Submitted to the Annals of Statistics

SUPPLEMENT TO "FOUNDATIONS OF STRUCTURAL CAUSAL MODELS WITH CYCLES AND LATENT VARIABLES"

BY STEPHAN BONGERS[1,*], PATRICK FORRÉ[1,†], JONAS PETERS[2,‡] AND JORIS M. MOOIJ[3,*,†]

[1]Informatics Institute, University of Amsterdam, s.r.bongers@uva.nl; p.d.forre@uva.nl

[2]Department of Mathematical Sciences, University of Copenhagen, jonas.peters@math.ku.dk

[3]Korteweg-De Vries Institute, University of Amsterdam, j.m.mooij@uva.nl

This Supplementary Material contains a summary of the basic terminology and results for causal graphical models (Appendix A), additional (unique) solvability properties (Appendix B), some results for linear SCMs (Appendix C), other examples (Appendix D), the proofs of all the theoretical results (Appendix E) and the measurable selection theorems (Appendix F) that are used in several proofs.

APPENDIX A: CAUSAL GRAPHICAL MODELS

In this appendix, we provide a summary of the basic terminology and results for causal graphical models. In Appendix A.1 we provide the terminology for directed (mixed) graphs. In Appendix A.2 we give an introduction and an intuitive derivation of Markov properties for SCMs with cycles. In Appendix A.3 we provide a definition of modular SCMs and show how they relate to SCMs. In Appendix A.4 we provide an overview of the causal graphical models related to SCMs. The proofs of the theoretical results in this appendix are given in Appendix E.

A.1. Directed (mixed) graphs. In this subsection, we introduce the terminology for directed (mixed) graphs, where we do allow for cycles [8, 15, 20, 23].

DEFINITION A.1 (Directed (mixed) graph).

1. *A* directed graph *is a pair* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, where* $\mathcal{V}$ *is a set of nodes and* $\mathcal{E}$ *is a set of directed edges, which is a subset* $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ *of ordered pairs of nodes. Each element* $(i, j) \in \mathcal{E}$ *can be represented by the directed edge* $i \to j$ *or equivalently* $j \leftarrow i$*. In particular,* $(i, i) \in \mathcal{E}$ *represents a* self-cycle $i \to i$*.*
2. *A* directed mixed graph *is a triple* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$*, where the pair* $(\mathcal{V}, \mathcal{E})$ *forms a directed graph and* $\mathcal{B}$ *is a set of bidirected edges, which is a subset* $\mathcal{B} \subseteq \{\{i, j\} : i, j \in \mathcal{V}, i \neq j\}$ *of unordered (distinct) pairs of nodes. Each element* $\{i, j\} \in \mathcal{B}$ *can be represented by the bidirected edge* $i \leftrightarrow j$ *or equivalently* $j \leftrightarrow i$*. Note that a directed graph can be considered as a directed mixed graph without bidirected edges.*
3. *Let* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ *be a directed mixed graph. A directed mixed graph* $\tilde{\mathcal{G}} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}}, \tilde{\mathcal{B}})$ *is a* subgraph of $\mathcal{G}$ *if* $\tilde{\mathcal{V}} \subseteq \mathcal{V}$*,* $\tilde{\mathcal{E}} \subseteq \mathcal{E}$ *and* $\tilde{\mathcal{B}} \subseteq \mathcal{B}$*, in which case we write* $\tilde{\mathcal{G}} \subseteq \mathcal{G}$*. For a subset* $\mathcal{W} \subseteq \mathcal{V}$*, we define the* induced subgraph of $\mathcal{G}$ on $\mathcal{W}$ *by* $\mathcal{G}_{\mathcal{W}} := (\mathcal{W}, \tilde{\mathcal{E}}, \tilde{\mathcal{B}})$*, where* $\tilde{\mathcal{E}}$ *and* $\tilde{\mathcal{B}}$

*Supported in part by NWO, the Netherlands Organization for Scientific Research (VIDI grant 639.072.410 and VENI grant 639.031.036).

†Supported in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n[O] 639466).

‡Supported by research grants from VILLUM FONDEN (18968) and the Carlsberg Foundation.

are the set of directed and bidirected edges in $\mathcal{E}$ and $\mathcal{B}$, respectively, that lie in $\mathcal{W} \times \mathcal{W}$ and $\{\{i,j\} : i,j \in \mathcal{W}, i \neq j\}$, respectively.

4. *A* walk *between $i,j \in \mathcal{V}$ in a directed mixed graph $\mathcal{G}$ is a tuple $(i_0, \epsilon_1, i_1, \epsilon_2, i_2, \ldots, \epsilon_n, i_n)$ of alternating nodes and edges in $\mathcal{G}$ for some $n \geq 0$, where all $i_0, \ldots, i_n \in \mathcal{V}$, all $\epsilon_1, \ldots, \epsilon_n \in \mathcal{E} \cup \mathcal{B}$ such that $\epsilon_k \in \{i_{k-1} \to i_k, i_{k-1} \leftarrow i_k, i_{k-1} \leftrightarrow i_k\}$ for all $k = 1, \ldots, n$, and it starts with node $i_0 = i$ and ends with node $i_n = j$. Note that $n = 0$ corresponds with a trivial walk consisting of a single node. If all nodes $i_0, \ldots, i_n$ are distinct, it is called a* path. *A walk (path) of the form $i \to \cdots \to j$, that is, ϵ_k is $i_{k-1} \to i_k$ for all $k = 1, 2, \ldots, n$, is called a* directed walk (path) *from i to j.*
5. *A* cycle *through $i \in \mathcal{V}$ in a directed mixed graph $\mathcal{G}$ is a directed path from i to some node j extended with the edge $j \to i \in \mathcal{E}$. In particular, a self-cycle $i \to i \in \mathcal{E}$ is a cycle. Note that a path cannot contain any cycles. A directed graph and a directed mixed graph are said to be* acyclic *if they contain no cycles, and are then referred to as a* directed acyclic graph (DAG) *and an* acyclic directed mixed graph (ADMG)*, respectively.*
6. *For a directed mixed graph $\mathcal{G}$ and a node $i \in \mathcal{V}$ we define the set of* parents *of i by $\mathrm{pa}_{\mathcal{G}}(i) := \{j \in \mathcal{V} : j \to i \in \mathcal{E}\}$, the set of children of i by $\mathrm{ch}_{\mathcal{G}}(i) := \{j \in \mathcal{V} : i \to j \in \mathcal{E}\}$, the set of* ancestors *of i by*

