulated data from the bathtub model until the system reached equilibrium and then applied the PC algorithm to the equilibrium data. Although ref. [14] interprets the learned graphical representation as the “causal graph,” this graph in itself does not have a straightforward causal interpretation. See Appendix F and the discussion in ref. [12] for more details.

$$X_O(t) = X_O(t_0) + \int_{t_0}^t U_4(U_5 I_K(\tau) X_P(\tau) - X_O(\tau)) d\tau. \quad (24)$$

Rewriting the differential equations as integral equations has two advantages: (i) there is no need to introduce the derivatives as if they were (variation) independent processes; (ii) it makes the dependence on the initial conditions  $X_D(t_0)$ ,  $X_P(t_0)$ , and  $X_O(t_0)$  explicit. Equations (3)–(6) describing the dynamical system can be solved in steps by following the topological ordering of the clusters in the dynamic causal ordering graph in Figure 5(j). First, solve  $f_I$  for  $X_I$ , resulting in  $X_I(t) = U_I$ . The cluster  $\{g_D, g_P, g_O, v_D, v_P, v_O\}$  has to be dealt with as a single unit, which means we have to solve the subsystem of three differential equations  $\{g_D, g_P, g_O\}$  (i.e., equations (4)–(6)) for its solution with components  $(X_D(t), X_P(t), X_O(t))$ . By applying the Picard–Lindelöf theorem (see, e.g., [50]), one obtains that this subsystem has a unique solution on a time interval  $[t_0, \infty)$  for any initial condition  $(X_D(t_0), X_P(t_0), X_O(t_0))$ , provided that  $X_I(t)$  and  $I_K(t)$  are continuous and that the input signal  $I_K(t)$  is bounded. Thus, equations (3)–(6) have a unique global solution for any value of the exogenous variables  $U_I, U_1, U_2, U_3, U_4, U_5$ , any initial condition  $(X_D(t_0), X_P(t_0), X_O(t_0))$ , and any continuous and bounded input signal  $I_K(t)$ . The approach of [12] can in this way be extended to yield both a dynamic causal interpretation and a Markov property (by using the trick of [34] to interpret path-continuous stochastic processes as random variables).<sup>18</sup>

Important to note here is that this explicit solution procedure shows that at equilibrium, the value of the input signal  $I_K$  may affect the value of  $X_D$  and  $X_P$ , but cannot affect the values of  $X_O$  and  $X_I$ , while there can be transient effects of the input signal  $I_K(t)$  on  $X_D(t)$ ,  $X_P(t)$ , and  $X_O(t)$ , but not on  $X_I(t)$ . Furthermore, under appropriate local solvability conditions for each cluster, these observations can directly be read off from the (equilibrium and dynamic) causal ordering graphs.

## 4 Identification of perfect adaptation

In Section 3.1, we identified perfect adaptation in three simple models through simulations. Here, we consider how to identify models that are capable of perfect adaptation without requiring simulations or explicit calculations. In Section 4.1, we will put the graphical representations of Section 3.2 to use for identifying perfect adaptation in dynamical models. We discuss possibilities for the identification of perfect adaptation from equilibrium data in Section 4.2.

### 4.1 Identification of perfect adaptation via causal ordering

The identification of perfect adaptation via causal ordering makes use of the causal semantics of the equilibrium causal ordering graph. The following lemma states that a change in the input signal has no effect on the value of a variable if there is no directed path from the input vertex to that variable in the equilibrium causal ordering graph.

**Lemma 1.** *Consider a model consisting of static equations, a set of first-order differential equations in canonical form, and an input signal. Assume that the equilibrium bipartite graph has a perfect matching and that the static equations and equilibrium equations derived from the first-order differential equations are uniquely solvable with respect to the equilibrium causal ordering graph for all relevant values of the input signal. If there is no directed path from the input vertex to a variable vertex in the equilibrium causal ordering graph, then the value of the input signal does not influence the equilibrium distribution of that variable.*

<sup>18</sup> A more formal and rigorous treatment is left as future work.

**Proof.** The statement follows directly from Theorem 20 in ref. [12].  $\square$

To establish perfect adaptation, we assume that the presence of a directed path in the dynamic causal ordering graph implies the presence of a transient causal effect.

**Assumption 1.** Consider a model consisting of static equations, a set of first-order differential equations in canonical form, and an input signal. Assume that the dynamic bipartite graph has a perfect matching that extends the natural labelling. If there is a directed path from the input vertex to a variable vertex in the dynamic causal ordering graph, then there will be a response of that variable to changes in the input signal some time later.

Intuitively, this assumption may seem plausible, as the presence of the directed path in the dynamic causal ordering graph implies that the input signal enters into the construction of the solution of the variable, as sketched in Section 3.3. Unless a perfect cancellation occurs, one then expects a generic effect on the solution some time after the change in the input signal. Assumption 1 can be seen as a consequence of a certain faithfulness assumption.<sup>19</sup> We conjecture that this assumption is generically satisfied for a large class of dynamical systems (e.g., it might hold for almost all parameter values with respect to the Lebesgue measure on a suitable parameter space).<sup>20</sup>

By combining Lemma 1 and Assumption 1, we immediately obtain the following result.

**Theorem 1.** *Consider a model that satisfies the conditions of Lemma 1 and assume that the associated dynamic bipartite graph has a perfect matching that extends the natural labelling. Under Assumption 1, the presence of a directed path from the input signal  $I$  to a variable  $X_v$  in the dynamic causal ordering graph and the absence of such a path in the equilibrium causal ordering graph together imply that  $X_v$  perfectly adapts to changes in the input signal  $I$  if the system equilibrates.*

Theorem 1 can be directly applied to the equilibrium and dynamic causal ordering graphs in Figure 5 to identify perfect adaptation. For example, we see that there is a directed path from the input signal  $I_K$  to  $v_O$  in the dynamic causal ordering graph of the filling bathtub in Figure 5(j), while no such path exists in the equilibrium causal ordering graph in Figure 5(g). It follows from Theorem 1 that  $X_O$  perfectly adapts to changes in the input signal  $I_K$ . This is in agreement with the simulation in Figure 3(a). Similarly, we can verify that the amount of infected cells  $X_I$  in the viral infection model perfectly adapts to changes in the input signal  $I_\sigma$  and that  $X_C$  perfectly adapts to  $I$  in the reaction network with negative feedback. Hence, perfect adaptation in the bathtub model, the viral infection model, and the reaction network with negative feedback can be identified by applying the graphical criteria in Theorem 1 to the respective causal ordering graphs. It is important to keep in mind, though, that what we can identify in this way is only the *possibility* of perfectly adaptive behaviour, relying on the implicit assumption that the system will actually equilibrate.

In Appendix D, we show that the sufficient conditions in Theorem 1 for the identification of perfect adaptation are not necessary. More specifically, we construct graphical representations for a dynamical model of a basic enzymatic reaction that achieves perfect adaptation but does not satisfy the conditions in Theorem 1. Interestingly, though, after rewriting the equations, the perfectly adaptive behaviour of these systems can be captured via Theorem 1. Further, in Appendix E, we show that the biochemical reaction network in Figure 4(b), which Ma et al. [17] identified as being capable of achieving perfect adaptation, does not satisfy the conditions in Theorem 1 either. We show that a change of variables enables one to still capture the perfectly adaptive behaviour of this system via Theorem 1.

<sup>19</sup> Indeed, it appears that Assumption 1 follows from the faithfulness assumption that corresponds with a Markov property that was derived for structural dynamical causal models [34].

<sup>20</sup> Proving this in sufficient generality seems not straightforward; indeed, even the well-known result that  $d$ -faithfulness is a generic property has only been shown so far for Bayesian networks with discrete variables and for linear-Gaussian structural equation models [51].



## 4.2 Identification of perfect adaptation from data

So far we have only considered how perfect adaptation can be identified in mathematical models. In this section, we focus on methods for identifying perfect adaptation from data that are generated by perfectly adapted dynamical systems under experimental conditions. The most straightforward approach to detect perfect adaptation is to collect time-series data while experimentally changing the input signal to the system. One can then simply observe whether the variables in the system revert to their original values. However, this type of experimentation is not always feasible. Another way to identify feedback loops that achieve perfect adaptation uses a combination of observational equilibrium data, background knowledge, and experimental data. Our second main result, Theorem 2, gives sufficient conditions under which we can identify a system that is capable of perfect adaptation from experimental equilibrium data.

**Theorem 2.** *Consider a set of first-order dynamical equations in canonical form for variables  $V$ , satisfying the conditions of Theorem 1, with equilibrium equations  $F$  under the natural labelling. Consider a soft intervention targeting an equation  $f_i \in F$ . Assume that the system is uniquely solvable with respect to the equilibrium causal ordering graph both before and after the intervention and that the intervention alters the equilibrium distribution of all descendants of  $f_i$  in the equilibrium causal ordering graph. If either*

- (1) *The soft intervention does not change the equilibrium distribution of  $X_i$  or*
  - (2) *the soft intervention alters the equilibrium distribution of a variable corresponding to a nondescendant of  $v_i$  in the equilibrium Markov ordering graph,*
- (or both), then the system is capable of perfect adaptation.*

**Proof.** The proof is given in Appendix A. □

Theorem 2 applies in particular to experiments on the filling bathtub, viral infection, and chemical reaction systems (for the corresponding graphs, see Figure 5). For example, a soft intervention targeting  $f_O$  in the bathtub example has no effect on the outflow rate at equilibrium  $X_O$ , because there is no directed path from  $f_O$  to  $v_O$  in Figure 5(g), and an intervention targeting  $f_C$  has no effect on the equilibrium concentration  $X_C$  in the reaction network because there is no directed path from  $f_C$  to  $v_C$  in Figure 5(i). In both cases, the first condition of Theorem 2 is satisfied. For the viral infection model, we see that a soft intervention targeting  $f_E$  has an effect on the amount of infected cells  $X_I$  at equilibrium (since there is a directed path from  $f_E$  to  $v_I$  in Figure 5(h)), while there is no directed path from  $v_E$  to  $v_I$  in Figure 5(n). In this case, the second condition of Theorem 2 is satisfied.

We can devise the following scheme to detect perfectly adapted dynamical systems from data and background knowledge. The procedure relies on several assumptions, including  $d$ -faithfulness of the equilibrium distribution to the equilibrium Markov ordering graph. We start by collecting observational equilibrium data and use a causal discovery algorithm (such as LCD or FCI) to learn a (partial) representation of the equilibrium Markov ordering graph, assuming the observational distribution at equilibrium to be  $d$ -faithful with respect to the equilibrium Markov ordering graph. We then consider a soft intervention that changes a known equation in the first-order differential equation model (i.e., it targets a known equilibrium equation). If this intervention does not change the distribution of the variable corresponding to this target using the natural labelling, or if it changes the distribution of identifiable non-descendants of the variable corresponding to the target according to the learned Markov equivalence class, we can apply Theorem 2 to identify the dynamical system as being capable of perfect adaptation. This way, we could identify perfect adaptation in specific cases such as the filling bathtub, viral infection, and reaction network by exploiting a combination of background knowledge and experimental data. Another example of a possible application of Theorem 2 is given in Section 5.3.

## 5 Perfect adaptation in protein signalling

In this section, we apply the ideas developed in the previous sections to a biological system that has been intensely studied during the past decades to emphasize the practical relevance of perfect adaptation. The so-called RAS-RAF-MEK-ERK signalling cascade is a textbook example of a *protein signalling network*, which forms an important ingredient of the “machinery” of cells in living organisms. The molecular pathways in such a network fulfil various important functions, for instance, the transmission and processing of information. Systems biologists make use of dynamical systems to model such networks both qualitatively and quantitatively. Because of the high complexity of protein signalling networks, which typically consist of many different interacting components, this has also been considered a promising application domain for causal discovery methods.

In an influential article, Sachs et al. [18] applied causal discovery to reconstruct a protein signalling network from experimental data. Over the years, the dataset of [18] has become an often used “benchmark” for assessing the accuracy of causal discovery algorithms, where the “consensus network” in ref. [18] is usually considered as the perfect ground truth. The apparent successes of causal discovery on this particular dataset may have led to the impression that causal discovery algorithms can in general successfully discover the causal semantics of complex protein signalling networks from real-world data. However, this success has hitherto not been repeated on other, similar datasets, to the best of our knowledge. Indeed, modelling and understanding such systems and inferring their behaviour and structure from data still pose many challenges, for instance, because of feedback loops and the inherent dynamical nature of such systems [52].

In this section, we focus on understanding the properties of the equilibrium distribution of a simple model of the RAS-RAF-MEK-ERK signalling pathway, and specifically we investigate the phenomenon of perfect adaptation. Like many other biological systems, protein signalling networks often show adaptive behaviour that helps to ensure a certain robustness of their functionality against various disturbances and perturbations [43]. By using the technique of causal ordering to analyze the conditional independences and causal relations that are implied by the model at equilibrium, we elucidate the causal interpretation of the output of constraint-based causal discovery algorithms when they are applied to equilibrium protein expression data if the parameters are such that the system shows perfect adaptation.

We test some of the model’s predictions on real-world data and compare with another model that has been proposed. We do not claim that the perfectly adaptive model that we analyze here is a realistic model of the protein signalling pathway. Although we will show in Section 5.4 that the model is able to explain certain observations in real-world data, this is not that surprising for a model with that many parameters.<sup>21</sup> Instead, our goal is to demonstrate that in systems with perfect adaptation, the standard interpretation of the output of causal discovery algorithms may not apply.<sup>22</sup> This could explain why the output of certain causal discovery algorithms applied to the data of [18] appears to be at odds with the biological consensus network presented in ref. [18], see, for example, [47] and [9].

This section is structured as follows. In Section 5.1 we introduce the perfectly adaptive model for the signalling pathway. We proceed with the associated graphical representations in Section 5.2. Then, in Section 5.3, we study the model’s predictions under a soft intervention and verify these in simulations. In Section 5.4 we take a closer look at some real-world data, more specifically, the data from [18], and compare the model’s predictions with the data. In Section 5.5, we explain how the phenomenon of perfect adaptation may lead to unexpected outcomes of causal discovery methods. In the end, we will have to conclude that the causal structure of the RAS-RAF-MEK-ERK cascade seems far from understood, and that it seems unlikely that the data in ref. [18] is sufficiently rich to be able to draw strong conclusions regarding the causal behaviour of the signalling network.

---

<sup>21</sup> As John von Neumann once put it: “With four parameters I can fit an elephant, and with five I can make him wiggle his trunk.”

<sup>22</sup> This was already pointed out in ref. [14] for the example of the filling bathtub, but our work shows how widespread this phenomenon may be, and thereby emphasizes its practical relevance for causal discovery.

## 5.1 Dynamical model

We adapt the mathematical model of [19] for the RAS-RAF-MEK-ERK signalling cascade, as in ref. [16].<sup>23</sup> Let  $V = \{v_s, v_r, v_m, v_e\}$  be an index set for endogenous variables that represent the equilibrium concentrations  $X_s$ ,  $X_r$ ,  $X_m$ , and  $X_e$  of active (phosphorylated) RAS, RAF, MEK, and ERK proteins, respectively. We model their dynamics as follows:

$$\dot{X}_s(t) = \frac{I(t)k_{Is}(T_s - X_s(t))}{(K_{Is} + (T_s - X_s(t)))\left(1 + \left(\frac{X_e(t)}{K_e}\right)^{\frac{3}{2}}\right)} - F_s k_{F_s s} \frac{X_s(t)}{K_{F_s s} + X_s(t)}, \quad (25)$$

$$\dot{X}_r(t) = \frac{X_s(t)k_{sr}(T_r - X_r(t))}{K_{sr} + (T_r - X_r(t))} - F_r k_{F_r r} \frac{X_r(t)}{K_{F_r r} + X_r(t)}, \quad (26)$$

$$\dot{X}_m(t) = \frac{X_r(t)k_{rm}(T_m - X_m(t))}{K_{rm} + (T_m - X_m(t))} - F_m k_{F_m m} \frac{X_m(t)}{K_{F_m m} + X_m(t)}, \quad (27)$$

$$\dot{X}_e(t) = \frac{X_m(t)k_{me}(T_e - X_e(t))}{K_{me} + (T_e - X_e(t))} - F_e k_{F_e e} \frac{X_e(t)}{K_{F_e e} + X_e(t)}, \quad (28)$$

where we assume that  $I(t)$  is an external stimulus or perturbation. Roughly speaking, there is a signalling pathway that goes from  $I(t)$  to  $X_s(t)$  to  $X_r(t)$  to  $X_m(t)$  to  $X_e(t)$  with negative feedback from  $X_e(t)$  on  $X_s(t)$ . As we did for the reaction network with negative feedback in Section 3.1.3, we will consider the system under certain saturation conditions. Specifically, for  $(T_e - X_e(t)) \gg K_{me}$  and  $X_e(t) \gg K_{F_e e}$ , the following approximation holds:

$$\dot{X}_e(t) \approx X_m(t)k_{me} - F_e k_{F_e e}. \quad (29)$$

We let  $f_s$ ,  $f_r$ ,  $f_m$ , and  $f_e$  represent the equilibrium equations corresponding to the dynamical equations in (25), (26), (27), and (29) respectively, where we assume the input signal to have a constant (possibly random) value  $I$ .

We simulated the model under these saturation conditions (picking values for  $K_{me}$  and  $K_{F_e e}$  close to zero) until it reached equilibrium, and then we recorded the changes in the concentrations  $X_s(t)$ ,  $X_r(t)$ ,  $X_m(t)$ , and  $X_e(t)$  after a change in the input signal  $I$ . The details of this simulation can be found in Appendix B. The results in Figure 3 show that active RAS, RAF, and MEK revert to their original values after an initial response, while the equilibrium value of active ERK changes. Apparently, the equilibrium concentrations  $X_s$ ,  $X_r$ , and  $X_m$  perfectly adapt to the input signal  $I$ , while the equilibrium concentration  $X_e$  of active ERK depends on the input signal  $I$ .

## 5.2 Graphical representations

We consider graphical representations of the protein signalling pathway. By using the natural labelling, we construct the dynamic bipartite graph in Figure 7(a) from the first-order differential equations, with the input signal  $I$  included. The associated dynamic causal ordering graph is given in Figure 7(b).

Under saturation conditions, the equilibrium equations  $f_s$ ,  $f_r$ ,  $f_m$ , and  $f_e$  obtained by setting equations (25), (26), (27), and (29), respectively, to zero have the bipartite structure in Figure 7(c). Note that there is no edge  $(f_e - v_e)$  in the equilibrium bipartite graph because  $X_e(t)$  does not appear in the approximation (29) of

<sup>23</sup> For simplicity, we omitted the feedback mechanism through RAF kinase inhibitor protein (RKIP). In the differential equation for the activated MEK, we therefore discarded the dependence on RKIP. The goal here is not to give the most realistic model but to elucidate the phenomenon of perfect adaptation and the causal interpretation of the equilibrium Markov ordering graph for perfectly adapted dynamical systems.

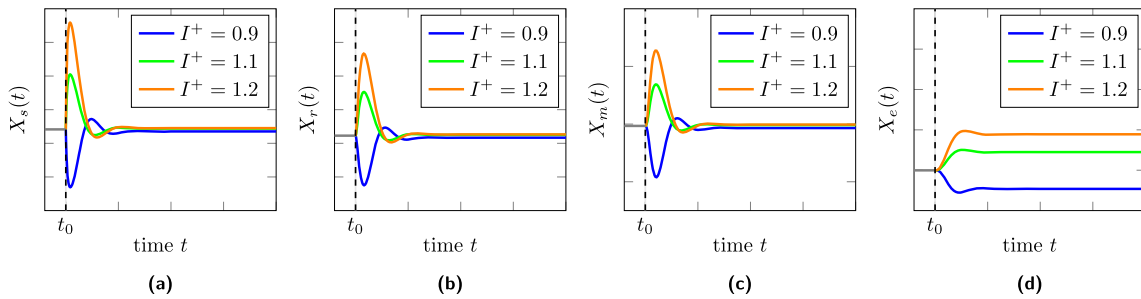
(28). The associated equilibrium causal ordering graph is shown in Figure 7(d), where the cluster  $\{I\}$  is added with an edge towards the cluster  $\{v_e, f_s\}$  because  $I$  appears in equation (25) and in no other equations. So far we have treated all symbols in equations (25)–(28) as deterministic parameters. Let  $w_s, w_r, w_m,$  and  $w_e$  represent independent exogenous random variables appearing in the equilibrium equations  $f_s, f_r, f_m,$  and  $f_e,$  respectively. After adding them to the causal ordering graph with edges to their respective clusters, we construct the equilibrium Markov ordering graph for the equilibrium distribution in Figure 7(e).

There is a directed path from the input vertex  $I$  to  $v_s, v_r,$  and  $v_m$  in the dynamic causal ordering graph (Figure 7(b)), while the equilibrium causal ordering graph has no directed paths from  $I$  to  $v_s, v_r,$  or  $v_m$  (Figure 7(d)). Under Assumption 1, we can apply Theorem 1 to conclude that  $X_s, X_r,$  and  $X_m$  will perfectly adapt to a persistent change in the input signal  $I$ . This is in line with what we observed in the simulations (Figure 6).

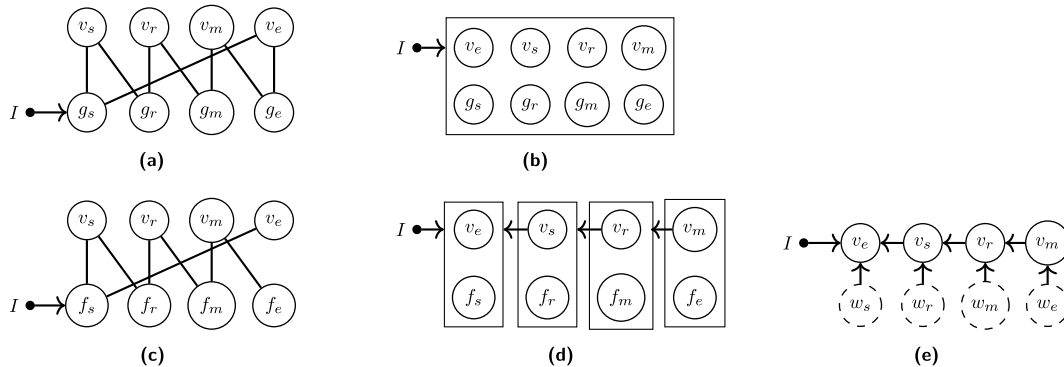
The  $d$ -separations in the equilibrium Markov ordering graph (Figure 7(e)) imply conditional independences between the corresponding variables at equilibrium. For example, from the graph in Figure 7(e), we read off the following (implied) conditional independences:

$$I \perp\!\!\!\perp v_r, \quad I \perp\!\!\!\perp v_m, \quad v_e \perp\!\!\!\perp v_m \mid v_r.$$

We verified that these conditional independences indeed appear in the simulated equilibrium distribution of the model (see Appendix C for details).