   $$\mathrm{an}_{\mathcal{G}}(i) := \{j \in \mathcal{V} : \textit{there is a directed path from } j \textit{ to } i \textit{ in } \mathcal{G}\}$$

   and the set of descendants *of i by*

   $$\mathrm{de}_{\mathcal{G}}(i) := \{j \in \mathcal{V} : \textit{there is a directed path from } i \textit{ to } j \textit{ in } \mathcal{G}\}.$$

   Note that we have $\{i\} \cup \mathrm{pa}_{\mathcal{G}}(i) \subseteq \mathrm{an}_{\mathcal{G}}(i)$ and $\{i\} \cup \mathrm{ch}_{\mathcal{G}}(i) \subseteq \mathrm{de}_{\mathcal{G}}(i)$. We can apply all these definitions to subsets $\mathcal{U} \subseteq \mathcal{V}$ by taking unions, for example $\mathrm{pa}_{\mathcal{G}}(\mathcal{U}) := \cup_{i \in \mathcal{U}} \mathrm{pa}_{\mathcal{G}}(i)$. A subset $\mathcal{A} \subseteq \mathcal{V}$ is called an ancestral subset *in $\mathcal{G}$ if $\mathcal{A} = \mathrm{an}_{\mathcal{G}}(\mathcal{A})$, that is, $\mathcal{A}$ is closed under taking ancestors of $\mathcal{A}$ in $\mathcal{G}$.*
7. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph. We call $\mathcal{G}$* strongly connected *if for every pair of distinct nodes $i,j \in \mathcal{V}$, the graph contains a cycle that passes through both i and j. The* strongly connected component of *$i \in \mathcal{V}$, denoted by $\mathrm{sc}_{\mathcal{G}}(i)$, is the maximal subset $\mathcal{S} \subseteq \mathcal{V}$ such that $i \in \mathcal{S}$ and the induced subgraph $\mathcal{G}_{\mathcal{S}}$ is strongly connected. Equivalently, $\mathrm{sc}_{\mathcal{G}}(i) = \mathrm{an}_{\mathcal{G}}(i) \cap \mathrm{de}_{\mathcal{G}}(i)$.*
8. *A* loop *in a directed mixed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ is a subset $\mathcal{O} \subseteq \mathcal{V}$ that is strongly connected in the induced subgraph $\mathcal{G}_{\mathcal{O}}$ of $\mathcal{G}$ on $\mathcal{O}$.*
9. *For a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we define the* graph of strongly connected components of *$\mathcal{G}$ as the directed graph $\mathcal{G}^{\mathrm{sc}} := (\mathcal{V}^{\mathrm{sc}}, \mathcal{E}^{\mathrm{sc}})$, where $\mathcal{V}^{\mathrm{sc}}$ are the strongly connected components of $\mathcal{G}$, that is, $\mathcal{V}^{\mathrm{sc}}$ are the equivalence classes in $\mathcal{V}/\!\sim$ with the equivalence relation $i \sim j$ if and only if $i \in \mathrm{sc}_{\mathcal{G}}(j)$, and $\mathcal{E}^{\mathrm{sc}} = (\mathcal{E} \setminus \{i \to i : i \in \mathcal{V}\})/\!\sim$ with the equivalence relation $(i \to j) \sim (i' \to j')$ if and only if $i \sim i'$ and $j \sim j'$.*

We omit the subscript $\mathcal{G}$ whenever it is clear which directed (mixed) graph $\mathcal{G}$ we are referring to.

LEMMA A.2 (DAG of strongly connected components). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed graph. Then $\mathcal{G}^{\mathrm{sc}}$, the graph of strongly connected components of $\mathcal{G}$, is a DAG.*

A.2. Markov properties. In this subsection, we give a short overview of Markov properties for SCMs with cycles. We will make use of the Markov properties that were recently developed by Forré and Mooij [8] for HEDGes, a graphical representation that is similar to the augmented graph of SCMs. We briefly summarize some of their main results and apply them to the class of SCMs. We also provide a shorter and more intuitive derivation so that this subsection can act as an entry point for the reader into the more extensive discussion of Markov properties provided in [8].

Markov properties associate a set of conditional independence relations to a graph. The directed global Markov property for directed acyclic graphs, also known as the d-separation criterion [19], is one of the most widely used. It directly extends to a similar property for acyclic directed mixed graphs (ADMGs) [23]. It does not hold in general for cyclic SCMs, however, as was already observed earlier [26, 27]. Under some conditions (roughly speaking, linearity or discrete variables) the directed global Markov property can be shown to hold also in the presence of cycles [8].

Inspired by work of Spirtes [26], Forré and Mooij [8] recognized that in the general cyclic case a different extension of d-separation, termed σ-separation, is needed, leading to the general directed global Markov property. One key result in [8] implies that under the assumption of unique solvability w.r.t. each strongly connected component of its graph, the observational distribution of an SCM satisfies the general directed global Markov property w.r.t. its graph. The solvability assumptions are in general not preserved under interventions. Under the stronger assumption of simplicity, however, they are, and one obtains the corollary that also all interventional and counterfactual distributions of a simple SCM satisfy the general directed global Markov property w.r.t. to their corresponding graphs.

For a more extensive study of different Markov properties that can be associated to SCMs we refer the reader to [8].

A.2.1. *The directed global Markov property.* Conditional independencies in the observational distribution of an acyclic SCM can be read off from its graph by using the graphical criterion called d-separation [20]. The directed global Markov property associates a conditional independence relation in the observational distribution of the SCM to each d-separation entailed by the graph. Here, we use a formulation of d-separation that generalizes d-separation for DAGs [19] and m-separation for ADMGs [23] and mDAGs [7].

DEFINITION A.3 (Collider). *Let $\pi = (i_0, \epsilon_1, i_1, \epsilon_2, i_2, \dots, \epsilon_n, i_n)$ be a walk (path) in a directed mixed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$. A node i_k on π is called a* collider on π *if it is a non-endpoint node ($1 \leq k < n$) and the two edges $\epsilon_k, \epsilon_{k+1}$ meet head-to-head on i_k (i.e., if the subwalk $(i_{k-1}, \epsilon_k, i_k, \epsilon_{k+1}, i_{k+1})$ is of the form $i_{k-1} \to i_k \leftarrow i_{k+1}$, $i_{k-1} \leftrightarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \to i_k \leftrightarrow i_{k+1}$ or $i_{k-1} \leftrightarrow i_k \leftrightarrow i_{k+1}$). The node i_k is called a* non-collider on π *otherwise, that is, if it is an endpoint node ($k = 0$ or $k = n$) or if the subwalk $(i_{k-1}, \epsilon_k, i_k, \epsilon_{k+1}, i_{k+1})$ is of the form $i_{k-1} \to i_k \to i_{k+1}$, $i_{k-1} \leftarrow i_k \leftarrow i_{k+1}$, $i_{k-1} \leftarrow i_k \to i_{k+1}$, $i_{k-1} \leftrightarrow i_k \to i_{k+1}$ or $i_{k-1} \leftarrow i_k \leftrightarrow i_{k+1}$.*

Note in particular that the end points of a walk are non-colliders on the walk.

DEFINITION A.4 (d-separation). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph and let $C \subseteq \mathcal{V}$ be a subset of nodes. A walk (path) $\pi = (i_0, \epsilon_1, i_1, \dots, i_n)$ in $\mathcal{G}$ is said to be C-d-*blocked or d-blocked by C *if*

1. *it contains a collider $i_k \notin \mathrm{an}_{\mathcal{G}}(C)$, or*
2. *it contains a non-collider $i_k \in C$.*

*The walk (path) π is said to be C-d-*open *if it is not d-blocked by C. For two subsets of nodes $A, B \subseteq \mathcal{V}$, we say that A* is d-separated from B given C in $\mathcal{G}$ *if all paths between any node in A and any node in B are d-blocked by C, and write*

$$A \underset{\mathcal{G}}{\overset{d}{\perp}} B \,|\, C \,.$$

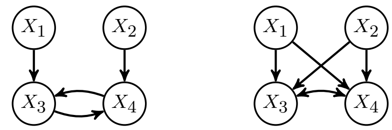

Fig 1: The graphs of the observationally equivalent SCMs $\mathcal{M}$ (left) and $\tilde{\mathcal{M}}$ (right) of Example A.8 and A.10.

The next lemma is a straightforward generalization of Lemma 3.3 in [9] to the cyclic setting. It implies that it suffices to formulate d-separation in terms of paths rather than walks.

LEMMA A.5. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph, $C \subseteq \mathcal{V}$ and $i, j \in \mathcal{V}$. There exists a C-d-open walk between i and j in $\mathcal{G}$ if and only if there exists a C-d-open path between i and j in $\mathcal{G}$.*

DEFINITION A.6 (Directed global Markov property). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph and $\mathbb{P}_{\mathcal{V}}$ a probability distribution on $\boldsymbol{\mathcal{X}}_{\mathcal{V}} = \prod_{i\in\mathcal{V}} \mathcal{X}_i$, where each $\mathcal{X}_i$ is a standard probability space. The probability distribution $\mathbb{P}_{\mathcal{V}}$ satisfies the* directed global Markov property *relative to $\mathcal{G}$ if for all subsets $A, B, C \subseteq \mathcal{V}$ we have*

$$A \overset{d}{\underset{\mathcal{G}}{\perp}} B \,|\, C \quad \Longrightarrow \quad \boldsymbol{X}_A \underset{\mathbb{P}_{\mathcal{V}}}{\perp\!\!\!\perp} \boldsymbol{X}_B \,|\, \boldsymbol{X}_C \,,$$

that is, $(X_i)_{i\in A}$ and $(X_i)_{i\in B}$ are conditionally independent given $(X_i)_{i\in C}$ under $\mathbb{P}_{\mathcal{V}}$, where we take the canonical projections $X_i : \boldsymbol{\mathcal{X}}_{\mathcal{V}} \to \mathcal{X}_i$ as random variables.

From the results in [8] it directly follows that for the observational distribution of an SCM, the directed global Markov property w.r.t. the graph of the SCM (also known as the d-separation criterion), holds under one of the following assumptions.