**Figure 6:** Perfect adaptation in the model for the RAS-RAF-MEK-ERK signalling pathway. After an initial response to a change of the input signal from value  $I^- = 1.0$  to a value  $I^+$  at time  $t = t_0$ , the equilibrium concentrations of active RAS, RAF, and MEK revert to their original values, while the concentration of active ERK changes. Shown are concentrations of active RAS (a), RAF (b), MEK (c) and ERK (d) against time.



**Figure 7:** Five graphs associated with the protein signalling pathway model under saturation conditions where indices  $s, r, m,$  and  $e$  correspond to concentrations of active RAS, RAF, MEK, and ERK, respectively. (a) Dynamic bipartite graph. (b) Dynamic causal ordering graph. (c) Equilibrium bipartite graph. (d) Equilibrium causal ordering graph. (e) Equilibrium Markov ordering graph.

### 5.3 Inhibiting the activity of MEK

A common biological experiment that is used to study protein signalling pathways is the use of an inhibitor that decreases the activity of a protein on the pathway. Such an inhibitor slows down the rate at which the active protein is able to activate another protein. Here, we consider inhibition of MEK activity. We can model this as a change of the parameters of the differential equations in which  $X_m(t)$  appears. We can interpret this experiment as a soft intervention on differential equation  $g_e$  in the dynamic model and on equation  $f_e$  at equilibrium, decreasing the rate  $k_{me}$  at which ERK is activated. Since there is a directed path from  $f_e$  to  $v_m$ ,  $v_r$ ,  $v_s$ , and  $v_e$  in the equilibrium causal ordering graph in Figure 7(d), we expect that a change in an input signal  $I_e$  on  $f_e$  (e.g., a change in the parameter  $k_{me}$  in the case of the MEK inhibition) affects the equilibrium concentrations of active MEK, RAF, RAS, and ERK.

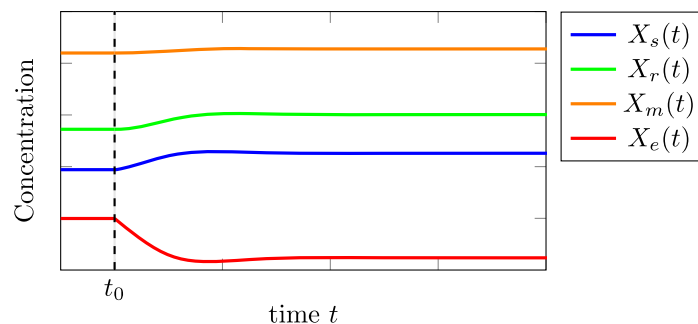
We assessed the effect of decreasing the activity of MEK on the equilibrium concentrations of RAS, RAF, MEK, and ERK. To that end, we simulated the perfectly adapted model (with parameters as described in Appendix B, in particular,  $k_{me} = 1.0$ ) until it reached equilibrium. We then decreased the parameter that controls the activity of MEK to  $k_{me} = 0.98$ . The recorded responses of the concentrations of active RAS, RAF, MEK, and ERK are displayed in Figure 8. From this, we confirm our prediction that inhibition of MEK activity affects the equilibrium concentrations of RAS, RAF, MEK, and ERK. A qualitative aspect of this change that one cannot read off from the graph is that the MEK inhibition *increases* (rather than decreases) the concentrations of active MEK, RAF, and RAS according to the model for this choice of the parameters.

Note that RAS, RAF, and MEK are non-descendants of ERK in the equilibrium Markov ordering graph in Figure 7(e), so that under the assumptions in Theorem 2, we could actually use this experiment to detect perfect adaptation in the protein pathway.

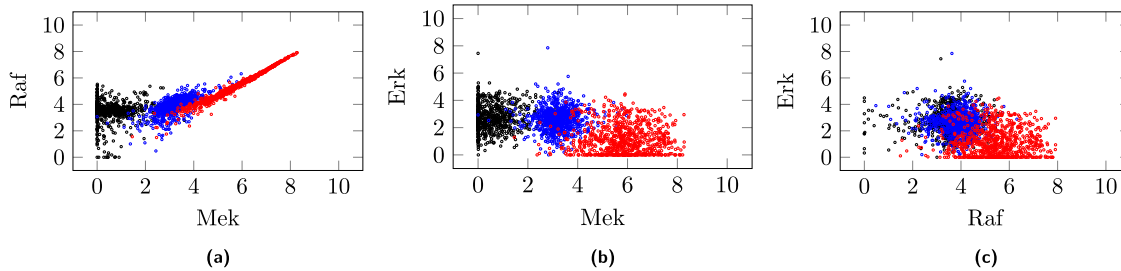
### 5.4 Testing model predictions on real-world data

In this subsection, we verify some of the predictions of the model we obtained in Sections 5.2 and 5.3 on real-world data. We will compare with predictions of the causal Bayesian network model proposed by ref. [18].

Figure 9 shows scatter plots for the (logarithms of) the expressions of active RAF, MEK, and ERK in the multivariate single-cell protein expression dataset that was used in ref. [18], for three (out of 14) different experimental conditions. The baseline condition (in blue) is the one where the cells were treated with anti-CD3 and anti-CD28, activators of the RAS-RAF-MEK-ERK signalling cascade. In another condition (in red), the cells were additionally exposed to U0126, a known inhibitor of MEK activity. By inspecting the scatter plots, we get a quick visual check of some of the predictions of the model. In particular, these plots clearly show that inhibition of MEK activity by administering U0126 results in an increase in the concentrations of



**Figure 8:** Simulation of the response of the concentrations of active RAS, RAF, MEK, and ERK after inhibition of the activity of MEK. The system starts out in equilibrium with  $k_{me} = 1.0$ . The concentrations of RAS, RAF, MEK, and ERK are recorded after the parameter controlling MEK activity is decreased to  $k_{me} = 0.98$  from  $t = t_0$  on.



**Figure 9:** Scatter plots of the logarithms of active RAF, MEK, and ERK concentrations for the data in ref. [18]. The blue circles correspond to cells treated only with anti-CD3 and anti-CD28, which activate the signalling cascade. The red circles correspond to cells treated with anti-CD3, anti-CD28, and in addition, the MEK-activity inhibitor U0126. The inhibition of MEK results in an increase of MEK and RAF, whereas ERK is reduced. The black circles correspond to cells treated with  $\beta$ 2cAMP (but not anti-CD3 and anti-CD28), which seems to affect MEK, but leaves RAF and ERK invariant.

active RAF and active MEK and a reduction in the concentration of active ERK. Furthermore, we clearly see a strong dependence between RAF and MEK (in both experimental conditions), but there is no discernible dependence between RAF and ERK or between MEK and ERK (in either experimental condition). In the light of the supposedly direct effect of MEK on ERK, it is surprising that the data show no significant dependence between the two.<sup>24</sup> This apparent “faithfulness violation” is problematic for constraint-based causal discovery methods, as they will typically not identify the causal relation between MEK and ERK.

According to ref. [18], the biological consensus (at the time) was that there is a signalling pathway from RAF to MEK to ERK.<sup>25</sup> They propose to model this pathway as a causal Bayesian network  $\text{RAF} \rightarrow \text{MEK} \rightarrow \text{ERK}$ , and this is also the structure identified by their causal discovery algorithm. That model predicts that inhibiting the MEK activity can only affect ERK. In Table 1, we summarize some of the observations made using the scatter plots in Figure 9, and whether they are in line with the predictions of the models (our perfectly adaptive model on the one hand, and the causal Bayesian network model on the other hand). We conclude that the predictions of the perfectly adaptive model are more in agreement with the data than those of the causal Bayesian network model.

Still, the perfectly adaptive model does not explain all aspects of the data. For example, the effects of the  $\beta$ 2cAMP stimulation (black circles in Figure 9) are hard to explain with either model.  $\beta$ 2cAMP is an AMP analogue that is supposed to activate the RAS-RAF-MEK-ERK cascade by promoting active RAS. It seems counterintuitive that this would lead to a strong reduction of active MEK in comparison to the activation of the cascade by means of anti-CD3 and anti-CD28, while leading to the same levels of active RAF and ERK. For completeness, in Table 2, we have indicated for all 12 perturbations in ref. [18] the effects on the levels of active RAF, MEK, and ERK. It appears questionable whether a simple causal model (such as the perfectly adaptive model) could account for all the observed perturbation effects.

## 5.5 Caveats for causal discovery

Experiments in which the protein signalling network is perturbed in various ways are of crucial importance to obtaining a causal understanding of the system. While very sophisticated causal discovery algorithms are

<sup>24</sup> In addition, ERK levels are considerably lower than those of RAF and MEK. This might be something specific to this experimental setting (perhaps there was too much time between stimulation and measurement to see the ERK response), as in other experiments high ERK levels and strong correlations with MEK have been observed [53]. In the setting of the experiment of [18], the G06976 treatment condition shows that ERK levels can actually get as high as those of RAF and MEK.

<sup>25</sup> Nowadays, the biological consensus appears to be that RAF activates MEK, MEK activates ERK, and that it is very likely that there is negative feedback from ERK to RAF, although the molecular pathway of this feedback remains unknown [54].

**Table 1:** For various observations in the data of [18], we indicate whether they are predicted by the causal Bayesian network model  $\text{RAF} \rightarrow \text{MEK} \rightarrow \text{ERK}$  proposed by ref. [18] and by our perfectly adaptive equilibrium model (Section 5.2)

Observation	Causal Bayesian network model	Perfectly adaptive model
RAF and MEK are dependent in both conditions	+	+
MEK and ERK are independent in both conditions	-	-
Inhibition of MEK activity affects active RAF	-	+
Inhibition of MEK activity affects active MEK	-	+
Inhibition of MEK activity affects active ERK	+	+

**Table 2:** Qualitative effects of reagents on the measured abundances of active RAF, MEK, and ERK, as read off from the data in ref. [18]

Condition	Reagents	RAF	MEK	ERK
1	Anti-CD3 + anti-CD28	.....Baseline .....		
3	Anti-CD3 + anti-CD28 + AKT inhibitor	NA	NA	0
4	Anti-CD3 + anti-CD28 + G06976	++	++	++
5	Anti-CD3 + anti-CD28 + Psitectorigenin	0	0	--
6	Anti-CD3 + anti-CD28 + U0126	++	++	(+)
7	Anti-CD3 + anti-CD28 + LY294002	NA	NA	0
8	PMA	-	0	0
9	$\beta$ 2cAMP	(-)	--	(+)
2	Anti-CD3 + anti-CD28 + ICAM-2	.....Baseline .....		
10	Anti-CD3 + anti-CD28 + ICAM-2 + AKT inhibitor	0	0	--
11	Anti-CD3 + anti-CD28 + ICAM-2 + G06976	++	++	++
12	Anti-CD3 + anti-CD28 + ICAM-2 + Psitectorigenin	0	0	--
13	Anti-CD3 + anti-CD28 + ICAM-2 + U0126	0	(+)	--
14	Anti-CD3 + anti-CD28 + ICAM-2 + LY294002	0	0	--

Legend: -- strong decrease, - decrease, (-) slight decrease, 0 no change, (+) slight increase, + increase, and ++ strong increase. Most researchers only use conditions 1–9 for causal discovery. The data appear to contain an error: RAF and MEK values are identical for the first 848 samples in conditions 3 and 7; therefore, we disregarded those values.

available, we will here illustrate the key concepts by means of applying one of the simplest causal discovery algorithms based on conditional independences to equilibrium data from the model. We simulate the system in two different conditions, a baseline and a condition where the activity of MEK has been inhibited.