THEOREM A.7 (Directed global Markov property for SCMs [8]). *Let $\mathcal{M}$ be a uniquely solvable SCM that satisfies at least one of the following three conditions:*

1. *$\mathcal{M}$ is acyclic;*
2. *all endogenous spaces $\mathcal{X}_i$ are discrete and $\mathcal{M}$ is ancestrally uniquely solvable;*
3. *$\mathcal{M}$ is linear (see Definition C.1), each of its causal mechanisms $\{f_i\}_{i\in\mathcal{I}}$ has a nontrivial dependence on at least one exogenous variable, and $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$ has a density w.r.t. the Lebesgue measure on $\mathbb{R}^{\mathcal{J}}$.*

Then its observational distribution $\mathbb{P}^{\boldsymbol{X}}$ exists, is unique and satisfies the directed global Markov property relative to $\mathcal{G}(\mathcal{M})$ (see Definition A.6).

The acyclic case is well known and was first shown in the context of linear-Gaussian structural equation models [14, 29]. The discrete case fixes the erroneous theorem by Pearl and Dechter [21], for which a counterexample was found by Neal [18], by adding the ancestral unique solvability condition, and extends it to allow for bidirected edges in the graph. The linear case is an extension of existing results for the linear-Gaussian setting without bidirected edges [13, 26, 27] to a linear (possibly non-Gaussian) setting with bidirected edges in the graph.

The following counterexample of an SCM for which the directed global Markov property does not hold was already given in [26, 27].

EXAMPLE A.8 (Directed global Markov property does not hold for cyclic SCM). *Consider the SCM $\mathcal{M} = \langle \mathbf{4}, \mathbf{4}, \mathbb{R}^4, \mathbb{R}^4, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^4} \rangle$ with causal mechanism given by*

$$f_1(\boldsymbol{x},\boldsymbol{e}) = e_1\,, \quad f_2(\boldsymbol{x},\boldsymbol{e}) = e_2\,, \quad f_3(\boldsymbol{x},\boldsymbol{e}) = x_1x_4 + e_3\,, \quad f_4(\boldsymbol{x},\boldsymbol{e}) = x_2x_3 + e_4$$

and $\mathbb{P}_{\mathbb{R}^4}$ is the standard-normal distribution on $\mathbb{R}^4$. The graph of $\mathcal{M}$ is depicted in Figure 1 on the left. The model is uniquely solvable (it is even simple). One can check that for every solution $\boldsymbol{X}$ of $\mathcal{M}$, X_1 is not independent of X_2 given $\{X_3, X_4\}$. However, the variables X_1 and X_2 are d-separated given $\{X_3, X_4\}$ in $\mathcal{G}(\mathcal{M})$. Hence the global directed Markov property does not hold here.

In constraint-based approaches to causal discovery, one usually assumes the converse of the directed global Markov property to hold [20, 28].

DEFINITION A.9 (d-Faithfulness). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph and $\mathbb{P}_{\mathcal{V}}$ a probability distribution on $\boldsymbol{\mathcal{X}}_{\mathcal{V}} = \prod_{i\in\mathcal{V}} \mathcal{X}_i$, where each $\mathcal{X}_i$ is a standard probability space. The probability distribution $\mathbb{P}_{\mathcal{V}}$ is d-*faithful *to $\mathcal{G}$ if for all subsets $A, B, C \subseteq \mathcal{V}$ we have*

$$A \underset{\mathcal{G}}{\overset{d}{\perp}} B \,|\, C \quad \Longleftarrow \quad \boldsymbol{X}_A \underset{\mathbb{P}_{\mathcal{V}}}{\perp\!\!\!\perp} \boldsymbol{X}_B \,|\, \boldsymbol{X}_C\,,$$

where we take the canonical projections $X_i : \boldsymbol{\mathcal{X}}_{\mathcal{V}} \to \mathcal{X}_i$ as random variables.

In other words, the d-faithfulness assumption states that the graph explains, via d-separation, all the conditional independencies that are present in the observational distribution. Meek [17] showed that for multinomial and linear-Gaussian DAG (i.e., acyclic and causally sufficient SCMs) models, d-faithfulness holds for all parameter values up to a measure zero set (in a natural parameterization). Up to our knowledge no such results have been shown in more general parametric or nonparametric settings (neither in the acyclic case, nor in the cyclic one).

A.2.2. *The general directed global Markov property.* In [8] the general directed global Markov property is introduced, that is based on σ-separation, an extension of d-separation. This notion of σ-separation was derived from the notion of d-separation in the acyclification of the graph. The acyclification of a graph generalizes the idea of the collapsed graph for directed graphs, developed by Spirtes [26], to HEDGes. In particular, this notion can be applied to directed mixed graphs, and thus to the graphs of SCMs. The main idea of the acyclification is that under the condition that the SCM is uniquely solvable w.r.t. each strongly connected component, we can replace the causal mechanisms of these strongly connected components by their measurable solution functions, which results in an acyclic SCM. This acyclification preserves the solutions, and d-separation in the acyclification can directly be translated into σ-separation in the original graph. This then leads to the general directed global Markov property. We will discuss this now in more detail.

EXAMPLE A.10 (Construction of an observationally equivalent acyclic SCM). *Consider the SCM $\mathcal{M}$ of Example A.8 which is uniquely solvable w.r.t. all its strongly connected components, i.e., the subsets $\{1\}$, $\{2\}$ and $\{3, 4\}$. Replacing the causal mechanisms of these strongly connected components by their measurable solution functions gives the SCM $\tilde{\mathcal{M}}$ that is the same as $\mathcal{M}$ except that its causal mechanism $\tilde{\boldsymbol{f}}$ is given by*

$$\tilde{f}_1(\boldsymbol{x},\boldsymbol{e}) := e_1, \quad \tilde{f}_2(\boldsymbol{x},\boldsymbol{e}) := e_2, \quad \tilde{f}_3(\boldsymbol{x},\boldsymbol{e}) := \tfrac{x_1e_4+e_3}{1-x_1x_2}, \quad \tilde{f}_4(\boldsymbol{x},\boldsymbol{e}) := \tfrac{x_2e_3+e_4}{1-x_1x_2}\,.$$

By construction, $\mathcal{M}$ and $\tilde{\mathcal{M}}$ are observationally equivalent. Because $\tilde{\mathcal{M}}$ is acyclic (see Figure 1 on the right) we can apply the directed global Markov property to $\tilde{\mathcal{M}}$. The fact that X_1 and X_2 are not d-separated given $\{X_3, X_4\}$ in $\mathcal{G}(\tilde{\mathcal{M}})$ is in line with X_1 being dependent of X_2 given $\{X_3, X_4\}$ for every solution $\boldsymbol{X}$ of $\tilde{\mathcal{M}}$ (and hence of $\mathcal{M}$).

One of the key insights in [8] is that this example can easily be generalized as follows.

DEFINITION A.11 (Acyclification of an SCM). *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ be an SCM that is uniquely solvable w.r.t. each strongly connected component of $\mathcal{G}(\mathcal{M})$. For each $i \in \mathcal{I}$, let g_i be the i^{th} component of a measurable solution function $\boldsymbol{g}_{\mathrm{sc}(i)} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathrm{sc}(i))\setminus \mathrm{sc}(i)} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathrm{sc}(i))} \to \boldsymbol{\mathcal{X}}_{\mathrm{sc}(i)}$ of $\mathcal{M}$ w.r.t.* $\mathrm{sc}(i)$*, where* pa *and* sc *denote the parents and strongly connected components according to $\mathcal{G}^a(\mathcal{M})$, respectively. We call the SCM $\mathcal{M}^{\mathrm{acy}} := \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \hat{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ with the* acyclified causal mechanism $\hat{\boldsymbol{f}} : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ *given by*

$$\hat{f}_i(\boldsymbol{x}, \boldsymbol{e}) = g_i(\boldsymbol{x}_{\mathrm{pa}(\mathrm{sc}(i))\setminus \mathrm{sc}(i)}, \boldsymbol{e}_{\mathrm{pa}(\mathrm{sc}(i))}), \quad i \in \mathcal{I},$$

an acyclification of $\mathcal{M}$. *We denote by* $\mathrm{acy}(\mathcal{M})$ *the equivalence class of the acyclifications of $\mathcal{M}$.*

Note that $\mathrm{acy}(\mathcal{M})$ is well-defined: all acyclifications of an SCM $\mathcal{M}$ belong to the same equivalence class of SCMs.

PROPOSITION A.12. *Let $\mathcal{M}$ be an SCM that is uniquely solvable w.r.t. each strongly connected component of $\mathcal{G}(\mathcal{M})$. Then an acyclification $\mathcal{M}^{\mathrm{acy}}$ of $\mathcal{M}$ is acyclic and observationally equivalent to $\mathcal{M}$.*

We can also define a graphical acyclification for directed mixed graphs, which is a special case of the operation defined in [8] for HEDGes.

DEFINITION A.13 (Acyclification of a directed mixed graph). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph. The* acyclification of $\mathcal{G}$ *maps $\mathcal{G}$ to the* acyclified graph $\mathcal{G}^{\mathrm{acy}} := (\mathcal{V}, \hat{\mathcal{E}}, \hat{\mathcal{B}})$ *with directed edges $j \to i \in \hat{\mathcal{E}}$ if and only if $j \in \mathrm{pa}_{\mathcal{G}}(\mathrm{sc}_{\mathcal{G}}(i)) \setminus \mathrm{sc}_{\mathcal{G}}(i)$ and bidirected edges $i \leftrightarrow j \in \hat{\mathcal{B}}$ if and only if there exist $i' \in \mathrm{sc}_{\mathcal{G}}(i)$ and $j' \in \mathrm{sc}_{\mathcal{G}}(j)$ with $i' = j'$ or $i' \leftrightarrow j' \in \mathcal{B}$.*

The following compatibility result is immediate from the definitions.

PROPOSITION A.14. *Let $\mathcal{M}$ be an SCM that is uniquely solvable w.r.t. each strongly connected component of $\mathcal{G}(\mathcal{M})$. Then $\mathcal{G}^a(\mathrm{acy}(\mathcal{M})) \subseteq \mathrm{acy}(\mathcal{G}^a(\mathcal{M}))$ and $\mathcal{G}(\mathrm{acy}(\mathcal{M})) \subseteq \mathrm{acy}(\mathcal{G}(\mathcal{M}))$.*

The following example illustrates that the graph of the acyclification of an SCM can be a strict subgraph of the acyclification of the graph of the SCM.

EXAMPLE A.15 (Graph of the acyclification of the SCM is a strict subgraph of the acyclification of its graph). *Consider the SCM $\mathcal{M} = \langle \mathbf{2}, \mathbf{1}, \mathbb{R}^2, \mathbb{R}, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}} \rangle$ with the causal mechanism defined by*

$$f_1(\boldsymbol{x}, e) = x_2 - e, \quad f_2(\boldsymbol{x}, e) = \tfrac{1}{2}x_1 + e$$

and $\mathbb{P}_{\mathbb{R}}$ the standard Gaussian measure on $\mathbb{R}$. The SCM $\mathcal{M}$ is uniquely solvable w.r.t. the (only) strongly connected component $\{1,2\}$. An acyclification of $\mathcal{M}$ is the acyclified SCM $\mathcal{M}^{\mathrm{acy}}$ with the acyclified causal mechanism $\hat{\boldsymbol{f}}$ defined by

$$\hat{f}_1(\boldsymbol{x}, e) = 0, \quad \hat{f}_2(\boldsymbol{x}, e) = e.$$

The graph $\mathcal{G}(\mathrm{acy}(\mathcal{M}))$ is a strict subgraph of $\mathrm{acy}(\mathcal{G}(\mathcal{M}))$ *as can be seen in Figure 2.*

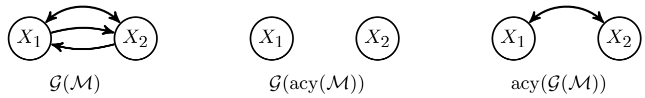

Fig 2: The graphs of the original SCM $\mathcal{M}$ (left), of the acyclified SCM (center), and of the acyclification of the graph of $\mathcal{M}$ (right) corresponding to Example A.15.