Consider observational equilibrium data from the protein signalling pathway model and also experimental equilibrium data from a setting where the MEK activity is inhibited. We introduce a *context* variable  $C$  that in this case simply indicates for each sample whether a MEK inhibition was applied. Because the decision whether to apply the MEK inhibitor to a sample occurs *before* the measurement, the equilibrium concentrations measured afterwards cannot causally affect this decision. Hence,  $C$  is an exogenous variable, i.e., it is not caused by other observed variables. This motivates the use of the LCD algorithm, which requires the existence of such a variable. For the MEK inhibition, the context variable represents a (soft) intervention on the equation  $f_e$  in the equilibrium causal ordering graph in Figure 7(d). The equilibrium Markov ordering graph that includes the context variable  $C$  (but where the independent exogenous random variables have been marginalized out) is given in Figure 10(a). To construct this graph, the context variable  $C$  is first added to the equilibrium causal ordering graph in Figure 7(d) as a singleton cluster with an edge towards the cluster  $\{v_m, f_e\}$ . The equilibrium Markov ordering graph is then constructed from the resulting directed cluster graph in the usual way. From Figure 10(a), we can read off (conditional) independences to find the LCD triples that are implied by the equilibrium equations of the model. We find that our model implies the following LCD triples:  $(C, v_m, v_r)$ ,  $(C, v_m, v_s)$ ,  $(C, v_m, v_e)$ ,  $(C, v_r, v_s)$ ,  $(C, v_r, v_e)$ , and  $(C, v_s, v_e)$ . We verified this by explicit simulation (details in Appendix C).

**Table 3:** Results of conditional independence tests on the (log-transformed) protein expression data of [18]

(Conditional) independence tested	$p$ Value
$v_r \perp v_m \mid C$	$<4.6 \times 10^{-324}$
$v_r \perp v_e \mid C$	0.79
$v_m \perp v_e \mid C$	0.35
$v_r \perp C$	$4.0 \times 10^{-172}$
$v_m \perp C$	$3.7 \times 10^{-245}$
$v_e \perp C$	$3.5 \times 10^{-138}$
$v_r \perp C \mid v_m$	$2.7 \times 10^{-8}$
$v_m \perp C \mid v_r$	$8.2 \times 10^{-168}$
$v_r \perp C \mid v_e$	$4.3 \times 10^{-210}$
$v_m \perp C \mid v_e$	$<4.6 \times 10^{-324}$
$v_e \perp C \mid v_m$	$1.5 \times 10^{-134}$
$v_e \perp C \mid v_r$	$2.6 \times 10^{-156}$

Specifically, we report the  $p$ -values of Kendall's test for partial correlation.



**Figure 10:** Equilibrium Markov ordering graphs of the protein signalling pathway with the context variable  $C$  included, corresponding to two hypothetical models. This context variable indicates whether a cell was treated with a MEK inhibitor or not. (a) Perfectly adaptive model (with feedback). (b) Causal Bayesian network model.

In particular, one of the LCD triples we obtained is  $(C, v_m, v_r)$ , whereas the triple  $(C, v_r, v_m)$  does not qualify as such. According to the standard interpretation of causal discovery algorithms, this would imply that MEK causes RAF rather than the other way around (see also Section 2.1), a surprising finding in the light of the textbook treatments of the RAS-RAF-MEK-ERK signalling pathway, and the “consensus network” according to ref. [18]. The equilibrium Markov ordering graph corresponding to the causal Bayesian network model of [18] is shown in Figure 10(b). As one can see from Figure 10(b), the causal Bayesian network model implies no LCD triples at all.

Apparently, the causal relationship  $\text{RAF} \rightarrow \text{MEK}$  appears to be reversed in the LCD triples found in the equilibrium data of the perfectly adaptive model. Similar observations of apparent “causal reversals” in protein interaction networks have been observed more often, see also [9,45–48]. We conclude that the mechanism of perfect adaptation provides one possible theoretical explanation of what might seem at first sight to be an incorrect reversal of a causal edge.

We investigated which of these LCD patterns can be found in the real-world data. In Table 3, we list the  $p$ -values of conditional independence tests applied to the data of [18].<sup>26</sup> By using a reasonable critical level for the test (say  $\alpha = 0.01$ ), we do not find any LCD pattern. Yet, the conditional dependence of RAF on the context variable when conditioning on MEK is much weaker than the conditional dependence of MEK on the context variable when conditioning on RAF. Strictly speaking, no conclusions should be drawn from this, but it does seem to suggest that also here the data are more in line with the perfectly adaptive model than with the causal Bayesian network model; indeed, if the adaptation is imperfect, for example, because

<sup>26</sup> We made use of Kendall's test for partial correlation because the implementation in the R package `ppcor` can deal with ties, while that for Spearman's test cannot.



the saturation conditions only hold approximately, one would expect to see a weak conditional dependence  $v_r \not\perp C \mid v_m$ .

## 6 Discussion and conclusion

Perfect adaptation is the phenomenon that a dynamical system initially responds to a change of input signal but reverts back to its original value as the system converges to equilibrium. We used the technique of causal ordering to obtain sufficient graphical conditions to identify perfect adaptation in a dynamical system described by a combination of equations and first-order differential equations. To represent the structure of the (non-equilibrium) dynamical system, we introduced the notions of the *dynamic bipartite graph* and the corresponding *dynamic causal ordering graph* obtained by the causal ordering algorithm. Moreover, we showed how perfect adaptation can be detected in equilibrium observational and experimental data for soft interventions with known targets. We illustrated our ideas on a variety of dynamical models and corresponding equilibrium equations. We believe that the methods presented in this work provide a useful tool for the characterization of a large class of dynamical systems that are able to achieve perfect adaptation and for the automated analysis of the behaviour of certain perfectly adapted dynamical systems.

In all examples that we discussed, the technique of causal ordering revealed the structure of a *given set* of equations. In some cases, more structure can be revealed by first rewriting the equations before applying the causal ordering algorithm. In Appendix D, we analyze a dynamical system describing a basic enzyme reaction, for which rewriting of the equilibrium equations reveals more structure (and hence yields a stronger Markov property). In Appendix E, we analyze the IFFLP network. We observe that a nonlinear transformation of the variables (and rewriting the equations correspondingly) reveals more structure and hence yields more conditional independences and less causal relations amongst the transformed variables. The further development of methods to analyze perfectly adapted dynamical systems that do not satisfy the conditions of Theorem 1 remains a challenge for future work. Also, the question of how one can more generally discover causal structure using nonlinear transformations of the variables is an interesting topic for future research that relates to what is nowadays known as “causal representation learning” in machine learning [55].

We also investigated the consequences of the phenomenon of perfect adaptation for causal discovery. We demonstrated that for perfectly adapted dynamical systems, the output of existing constraint-based causal discovery algorithms applied to equilibrium data may appear counterintuitive and at odds with our understanding of the mechanisms that drive the system. As we have illustrated in this work, careful application of the causal ordering algorithm enables a better theoretical understanding of these phenomena.

We applied our approach to a model for a well-known protein signalling pathway and tested the model’s predictions both in simulations and on real-world protein expression data. The challenges for causal discovery that are encountered in non-linear dynamical systems with feedback loops, possibly leading to context-specific perfectly adaptive behaviour, are substantial. If the behaviour of the model that we analyzed in Section 5 is representative of that of actual systems occurring *in vitro* and *in vivo*, then it seems unlikely that existing causal discovery methods based on causal Bayesian networks will lead to reasonable results. These observations further motivate the development of causal discovery algorithms based on bipartite graphical representations that would be more widely applicable than the existing ones based on causal Bayesian networks or simple SCMs.

**Acknowledgements:** We thank the reviewers and the editor for their constructive comments, which helped us improve our work.

**Funding information:** This work was supported by the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement 639466).

**Author contributions:** The idea to use causal ordering to study perfectly adapted dynamical systems was due to TB, as well as the development of the theoretical results. The notions of the dynamic bipartite graph and the dynamic causal ordering graph were proposed by JMM. Simulations were conducted by TB. Analysis of the protein signalling data and preparation of the manuscript was done by both authors. The SDC approach was applied for the sequence of authors. All authors have accepted responsibility for the entire content of this article and approved its submission.

**Conflict of interest:** JMM is a member of the Editorial Board of Journal of Causal Inference and was not involved in the review process of this article.

**Data availability statement:** The simulated datasets can be reproduced with the R code provided at <https://bitbucket.org/jorism/jci2023paper.git> as free and open source software. The protein expression dataset of [18] analyzed in Section 5.4 is publicly available as Supplementary Material to ref. [18] at [https://www.science.org/doi/suppl/10.1126/science.1105809/suppl\\_file/sachs.som.datasets.zip](https://www.science.org/doi/suppl/10.1126/science.1105809/suppl_file/sachs.som.datasets.zip).