Translating the notion of d-separation from the acyclified graph back to the original graph led to the notion of σ-separation.

DEFINITION A.16 (σ-separation [8]). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph and let $C \subseteq \mathcal{V}$ be a subset of nodes. A walk (path) $\pi = (i_0, \epsilon_1, i_1, \ldots, i_n)$ in $\mathcal{G}$ is said to be C-σ-*blocked or σ-blocked by C *if*

1. *its first node $i_0 \in C$ or its last node $i_n \in C$, or*
2. *it contains a collider $i_k \notin \text{an}_{\mathcal{G}}(C)$, or*
3. *it contains a non-endpoint non-collider $i_k \in C$ that points towards a neighboring node on π that lies in a different strongly connected component of $\mathcal{G}$, that is, such that $i_{k-1} \leftarrow i_k$ in π and $i_{k-1} \notin \text{sc}_{\mathcal{G}}(i_k)$, or $i_k \to i_{k+1}$ in π and $i_{k+1} \notin \text{sc}_{\mathcal{G}}(i_k)$.*

*The walk (path) π is said to be C-σ-*open *if it is not σ-blocked by C. For two subsets of nodes $A, B \subseteq \mathcal{V}$, we say that A* is σ-separated from B given C in $\mathcal{G}$ *if all paths between any node in A and any node in B are σ-blocked by C, and write*

$$A \overset{\sigma}{\underset{\mathcal{G}}{\perp}} B \,|\, C \,.$$

The only difference between σ-separation and d-separation is that d-separation does not have the extra condition on the non-collider that it has to point to a node in a different strongly connected component. It is therefore obvious that σ-separation reduces to d-separation for acyclic graphs, since $\text{sc}_{\mathcal{G}}(i) = \{i\}$ for each $i \in \mathcal{V}$ in that case.

Although for proofs it is often easier to make use of walks, it suffices to formulate σ-separation in term of paths rather than walks because of the following result, which is analogous to a similar result for d-separation (see Lemma A.5).

LEMMA A.17. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph, $C \subseteq \mathcal{V}$ and $i, j \in \mathcal{V}$. There exists a C-σ-open walk between i and j in $\mathcal{G}$ if and only if there exists a C-σ-open path between i and j in $\mathcal{G}$.*

It is clear from the definitions that σ-separation implies d-separation. The other way around does not hold in general, as can be seen in the following example.

EXAMPLE A.18 (d-separation does not imply σ-separation). *Consider the directed graph $\mathcal{G}$ as depicted in Figure 1 (left). Here X_1 is d-separated from X_2 given $\{X_3, X_4\}$, but X_1 is not σ-separated from X_2 given $\{X_3, X_4\}$.*

The following result in [8] relates σ-separation to d-separation.

PROPOSITION A.19. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph. Then for $A, B, C \subseteq \mathcal{V}$,*

$$A \overset{\sigma}{\underset{\mathcal{G}}{\perp}} B \,|\, C \iff A \overset{d}{\underset{\text{acy}(\mathcal{G})}{\perp}} B \,|\, C \,.$$

By replacing in Definition A.6 "d-separation" by "σ-separation", one obtains the formulation of what Forré and Mooij [8] termed the general directed global Markov property.

DEFINITION A.20 (General directed global Markov property [8]). *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ be a directed mixed graph and $\mathbb{P}_{\mathcal{V}}$ a probability distribution on $\boldsymbol{\mathcal{X}}_{\mathcal{V}} = \prod_{i\in\mathcal{V}} \mathcal{X}_i$, where each $\mathcal{X}_i$ is a standard probability space. The probability distribution $\mathbb{P}_{\mathcal{V}}$ satisfies the* general directed global Markov property *relative to $\mathcal{G}$ if for all subsets $A, B, C \subseteq \mathcal{V}$ we have*

$$A \underset{\mathcal{G}}{\overset{\sigma}{\perp}} B \,|\, C \quad \Longrightarrow \quad \boldsymbol{X}_A \underset{\mathbb{P}_{\mathcal{V}}}{\perp\!\!\!\perp} \boldsymbol{X}_B \,|\, \boldsymbol{X}_C \,,$$

that is, $(X_i)_{i\in A}$ and $(X_i)_{i\in B}$ are conditionally independent given $(X_i)_{i\in C}$ under $\mathbb{P}_{\mathcal{V}}$, where we take the canonical projections $X_i : \boldsymbol{\mathcal{X}}_{\mathcal{V}} \to \mathcal{X}_i$ as random variables.

The fact that σ-separation implies d-separation means that the directed global Markov property implies the general directed global Markov property. In other words, the general directed global Markov property is weaker than the directed global Markov property. It is actually strictly weaker, as we saw in Example A.18.

The following fundamental result, also known as the σ-separation criterion, follows directly from the theory in [8].