## References

- [1] Cooper GF. A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Min Knowl Discov.* 1997;1:203–24.
- [2] Richardson TS, Spirtes P. Automated discovery of linear feedback models. In: Gregory F. Cooper, Clark Glymour. *Computation, Causation and Discovery.* London, England: MIT Press; 1999. p. 253–302.
- [3] Pearl J. *Causality: models, reasoning, and inference.* Cambridge, UK: Cambridge University Press; 2000.
- [4] Spirtes P, Glymour C, Scheines R. *Causation, prediction, and search.* Cambridge, Massachusetts: MIT Press; 2000.
- [5] Zhang J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artif Intell.* 2008;172:1873–96.
- [6] Hyttinen A, Eberhardt F, Hoyer PO. Learning linear cyclic causal models with latent variables. *J Mach Learn Res.* 2012;13(1):3387–439.
- [7] Colombo D, Maathuis MH, Kalisch M, Richardson TS. Learning high-dimensional directed acyclic graphs with latent and selection variables. *Ann Stat.* 2012;40:294–321.
- [8] Forré P, Mooij JM. 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.* p. 269–78.
- [9] Mooij JM, Magliacane S, Claassen T. Joint causal inference from multiple contexts. *J Mach Learn Res.* 2020;21:1–108.
- [10] Mooij JM, Claassen T. Constraint-based causal discovery using partial ancestral graphs in the presence of cycles. In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI-20).* vol. 124. *Proceedings of Machine Learning Research (PMLR); 2020.* p. 1159–68.
- [11] Simon HA. Causal ordering and identifiability. In: *Studies in econometric methods.* New York: John Wiley & Sons; 1953. p. 49–74.
- [12] Blom T, van Diepen MM, Mooij JM. Conditional independences and causal relations implied by sets of equations. *J Mach Learn Res.* 2021;22(178):1–62.
- [13] Iwasaki Y, Simon HA. Causality and model abstraction. *Artif Intell.* 1994;67:143–94.
- [14] Dash D. Restructuring dynamic causal systems in equilibrium. In: *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS 2005); 2005.* p. 81–8.
- [15] De Boer RJ. Which of our modeling predictions are robust? *PLOS Comput Biol.* 2012;8:e1002593.
- [16] Blom T, Mooij JM. Robustness of model predictions under extension. In: Cussens J, Zhang K, editors. *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence (UAI-22).* vol. 180 of *Proceedings of Machine Learning Research.* PMLR; 2022. p. 213–22.
- [17] Ma W, Trusina A, El-Samad H, Lim WA, Tang C. Defining network topologies that can achieve biochemical adaptation. *Cell.* 2009;138:760–73.
- [18] Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. *Science.* 2005;308:523–9.

- [19] Shin SY, Rath O, Choo SM, Fee F, McFerran B, Kolch W, et al. Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-Mek-Erk signal transduction pathway. *J Cell Sci.* 2009;122:425–35.
- [20] Lauritzen SL, Dawid AP, Larsen BN, Leimer HG. Independence properties of directed Markov fields. *Networks.* 1990;20:491–505.
- [21] Lacerda G, Spirtes P, Ramsey J, Hoyer PO. Discovering cyclic causal models by independent components analysis. In: *Proceedings of the 24th Annual Conference on Uncertainty in Artificial Intelligence (UAI-08)*; 2008. p. 1159–68.
- [22] Strobl EV. A constraint-based algorithm for causal discovery with cycles, latent variables and selection bias. *Int J Data Sci Anal.* 2019;8:33–56.
- [23] Bongers S, Forré P, Peters J, Mooij JM. Foundations of structural causal models with cycles and latent variables. *Ann Stat.* 2021;49(5):2885–915.
- [24] Spirtes P. Directed cyclic graphical representations of feedback models. In: *Proceedings of the 11th Annual Conference on Uncertainty in Artificial Intelligence (UAI-1995)*; 1995. p. 491–8.
- [25] Forré P, Mooij JM. Markov properties for graphical models with cycles and latent variables. 2017. arXiv:1710.08775v1 [math.ST]. <https://arxiv.org/abs/1710.08775v1>.
- [26] Richardson TS. *Models of feedback: interpretation and discovery.* PhD dissertation. Carnegie-Mellon University, 1996.
- [27] Nayak P. *Automated modeling of physical systems.* Berlin: Springer-Verlag; 1995.
- [28] Pothen A, Fan CJ. Computing the block triangular form of a sparse matrix. *ACM Trans Math Softw.* 1990;16:303–24.
- [29] Gonçalves B, Porto F. A note on the complexity of the causal ordering problem. *Artif Intell.* 2016;238:154–65.
- [30] Fisher FM. A correspondence principle for simultaneous equation models. *Econometrica.* 1970;38(1):73–92.
- [31] Voortman M, Dash D, Druzdzel MJ. Learning why things change: the difference-based causality learner. In: *Proceedings of the Twenty-Sixth Annual Conference on Uncertainty in Artificial Intelligence (UAI-10)*; 2010. p. 641–50.
- [32] Sokol A, Hansen NR. Causal interpretation of stochastic differential equations. *Electr J Probabil.* 2014;19:1–24.
- [33] Rubenstein PK, Bongers S, Schölkopf B, Mooij JM. From deterministic ODEs to dynamic structural causal models. In: *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*; 2018. p. 114–23.
- [34] Bongers S, Blom T, Mooij JM. Causal modeling of dynamical systems. 2022 Mar. Preprint. arXiv:1803.08784v4 [cs.AI]. <https://arxiv.org/abs/1803.08784v4>.
- [35] Mogensen SW, Malinsky D, Hansen NR. Causal learning for partially observed stochastic dynamical systems. In: *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*; 2018. p. 350–60.
- [36] Mooij JM, Janzing D, Schölkopf B. 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)*; 2013. p. 440–8.
- [37] Lauritzen SL, Richardson TS. Chain graph models and their causal interpretations. *J R Stat Soc Ser B (Stat Methodol).* 2002;64:321–61.
- [38] Mooij JM, Janzing D, Heskes T, Schölkopf B. On causal discovery with cyclic additive noise models. *Adv Neural Inform Process Syst (NIPS 2011).* 2011;24:639–47.
- [39] Blom T, Bongers S, Mooij JM. Beyond structural causal models: causal constraints models. In: *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference (UAI-19)*. vol. 115 of *Proceedings of Machine Learning Research*; 2020. p. 585–94.
- [40] Dawid AP. Beware of the DAG! In: *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*. vol. 6 of *Proceedings of Machine Learning Research*; 2010. p. 59–86.
- [41] Araujo RP, Liotta LA. The topological requirements for robust perfect adaptation in networks of any size. *Nature Commun.* 2018;9:29717141.
- [42] Muzzey D, Gómez-Urbe CA, Mettetal JT, van Oudenaarden A. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell.* 2009;138:160–71.
- [43] Ferrell JE. Perfect and near-perfect adaptation in cell signaling. *Cell Sys.* 2016;2:62–67.
- [44] Krishnan J, Floros I. Adaptive information processing of network modules to dynamic and spatial stimuli. *BMC Syst Biol.* 2019;13:30866946.
- [45] Triantafillou S, Lagani V, Heinze-Deml C, Schmidt A, Tegner J, Tsamardinos I. Predicting causal relationships from biological data: applying automated causal discovery on mass cytometry data of human immune cells. *Sci Rep.* 2017;7:12724.
- [46] Mooij JM, Heskes T. Cyclic causal discovery from continuous equilibrium data. In: *Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13)*. Arlington, Virginia: AUAI Press; 2013. p. 431–9.
- [47] Ramsey J, Andrews B. FASK with interventional knowledge recovers edges from the Sachs model. 2018; arXiv:1805.03108 [q-bio.MN]. <https://arxiv.org/abs/1805.03108>.
- [48] Boeken PA, Mooij JM. A Bayesian nonparametric conditional two-sample test with an application to local causal discovery. In: *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI-21)*. vol. 161 of *Proceedings of Machine Learning Research*; 2021. p. 1565–75.
- [49] Blom T, Klimovskaia A, Magliacane S, Mooij JM. An upper bound for random measurement error in causal discovery. In: *Proceedings of the 34th Annual Conference on Uncertainty in Artificial Intelligence (UAI-18)*; 2018. p. 570–9.
- [50] Coddington EA, Levinson N. *Theory of ordinary differential equations.* New York: McGraw-Hill; 1955.

- [51] Meek C. Strong completeness and and faithfulness in Bayesian networks. In: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95); 1995. p. 411–8.
- [52] Sachs K, Itani S, Fitzgerald J, Schoeberl B, Nolan GP, Tomlin CJ. Single timepoint models of dynamic systems. *Interface Focus*. 2013;3:24511382.
- [53] Filippi S, Barnes CP, Kirk PD, Kudo T, Kunida K, McMahon SS, et al. Robustness of MEK-ERK dynamics and origins of cell-to-cell variability in MAPK signaling. *Cell Reports*. 2016 Jun;15:2524–35.
- [54] Fritsche-Guenther R, Witzel F, Sieber A, Herr R, Schmidt N, Braun S, et al. Strong negative feedback from Erk to Raf confers robustness to MAPK signalling. *Mol Syst Biol*. 2011;7(1):489.
- [55] Chalupka K, Eberhardt F, Perona P. Causal feature learning: an overview. *Behaviormetrika*. 2017;44:137–67.
- [56] Murray JD. *Mathematical biology I: an introduction*. 3rd edition. New York: Springer-Verlag; 2002.
- [57] Dash D, Druzdel MJ. A note on the correctness of the causal ordering algorithm. *Artif Intell*. 2008;172:1800–8.

## Appendix

### A Proof of Theorem 2

**Theorem 2.** Consider a set of first-order dynamical equations in canonical form for variables  $V$ , satisfying the conditions of Theorem 1, with equilibrium equations  $F$  under the natural labelling. Consider a soft intervention targeting an equation  $f_i \in F$ . Assume that the system is uniquely solvable with respect to the equilibrium causal ordering graph both before and after the intervention and that the intervention alters the equilibrium distribution of all descendants of  $f_i$  in the equilibrium causal ordering graph. If either

- (1) the soft intervention does not change the equilibrium distribution of  $X_i$  or
  - (2) the soft intervention alters the equilibrium distribution of a variable corresponding to a non-descendant of  $v_i$  in the equilibrium Markov ordering graph,
- (or both), then the system is capable of perfect adaptation.

**Proof.** If condition 1 holds there is no directed path from  $f_i$  to  $v_i$  in the equilibrium causal ordering graph, by the assumption that the soft intervention on  $f_i$  changes the equilibrium distribution of all its descendants. By definition of the dynamic bipartite graph, there is a directed path from  $g_i$  to  $v_i$  in the dynamic causal ordering graph, because  $g_i$  and  $v_i$  end up in the same cluster (note that this follows by using the natural labelling as perfect matching and the result that the causal ordering graph does not depend on the chosen perfect matching [12]). It follows from Theorem 1 that  $X_i$  perfectly adapts to an input signal  $I_{f_i}$  on  $f_i$  (i.e., a soft intervention targeting  $\dot{X}_i(t)$  and thus the equilibrium equation  $f_i$ ).

Suppose that condition 1 does not hold while condition 2 does hold. By Theorem 4 in ref. [12] (which roughly states that the presence of a causal effect at equilibrium implies the presence of a corresponding directed path in the equilibrium causal ordering graph) we have that  $f_i$  is an ancestor of  $v_i$  and some  $v_h$  in the equilibrium causal ordering graph, while  $v_i$  is not an ancestor of  $v_h$  in the equilibrium Markov ordering graph. For a perfect matching  $M$  of the equilibrium bipartite graph, let  $v_j = M(f_i)$ . Then  $v_j$  is in the same cluster as  $f_i$  in the equilibrium causal ordering graph by construction. Note that  $j = i$  would give a contradiction, as then  $v_i$  would be an ancestor of  $v_h$  in the equilibrium Markov ordering graph. Suppose that the vertex  $f_j$ , which is associated with  $v_j$  through the natural labelling, is matched to a non-ancestor of  $v_j$  in the equilibrium causal ordering graph. Because of the edge  $(g_j - v_j)$  in the dynamic bipartite graph, it follows from Theorem 1 that  $X_j$  perfectly adapts to an input signal  $I_{f_j}$  on  $f_j$ . Therefore, the system is able to achieve perfect adaptation. Now suppose that  $f_j$  is matched to an ancestor  $v_k$  of  $v_j$  in the equilibrium causal ordering graph, and consider the vertex  $f_k$ . The previous argument can be repeated to show perfect adaptation for  $X_k$  is present when  $f_k$  is matched to a non-ancestor of  $v_k$  in the equilibrium causal ordering graph. Otherwise,  $f_k$  must be matched to an ancestor of  $v_k$ . Note that the ancestors of  $v_k$  are a subset of the ancestors of  $v_j$ , which in turn are a subset of the ancestors of  $v_i$ . In a finite system of equations,  $v_i$  has a finite set of ancestors and therefore we eventually find, by repeating our argument, a vertex  $f_m$  that cannot be matched to an ancestor of  $v_m$  because  $v_m$  has no ancestors that are not matched to one of the vertices  $f_i, f_j, f_k, \dots$  that were considered up to that point. Because  $f_m$  is matched to a non-ancestor, we then find that  $X_m$  perfectly adapts to an input signal on  $I_{f_m}$  as before.  $\square$

### B Simulation settings

For the simulations in Figures 3,6, and 8 of the model of a filling bathtub, the viral infection model, the reaction network with a feedback loop, and the protein pathway we used the settings outlined below. Since we only simulated a single response, we used constant values for the exogenous random variables as well.