THEOREM A.21 (General directed global Markov property for SCMs). *Let $\mathcal{M}$ be an SCM that is uniquely solvable w.r.t. each strongly connected component of $\mathcal{G}(\mathcal{M})$. Then its observational distribution $\mathbb{P}^{\boldsymbol{X}}$ exists, is unique and it satisfies the general directed global Markov property relative to $\mathcal{G}(\mathcal{M})$.*[1]

The proof is based on the reasoning that, for $A, B, C \subseteq \mathcal{I}$, if A is σ-separated from B given C in $\mathcal{G}(\mathcal{M})$, then A is d-separated from B by C in $\mathrm{acy}(\mathcal{G}(\mathcal{M}))$ and hence in $\mathcal{G}(\mathrm{acy}(\mathcal{M}))$, and since $\mathrm{acy}(\mathcal{M})$ is acyclic and observationally equivalent to $\mathcal{M}$, it follows from the directed global Markov property applied to $\mathrm{acy}(\mathcal{M})$ that $\boldsymbol{X}_A \perp\!\!\!\perp_{\mathbb{P}^{\boldsymbol{X}}} \boldsymbol{X}_B \,|\, \boldsymbol{X}_C$ for every solution $\boldsymbol{X}$ of $\mathcal{M}$. Note that the ancestral unique solvability condition for the discrete case is strictly weaker than the condition of unique solvability w.r.t. each strongly connected component in Theorem A.21. For the linear case, the condition of unique solvability is equivalent to the condition of unique solvability w.r.t. each strongly connected component (see Proposition C.4).

The results in Theorems A.7 and A.21 are not preserved under perfect intervention, because intervening on a strongly connected component could split it into several strongly connected components with different solvability properties. As the class of simple SCMs is preserved under perfect intervention and the twin operation (Proposition 8.2), we obtain the following corollary.

COROLLARY A.22 (Global Markov properties for simple SCMs). *Let $\mathcal{M}$ be a simple SCM. Then the:*

1. *observational distribution,*

[1] Since [8] also provides results under the weaker condition that an SCM is solvable (not necessarily uniquely) w.r.t. each strongly connected component of $\mathcal{G}(\mathcal{M})$, one might believe that Theorem A.21 could be generalized to stating that in that case, any of its observational distributions satisfies the general directed global Markov property. However, that is not true: consider for example the SCM $\mathcal{M} = \langle \mathbf{2}, \emptyset, \mathbb{R}^2, \mathbf{1}, \boldsymbol{f}, \mathbb{P}_{\mathbf{1}} \rangle$ with $f_1(\boldsymbol{x}) = x_1$ and $f_2(\boldsymbol{x}) = x_2$. Then $\mathcal{M}$ is solvable w.r.t. each of its strongly connected components $\{1\}$ and $\{2\}$. The solution with $X_1 = X_2$ shows a dependence between X_1 and X_2 and thus $X_1 \perp\!\!\!\perp X_2$ does not hold. In general, all strongly connected components that admit multiple solutions may be dependent on any other variable(s) in the model.

2. *interventional distribution after perfect intervention on* $I \subset \mathcal{I}$,
3. *counterfactual distribution after perfect intervention on* $\tilde{I} \subseteq \mathcal{I} \cup \mathcal{I}'$,

all exist, are unique and satisfy the general directed global Markov property relative to $\mathcal{G}(\mathcal{M})$, $\mathrm{do}(I)(\mathcal{G}(\mathcal{M}))$ *and* $\mathrm{do}(\tilde{I})(\mathrm{twin}(\mathcal{G}(\mathcal{M})))$, *respectively. Moreover, if* $\mathcal{M}$ *satisfies at least one of the three conditions (1), (2), (3) of Theorem A.7, then they also satisfies the directed global Markov property relative to* $\mathcal{G}(\mathcal{M})$, $\mathrm{do}(I)(\mathcal{G}(\mathcal{M}))$ *and* $\mathrm{do}(\tilde{I})(\mathrm{twin}(\mathcal{G}(\mathcal{M})))$, *respectively.*

Similar to d-faithfulness, σ-faithfulness[2] is defined as follows.

DEFINITION A.23 (σ-Faithfulness). *Let* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{B})$ *be a directed mixed graph and* $\mathbb{P}_{\mathcal{V}}$ *a probability distribution on* $\boldsymbol{\mathcal{X}}_{\mathcal{V}} = \prod_{i \in \mathcal{V}} \mathcal{X}_i$, *where each* $\mathcal{X}_i$ *is a standard probability space. The probability distribution* $\mathbb{P}_{\mathcal{V}}$ *is* σ-faithful *to* $\mathcal{G}$ *if for all subsets* $A, B, C \subseteq \mathcal{V}$ *we have*

$$A \underset{\mathcal{G}}{\overset{\sigma}{\perp}} B \mid C \quad \Longleftarrow \quad \boldsymbol{X}_A \underset{\mathbb{P}_{\mathcal{V}}}{\perp\!\!\!\perp} \boldsymbol{X}_B \mid \boldsymbol{X}_C \,,$$

where we take the canonical projections $X_i : \boldsymbol{\mathcal{X}}_{\mathcal{V}} \to \mathcal{X}_i$ *as random variables.*

In other words, the graph explains, via σ-separation, all the conditional independencies that are present in the observational distribution. Although it has been conjectured [27] that under certain conditions σ-faithfulness should hold, formulating and proving such completeness results is an open problem to the best of our knowledge.

A.3. Modular SCMs. In this subsection, we relate the class of (simple) SCMs to that of modular SCMs. Modular SCMs introduced by Forré and Mooij [8] are causal graphical models on which marginalizations and interventions are defined and they satisfy the general directed global Markov property. For a comprehensive account on modular SCMs we refer the reader to [8].

A.3.1. *Definition of a modular SCM.* In contrast to an SCM from which a graph can be derived, a modular SCM is defined in terms of a graphical object, which Forré and Mooij [8] call a directed graph with hyperedges (HEDG). The hyperedges of a HEDG are described in terms of a simplicial complex.