**Filling bathtub.** First, we recorded the behaviour of the system for the parameters  $I_K = 1.2$ ,  $U_I = 5.0$ ,  $U_1 = 1.1$ ,  $U_2 = 1.0$ ,  $U_3 = 1.2$ ,  $U_4 = 1.0$ ,  $U_5 = 0.8$ , and  $g = 1.0$  until it reached equilibrium. We then changed the input parameter  $I_K$  to 0.8, 1.0, and 1.3 and recorded the response until the system reverted to equilibrium.

**Viral infection.** For the parameter settings  $I_\sigma = 1.6$ ,  $d_T = 0.9$ ,  $\beta = 0.9$ ,  $d_I = 0.3$ ,  $k = 1.5$ ,  $a = 0.1$ , and  $d_E = 0.25$ , we simulated the model until it reached equilibrium. We changed the input parameter  $I_\sigma$  to 1.1, 1.3, and 2.0 and recorded the response until equilibrium was reached.

**Reaction network.** We simulated the model until it reached equilibrium with parameters  $I = 1.5$ ,  $K_{IA} = 1.4$ ,  $K_{IA} = 0.8$ ,  $F_A = 1.1$ ,  $k_{F_A A} = 0.9$ ,  $K_{F_A A} = 1.2$ ,  $k_{CB} = 0.6$ ,  $K_{CB} = 0.0001$ ,  $F_B = 0.7$ ,  $k_{F_B B} = 0.7$ ,  $K_{F_B B} = 0.0001$ ,  $k_{AC} = 2.1$ ,  $K_{AC} = 1.5$ ,  $k_{BC} = 0.7$ , and  $K_{BC} = 0.6$ . The settings were chosen in such a way that the saturation conditions  $(1 - X_B(t)) \gg K_{CB}$  and  $X_B(t) \gg K_{F_B B}$  were satisfied. We then changed the input signal to 0.25, 1.0, and 10.0 and recorded the response.

**Protein pathway.** The parameter settings of the simulation were  $I = 1.0$ ,  $k_{I_S} = 1.0$ ,  $T_S = 1.0$ ,  $K_{I_S} = 1.0$ ,  $K_e = 1.2$ ,  $F_S = 1.0$ ,  $k_{F_S S} = 1.0$ ,  $K_{F_S S} = 0.9$ ,  $k_{S_r} = 1.0$ ,  $K_{S_r} = 1.0$ ,  $T_r = 1.0$ ,  $F_r = 0.3$ ,  $k_{F_r r} = 1.0$ ,  $K_{F_r r} = 0.8$ ,  $k_{r_m} = 1.0$ ,  $K_{r_m} = 0.9$ ,  $T_m = 1.0$ ,  $F_m = 0.2$ ,  $k_{F_m m} = 1.0$ ,  $K_{F_m m} = 1.2$ ,  $k_{m_e} = 1.0$ ,  $K_{m_e} = 0.0001$ ,  $T_e = 1.0$ ,  $F_e = 0.7$ ,  $k_{F_e e} = 1.2$ , and  $K_{F_e e} = 0.0001$ . This ensured that the saturation conditions  $(T_e - X_e(t)) \gg K_{m_e}$  and  $X_e(t) \gg K_{F_e e}$  were satisfied. For Figure 6, we changed the input signal  $I$  to 0.9, 1.1, 1.2, respectively, after the system had reached equilibrium, and continued the simulation. For Figure 8, we reduced the parameter value of  $k_{m_e}$  to 0.98 at some point in time after the system had equilibrated.

The same qualitative behaviour as reported here can be observed for a range of parameter values. The behaviour of the protein pathway is rather complex; in particular,  $X_e(t)$  may show a switch-like behaviour (either taking on a value close to 0, or close to  $T_e$ ), and therefore, some parameter tuning is required if one wants to ensure that the assumed saturation conditions  $((T_e - X_e(t)) \gg K_{m_e}$  and  $X_e(t) \gg K_{F_e e})$  hold.

## C Conditional independences and causal discovery

The equilibrium Markov ordering graph in Figure 7(e) was derived from the equilibrium equations of the protein pathway model under saturation conditions. From this, we can read off the following  $d$ -separations:

$$\begin{aligned}
& I \perp^d v_s, \quad I \perp^d v_s \mid v_r, \quad I \perp^d v_s \mid v_m, \quad I \perp^d v_s \mid \{v_r, v_m\}, \\
& I \perp^d v_r, \quad I \perp^d v_r \mid v_s, \quad I \perp^d v_r \mid v_m, \quad I \perp^d v_r \mid \{v_s, v_m\}, \\
& I \perp^d v_m, \quad I \perp^d v_m \mid v_s, \quad I \perp^d v_m \mid v_r, \quad I \perp^d v_m \mid \{v_s, v_r\}, \\
& v_e \perp^d v_r \mid v_s, \quad v_e \perp^d v_r \mid \{v_s, v_m\}, \quad v_e \perp^d v_r \mid \{v_s, I\}, \quad v_e \perp^d v_r \mid \{v_s, v_m, I\}, \\
& v_e \perp^d v_m \mid v_s, \quad v_e \perp^d v_m \mid v_r, \quad v_e \perp^d v_m \mid \{v_s, v_r\}, \\
& v_e \perp^d v_m \mid \{v_s, I\}, \quad v_e \perp^d v_m \mid \{v_r, I\}, \quad v_e \perp^d v_m \mid \{v_s, v_r, I\}, \\
& v_s \perp^d v_m \mid v_r, \quad v_s \perp^d v_m \mid \{v_e, v_r\}, \quad v_s \perp^d v_m \mid \{v_r, I\}, \quad v_s \perp^d v_m \mid \{v_e, v_r, I\}.
\end{aligned}$$

It is easy to check that the equilibrium equations and endogenous variables in this model are uniquely solvable with respect to the causal ordering graph. Therefore, the aforementioned  $d$ -separations imply conditional independences between the variables in the model, assuming the exogenous variables to be independent.

To test whether the predicted conditional independences hold when the system is at equilibrium, we ran the simulation  $n = 500$  times until it reached equilibrium and recorded the equilibrium concentrations  $X_s$ ,  $X_r$ ,  $X_m$ , and  $X_e$ . Parameters were as described in Appendix B, except that  $k_{I_S}$ ,  $k_{S_r}$ ,  $k_{r_m}$ , and  $k_{m_e}$  were all drawn randomly from a uniform distribution on  $[1.0, 1.1]$ , and the input signal  $I$  was drawn randomly from a

**Table A1:** The conditional independences in the simulation of the protein pathway (described in Section 5.2 and Appendix C) were assessed using Spearman’s rank correlations. With a  $p$ -value threshold of 0.01,  $d$ -separations with a separating set of size 0 or 1 coincide with conditional independence test results

Independence test	Correlation	$p$ -value	$d$ -separation
$I \perp\!\!\!\perp X_S$	-0.010	0.82	Yes
$I \perp\!\!\!\perp X_r$	-0.003	0.94	Yes
$I \perp\!\!\!\perp X_m$	+0.012	0.80	Yes
$I \perp\!\!\!\perp X_e$	+0.408	$<2.2 \times 10^{-16}$	No
$X_S \perp\!\!\!\perp X_r$	+0.976	$<2.2 \times 10^{-16}$	No
$X_S \perp\!\!\!\perp X_m$	+0.946	$<2.2 \times 10^{-16}$	No
$X_S \perp\!\!\!\perp X_e$	-0.891	$<2.2 \times 10^{-16}$	No
$X_r \perp\!\!\!\perp X_m$	+0.969	$<2.2 \times 10^{-16}$	No
$X_r \perp\!\!\!\perp X_e$	-0.868	$<2.2 \times 10^{-16}$	No
$X_m \perp\!\!\!\perp X_e$	-0.836	$<2.2 \times 10^{-16}$	No
$I \perp\!\!\!\perp X_S \mid X_r$	-0.033	0.46	Yes
$I \perp\!\!\!\perp X_S \mid X_m$	-0.066	0.14	Yes
$I \perp\!\!\!\perp X_r \mid X_S$	+0.032	0.48	Yes
$I \perp\!\!\!\perp X_r \mid X_m$	-0.059	0.19	Yes
$I \perp\!\!\!\perp X_m \mid X_S$	+0.066	0.14	Yes
$I \perp\!\!\!\perp X_m \mid X_r$	+0.060	0.18	Yes
$X_e \perp\!\!\!\perp X_r \mid X_S$	+0.020	0.66	Yes
$X_e \perp\!\!\!\perp X_m \mid X_S$	+0.045	0.32	Yes
$X_e \perp\!\!\!\perp X_m \mid X_r$	+0.041	0.36	Yes
$X_S \perp\!\!\!\perp X_m \mid X_r$	-0.007	0.87	Yes
$I \perp\!\!\!\perp X_e \mid X_S$	+0.876	$7.9 \times 10^{-160}$	No
$I \perp\!\!\!\perp X_e \mid X_r$	+0.814	$2.0 \times 10^{-119}$	No
$I \perp\!\!\!\perp X_e \mid X_m$	+0.760	$3.5 \times 10^{-95}$	No
$I \perp\!\!\!\perp X_S \mid X_e$	+0.850	$3.2 \times 10^{-140}$	No
$I \perp\!\!\!\perp X_r \mid X_e$	+0.772	$8.6 \times 10^{-100}$	No
$I \perp\!\!\!\perp X_m \mid X_e$	+0.703	$1.5 \times 10^{-75}$	No
$X_e \perp\!\!\!\perp X_S \mid X_r$	-0.405	$3.6 \times 10^{-21}$	No
$X_e \perp\!\!\!\perp X_S \mid X_m$	-0.562	$7.8 \times 10^{-43}$	No
$X_e \perp\!\!\!\perp X_S \mid I$	-0.971	$2.0 \times 10^{-310}$	No
$X_e \perp\!\!\!\perp X_r \mid X_m$	-0.425	$2.4 \times 10^{-23}$	No
$X_e \perp\!\!\!\perp X_r \mid I$	-0.949	$1.0 \times 10^{-250}$	No
$X_e \perp\!\!\!\perp X_m \mid I$	-0.921	$3.9 \times 10^{-205}$	No
$X_S \perp\!\!\!\perp X_r \mid I$	+0.976	$<4.6 \times 10^{-324}$	No
$X_S \perp\!\!\!\perp X_r \mid X_e$	+0.900	$1.5 \times 10^{-181}$	No
$X_S \perp\!\!\!\perp X_r \mid X_m$	+0.745	$2.0 \times 10^{-89}$	No
$X_S \perp\!\!\!\perp X_m \mid I$	+0.946	$1.9 \times 10^{-245}$	No
$X_S \perp\!\!\!\perp X_m \mid X_e$	+0.807	$1.2 \times 10^{-115}$	No
$X_r \perp\!\!\!\perp X_m \mid I$	+0.969	$4.6 \times 10^{-305}$	No
$X_r \perp\!\!\!\perp X_m \mid X_e$	+0.894	$2.3 \times 10^{-175}$	No
$X_r \perp\!\!\!\perp X_m \mid X_S$	+0.652	$1.1 \times 10^{-61}$	No

uniform distribution on  $[0.9, 1.1]$ . We tested all (conditional) independences with a maximum of one conditioning variable using Spearman’s rank correlation test with a  $p$ -value threshold of 0.01.<sup>27</sup> Table A1

<sup>27</sup> Because the LCD algorithm only uses conditional independence tests with a maximum of one variable in the conditioning test, we do not consider conditional independence tests with larger conditioning sets in this work. We did experiment with larger conditioning sets, but we were not able to retrieve all predicted conditional dependences with our parameter settings and only  $n = 500$  samples.

shows that the conditional independences with a maximum conditioning set of size one that are implied by the equilibrium Markov ordering graph are indeed present in the simulated data.

To verify the predicted LCD triples of Section 5.5, we ran the simulation  $n = 500$  times with  $k_{me} = C$  as the context variable. Parameters were as described in Appendix B, except that  $I$ ,  $k_{IS}$ ,  $k_{FS}$ ,  $k_{SR}$ ,  $k_{rm}$ , and  $k_{me}$  were all drawn randomly from a uniform distribution on  $[1.0, 1.1]$ , and the parameter  $k_{F_eE}$  from a uniform distribution on  $[1.2, 1.3]$ . Note that some parameter tuning is necessary if one wants to ensure that the equilibrium state satisfies the saturation conditions  $(T_e - X_e(t)) \gg K_{me}$  and  $X_e(t) \gg K_{F_eE}$ . On the other hand, the parameters need to have sufficient random variation, otherwise LCD may fail because of (almost) deterministic relations between variables. We ran the simulations until the system reaches equilibrium and recorded the equilibrium values of the variables. We then applied the LCD algorithm to search for LCD triples in this equilibrium data with context variable  $C = k_{me}$ . For the conditional independence tests, we used Spearman's rank correlation with a  $p$ -value threshold of 0.01. We found the expected LCD triples  $(C, v_m, v_r)$ ,  $(C, v_m, v_s)$ ,  $(C, v_m, v_e)$ ,  $(C, v_r, v_s)$ ,  $(C, v_r, v_e)$ ,  $(C, v_s, v_e)$ , and no others.

## D Rewriting equations may reveal additional structure

Theorem 1 specifies sufficient but not necessary conditions for the presence of perfect adaptation. The equilibrium distribution of some systems is not faithful to the equilibrium Markov ordering graph associated with the equilibrium equations of the model. Here, we will discuss a dynamical model for a basic enzymatic reaction, and we will demonstrate that this model is capable of perfect adaptation. However, it does not satisfy the conditions in Theorem 1, and the presence of directed paths in the equilibrium causal ordering graph does not imply the presence of a causal effect at equilibrium. We will also show that this may be addressed by appropriately rewriting the equations.

The basic enzyme reaction models a substrate  $S$  that reacts with an enzyme  $E$  to form a complex  $C$ , which is converted into a product  $P$  and the enzyme  $E$ . The dynamical equations for the concentrations  $X_S(t)$ ,  $X_E(t)$ ,  $X_C(t)$ , and  $X_P(t)$  are given by:

$$\dot{X}_S(t) = k_0 - k_1 X_S(t) X_E(t) + k_{-1} X_C(t), \quad (\text{A1})$$

$$\dot{X}_C(t) = k_1 X_S(t) X_E(t) - (k_{-1} + k_2) X_C(t), \quad (\text{A2})$$

$$\dot{X}_E(t) = -k_1 X_S(t) X_E(t) + (k_{-1} + k_2) X_C(t), \quad (\text{A3})$$

$$\dot{X}_P(t) = k_2 X_C(t) - k_3 X_P(t), \quad (\text{A4})$$

where  $k_{-1}$ ,  $k_0$ ,  $k_1$ ,  $k_2$ , and  $k_3$  are parameters taking value in  $\mathbb{R}_{>0}$  [39,56]. We treat the initial conditions  $X_S(0) = S_0$ ,  $X_C(0) = C_0$ ,  $X_E(0) = E_0$ , and  $X_P(0) = P_0$  as independent exogenous random variables, and consider the parameter  $k_1$  as input signal.

Figure A1(a) shows the dynamic bipartite graph corresponding to the dynamic equations, and Figure A1(b) the corresponding dynamic causal ordering graph. Since there is a directed path from  $k_1$  to  $v_P$  in Figure A1(b), we would expect that a change in  $k_1$  would generically lead to a response of  $X_P(t)$ . We verified this by simulating this model with  $k_{-1} = 1.0$ ,  $k_0 = 1.0$ ,  $k_1 = k_1^- = 10.0$ ,  $k_2 = 0.8$ , and  $k_3 = 2.5$  and with initial conditions  $S_0 = 0.45$ ,  $E_0 = 0.5$ ,  $C_0 = 1.25$ , and  $P_0 = 0.40$  until the system reached equilibrium. We then recorded the response of  $X_P(t)$  after changing the input signal  $k_1$  into different values  $k_1^+$ . Figure A1(c) shows that  $X_P(t)$  indeed responds to changes in the input signal  $k_1$ , but eventually returns to its original equilibrium value.

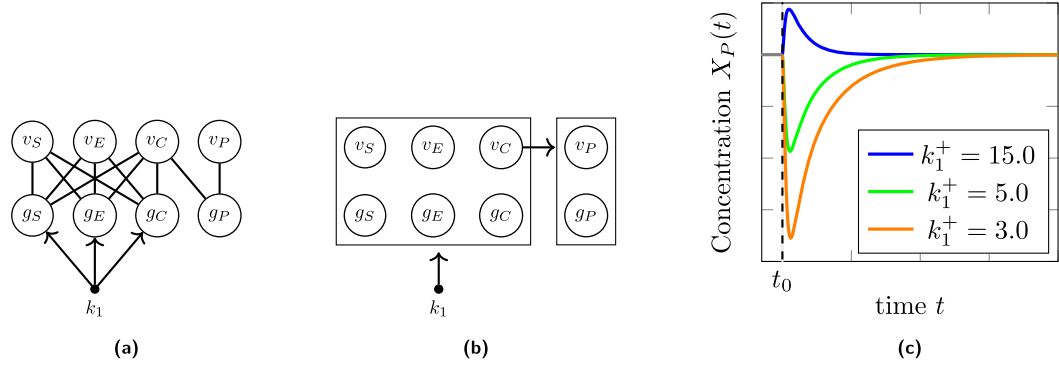
The equilibrium equations of the model are given by:

$$f_S : k_0 - k_1 X_S X_E + k_{-1} X_C = 0, \quad (\text{A5})$$

$$f_C : k_1 X_S X_E - (k_{-1} + k_2) X_C = 0, \quad (\text{A6})$$

$$f_E : -k_1 X_S X_E + (k_{-1} + k_2) X_C = 0, \quad (\text{A7})$$





**Figure A1:** The dynamic bipartite graph (a) and dynamic causal ordering graph (b) for the basic enzyme reaction modelled by equations (A1)–(A4). Panel (c) shows simulations that suggest that the concentration  $X_P$  perfectly adapts after an initial transient response to a persistent change in the parameter  $k_1$ .

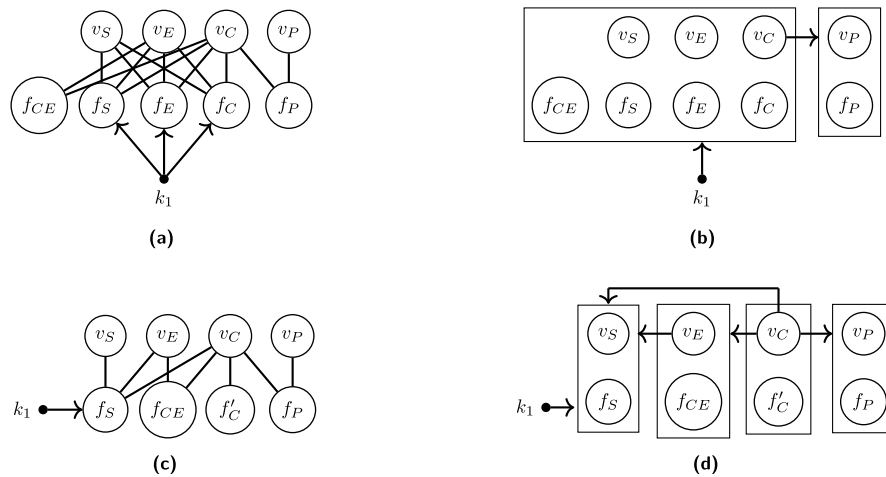
$$f_P : k_2 X_C - k_3 X_P = 0, \quad (\text{A8})$$

$$f_{CE} : X_C + X_E - (C_0 + E_0) = 0, \quad (\text{A9})$$

where the last equation is derived from the constant of motion  $X_C(t) + X_E(t)$  (see [39] for more details). The equilibrium bipartite graph (Figure A2(a)) does not have a perfect matching (as it contains more equations than endogenous variables), but we can apply the extended causal ordering algorithm [12] to construct the equilibrium causal ordering graph in Figure A2(b). There is a directed path from  $k_1$  to  $v_P$  in the equilibrium Markov ordering graph. Therefore, even though the basic enzyme reaction does achieve perfect adaptation, we see that it does not satisfy the conditions of Theorem 1. The basic enzyme reaction is an example of a system for which directed paths in the equilibrium causal ordering graph do not imply generic causal relations between variables.

By rewriting the equilibrium equations, we can achieve stronger conclusions for this particular case. For instance, we can consider the equation  $f'_C$ , obtained from summing equations  $f_S$  and  $f_C$ :

$$f'_C : k_0 - k_2 X_C = 0, \quad (\text{A10})$$



**Figure A2:** Different graphical representations of the basic enzyme reaction at equilibrium. Panels (a) and (b) show the equilibrium bipartite graph and equilibrium causal ordering graph, respectively, for the equilibrium equations (A5), (A6), (A7), (A8), and (A9). Panels (c) and (d) show the equilibrium bipartite graph and the equilibrium causal ordering graph, respectively, for the rewritten equilibrium equations (A5), (A8), (A9), and (A10). Rewriting the equilibrium equations yields a sparser equilibrium bipartite graph, and therefore, more structure is revealed in the equilibrium causal ordering graph.

in combination with  $f_S$ ,  $f_P$ , and  $f_{CE}$ . The equilibrium equations  $f_C$  and  $f_E$  can be dropped because they are linear combinations of the other equations. The graphs corresponding with these rewritten equations are shown in Figure A2(c) and (d). The equilibrium bipartite graph for the rewritten equations in Figure A2(c) turns out to be sparser than the one for the original equilibrium equations in Figure A2(a). The equilibrium causal ordering graph in Figure A2(d) for the rewritten equilibrium equations consequently reveals more structure than the one in Figure A2(b) for the original equilibrium equations.

The two equilibrium causal ordering graphs do not model the same set of perfect interventions. For example, the (non)effects of an intervention targeting the cluster  $\{v_S, f_S\}$  in the equilibrium causal ordering graph in Figure A2(d) (where  $f_S$  is replaced by an equation  $v_S = \xi_S$  setting  $v_S$  equal to a constant  $\xi_S \in \mathbb{R}_{>0}$ ) cannot be read off from the equilibrium causal ordering graph in Figure A2(b). Furthermore, there is no directed path from  $k_1$  to  $v_P$  in Figure A2(d), and hence, we can now make use of Theorem 1 to conclude that the concentration  $X_P(t)$  perfectly adapts to persistent changes of the parameter  $k_1$ .

## E Transforming variables may reveal structure

The IFFLP network in Ma et al. [17] that we briefly discussed in Section 3.1.3 could be a graphical representation of the following differential equations:

$$\dot{X}_A(t) = I(t)k_{IA}\frac{(1 - X_A(t))}{K_{IA} + (1 - X_A(t))} - F_A k_{F_{AA}} \frac{X_A(t)}{K_{F_{AA}} + X_A(t)}, \quad (\text{A11})$$

$$\dot{X}_B(t) = X_A(t)k_{AB}\frac{(1 - X_B(t))}{K_{AB} + (1 - X_B(t))} - F_B k_{F_{BB}} \frac{X_B(t)}{K_{F_{BB}} + X_B(t)}, \quad (\text{A12})$$

$$\dot{X}_C(t) = X_A(t)k_{AC}\frac{(1 - X_C(t))}{K_{AC} + (1 - X_C(t))} - X_B(t)k_{BC}\frac{X_C(t)}{K_{BC} + X_C(t)}, \quad (\text{A13})$$

where  $I(t)$  represents an external input into the system. This network is capable of perfect adaptation if the first term of  $\dot{X}_B(t)$  is in the saturated region  $(1 - X_B(t)) \gg K_{AB}$  and the second term is in the linear region  $X_B(t) \ll K_{F_{BB}}$ , which allows us to make the following approximation:

$$\frac{dX_B(t)}{dt} \approx X_A(t)k_{AB} - \frac{F_B k_{F_{BB}}}{K_{F_{BB}}} X_B(t). \quad (\text{A14})$$

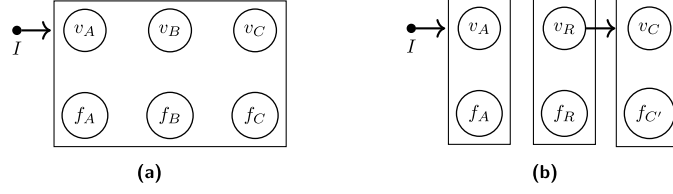
Therefore, an equilibrium solution  $X_B$  for  $B$  is proportional to the equilibrium solution  $X_A$  for  $A$ . Since both terms in equation (A13) are proportional to  $X_A$ , we find that the equilibrium solution  $X_C$  for  $C$  is a function of only the parameters  $k_{AC}$ ,  $K_{AC}$ ,  $k_{BC}$ , and  $K_{BC}$  (note that  $X_A$  factors out of the equilibrium equation corresponding to (A14)), and hence, it does not depend on the input parameter  $I(t)$ . Since a change in the input signal  $I(t)$  changes  $\dot{X}_A(t)$ , there is a transient effect on  $X_A(t)$ . Similarly there must also be a transient effect on both  $X_B(t)$  and  $X_C(t)$ . It follows that the system achieves perfect adaptation.

The equilibrium equations associated with equations (A11), the approximation (A14) to (A12), and (A13) are given by:

$$f_A : I k_{IA} \frac{(1 - X_A)}{K_{IA} + (1 - X_A)} - F_A k_{F_{AA}} \frac{X_A}{K_{F_{AA}} + X_A} = 0, \quad (\text{A15})$$

$$f_B : X_A k_{AB} - \frac{F_B k_{F_{BB}}}{K_{F_{BB}}} X_B = 0, \quad (\text{A16})$$

$$f_C : X_A k_{AC} \frac{(1 - X_C)}{K_{AC} + (1 - X_C)} - X_B k_{BC} \frac{X_C}{K_{BC} + X_C} = 0. \quad (\text{A17})$$



**Figure A3:** Equilibrium causal ordering graphs for the IFFLP network. (a) Modelled by equations (A15), (A16), and (A17) and variables  $X_A, X_B, X_C$ ; (b) modelled by equations (A18), (A19), and (A20) and variables  $X_A, X_R, X_C$ .

The associated equilibrium causal ordering graph in Figure A3(a) shows that there is a directed path from the input signal  $I$  to the cluster  $\{v_A, v_B, v_C\}$ . Therefore, the conditions of Theorem 1 are not satisfied for the system.

Interestingly, though, if we first make a change of variables from  $(X_A, X_B, X_C) \mapsto (X_A, X_R, X_C)$  with  $X_R := X_B/X_A$  (assuming that  $X_A \in \mathbb{R}_{>0}$ ), we can make use of the causal ordering approach. Indeed, we can rewrite the equilibrium equations as follows in terms of the new variables  $X_A, X_R, X_C$ :

$$f_A : Ik_{IA} \frac{(1 - X_A)}{K_{IA} + (1 - X_A)} - F_A k_{FAA} \frac{X_A}{K_{FAA} + X_A} = 0, \quad (\text{A18})$$

$$f_R : k_{AB} - \frac{F_B k_{FBB}}{K_{FBB}} X_R = 0, \quad (\text{A19})$$

$$f_{C'} : k_{AC} \frac{(1 - X_C)}{K_{AC} + (1 - X_C)} - X_R k_{BC} \frac{X_C}{K_{BC} + X_C} = 0. \quad (\text{A20})$$

This yields a sparser equilibrium bipartite graph. As shown in Figure A3(b), we can now read off from the corresponding equilibrium causal ordering graph that the input signal  $I$  does not affect the equilibrium values of  $X_R$  and  $X_C$ . Hence, in this parameterization of the model, Theorem 1 can be used to identify the perfect adaptive behaviour of the system.

## F Markov ordering graphs have no inherent causal interpretation

The causal interpretation of the equilibrium Markov ordering graph for the bathtub model is discussed at length in ref. [12]. The conclusion is that the Markov ordering graph *alone* does not contain enough information to read off the effects of interventions in an unambiguous way.<sup>28</sup> As a result, the Markov ordering graph does not have a straightforward causal interpretation in terms of interventions, contrary to what is sometimes claimed [13,14]. For the sake of completeness, we will summarize here the discussion of the causal interpretation of the equilibrium Markov ordering graph of the bathtub model.

**Example 2.** For the bathtub model (Section 3.1.1), consider an intervention targeting the dynamical equation  $g_D$  that also changes the associated equilibrium equation  $f_D$ . For example, consider putting the bathtub outside in the rain. This will not change the inflow through the faucet, but will add to the total amount of inflowing water into the tub and can be modelled by modifying equation (4) into:

$$\dot{X}_D(t) = U_1(X_I(t) - X_O(t) + U_R(t)),$$

<sup>28</sup> To arrive at this conclusion, we first need to explicitly state what we mean when we talk about ‘‘causal relations.’’ We follow the common interpretation in contemporary literature in terms of interventions. In the context of a specific model, this means that an intervention on the cause brings about a change in the effect.

where  $U_R(t)$  is a new exogenous variable that quantifies the amount of in-flowing water due to the rain (in the original model,  $U_R(t) = 0$ ). Through explicit calculations, it can be shown that this soft intervention has an effect on  $X_O$ ,  $X_P$ , and  $X_D$  at equilibrium.<sup>29</sup> If we were to read off the effects of a soft intervention targeting  $f_D$  by verifying whether there is a directed path from  $v_D$  in the equilibrium Markov ordering graph, we would erroneously conclude that this intervention *only* has an effect on  $X_D$ . In a similar fashion, it can be shown that the graph does not represent the effects of perfect interventions targeting  $\{f_D, v_D\}$  or  $\{f_D, v_O\}$  either. From the equilibrium causal ordering graph, one can read off that the absence and presence of (generic) effects of the perfect intervention  $\{f_P, v_D\}$  correspond with the absence and presence of directed paths in the equilibrium Markov ordering graph starting from  $v_D$ . Because the equilibrium Markov ordering graph alone does not specify to which experimental procedure a perfect intervention on one of its vertices corresponds to, we conclude that the equilibrium Markov ordering graph on its own does not possess a causal interpretation.

A correct interpretation of a directed edge ( $v_i \rightarrow v_j$ ) in the equilibrium Markov ordering graph would be that an intervention targeting equations in the cluster of  $v_i$  in *the causal ordering graph* generically will have an effect on the equilibrium distribution of  $v_j$ . However, the Markov ordering graph itself does not specify the variables and equations that share a cluster with the variables in the causal ordering graph. In many dynamical systems, though, the equilibrium equations  $f_i$  derived from differential equations for variables  $X_i(t)$  end up in the same cluster as the associated variable  $v_i$ . Under such an assumption, one could give an unambiguous causal interpretation to the Markov ordering graph in terms of perfect interventions.<sup>30</sup>

However, there is another obstacle to interpreting a directed edge ( $v_i \rightarrow v_j$ ) in the Markov ordering graph as a *direct* causal effect. This can go wrong in case causal cycles are present. Indeed, the Markov ordering graph for a simple SCM is the acyclification of its graph [23]. Therefore, if  $v_j$  is part of a feedback loop together with  $v_k$ , and the structural equation of  $v_k$  depends on  $v_i$ , the Markov ordering graph will contain the directed edge  $v_i \rightarrow v_j$  even if there is no direct effect of  $v_i$  on  $v_j$  (i.e., if the structural equation for  $v_j$  does not depend on  $v_i$ ).

Therefore, if one cannot rule out the possibility of feedback, one should avoid reading the Markov ordering graph as if the directed edges represent direct causal effects (and as if directed paths represent causal effects).

<sup>29</sup> Assuming that the system converges to equilibrium after the intervention, the equilibrium causal ordering graph, in Figure 5(g) tells us that soft interventions on  $f_D$  *generically* change the equilibrium distributions of  $X_O$ ,  $X_P$ , and  $X_D$ . We need explicit calculations to verify that the generic effects correspond to a change in distribution.

<sup>30</sup> For first-order dynamical systems in which each variable is self-regulating, each variable  $v_i$  shares a cluster with the equilibrium equation  $f_i$  associated with the time derivative  $\dot{X}_i(t)$  [16]. Note that the examples of perfectly adaptive systems that we have considered in this work do contain variables that are not self-regulating.