DEFINITION A.24 (Simplicial complex). *Let* $\mathcal{V}$ *be a finite set. A* simplicial complex $\mathcal{H}$ *over* $\mathcal{V}$ *is a set of subsets of* $\mathcal{V}$ *such that*

1. *all single element sets* $\{v\}$ *are in* $\mathcal{H}$ *for* $v \in \mathcal{V}$, *and*
2. *if* $\mathcal{F} \in \mathcal{H}$, *then also all subsets* $\tilde{\mathcal{F}} \subseteq F$ *are elements of* $\mathcal{H}$.

DEFINITION A.25 (Directed graph with hyperedges (HEDGes) [8]). *A* directed graph with hyperedges (HEDG) *is a triple* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{H})$, *where* $(\mathcal{V}, \mathcal{E})$ *is a directed graph and* $\mathcal{H}$ *a simplicial complex over the set of nodes* $\mathcal{V}$. *The elements* $\mathcal{F}$ *of* $\mathcal{H}$ *are called* hyperedges *of* $\mathcal{G}$. *The elements* $\mathcal{F}$ *of* $\mathcal{H}$ *that are inclusion-maximal elements of* $\mathcal{H}$ *are called* maximal hyperedges *and are denoted by* $\hat{\mathcal{H}}$.

[2] In [24] it is called "collapsed graph faithfulness".

A HEDG $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{H})$ can be represented as a directed graph $\bar{\mathcal{G}}:=(\mathcal{V},\mathcal{E})$ consisting of nodes $\mathcal{V}$ and directed edges $\mathcal{E}$, with additional maximal hyperedges $\mathcal{F}\in\hat{\mathcal{H}}$ with $|\mathcal{F}|\geq 2$ (i.e., not corresponding to single element sets $\{v\}\in\hat{\mathcal{H}}$), that point to their target nodes $v\in\mathcal{F}$. For a HEDG $\mathcal{G}$, we define $\mathrm{pa}_{\mathcal{G}}$, $\mathrm{ch}_{\mathcal{G}}$, etc., in terms of the underlying directed graph $\bar{\mathcal{G}}$, that is, $\mathrm{pa}_{\bar{\mathcal{G}}}$, $\mathrm{ch}_{\bar{\mathcal{G}}}$, etc., respectively.

A *loop* in a HEDG $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{H})$ is a subset $\mathcal{O}\subseteq\mathcal{V}$ that is a loop in the underlying directed graph $\bar{\mathcal{G}}=(\mathcal{V},\mathcal{E})$. In other words, a loop of $\mathcal{G}$ is a set of nodes $\mathcal{O}\subseteq\mathcal{V}$ such that for every two nodes $v,w\in\mathcal{O}$ there are directed paths $v\to\cdots\to w$ and $w\to\cdots\to v$ in $\mathcal{G}$ for which all the intermediate nodes lie in $\mathcal{O}$ (if any exist). In particular, a loop may consist of a single element $\{v\}$ for $v\in\mathcal{V}$. The set of loops in $\mathcal{G}$ is denoted by $\mathcal{L}(\mathcal{G})$.

In order to define a modular SCM one needs the notion of a compatible system of solution functions, which assigns to each loop a separate solution function such that all these solution functions are "compatible" with each other.

DEFINITION A.26 (Compatible system of solution functions[3]). *Let $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{H})$ be a HEDG. For every $v\in\mathcal{V}$ and maximal hyperedge $\mathcal{F}$ in $\hat{\mathcal{H}}$, let $\mathcal{X}_v$ and $\mathcal{E}_{\mathcal{F}}$ be standard measurable spaces. For a subset $\mathcal{O}\subseteq\mathcal{V}$ we define*[4]

$$\boldsymbol{\mathcal{X}}_{\mathcal{O}}:=\prod_{v\in\mathcal{O}}\mathcal{X}_v \quad \text{and} \quad \widehat{\boldsymbol{\mathcal{E}}}_{\mathcal{O}}:=\prod_{\substack{\mathcal{F}\in\hat{\mathcal{H}}\\ \mathcal{F}\cap\mathcal{O}\neq\emptyset}}\mathcal{E}_{\mathcal{F}}\,.$$

Consider a family of measurable mappings $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}$ indexed by $\mathcal{L}(\mathcal{G})$ which are of the form

$$\boldsymbol{g}_{\mathcal{O}}:\boldsymbol{\mathcal{X}}_{\mathrm{pa}_{\mathcal{G}}(\mathcal{O})\setminus\mathcal{O}}\times\widehat{\boldsymbol{\mathcal{E}}}_{\mathcal{O}}\to\boldsymbol{\mathcal{X}}_{\mathcal{O}}\,.$$

We call the family of measurable mappings $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}$ a compatible system of solution functions*, if for all $\mathcal{O},\tilde{\mathcal{O}}\in\mathcal{L}(\mathcal{G})$ with $\tilde{\mathcal{O}}\subseteq\mathcal{O}$ and for all $\widehat{\boldsymbol{e}}_{\mathcal{O}}\in\widehat{\boldsymbol{\mathcal{E}}}_{\mathcal{O}}$ and $\boldsymbol{x}_{\mathrm{pa}_{\mathcal{G}}(\mathcal{O})\cup\mathcal{O}}\in\boldsymbol{\mathcal{X}}_{\mathrm{pa}_{\mathcal{G}}(\mathcal{O})\cup\mathcal{O}}$ we have*

$$\boldsymbol{x}_{\mathcal{O}}=\boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}_{\mathcal{G}}(\mathcal{O})\setminus\mathcal{O}},\widehat{\boldsymbol{e}}_{\mathcal{O}}) \quad\Longrightarrow\quad \boldsymbol{x}_{\tilde{\mathcal{O}}}=\boldsymbol{g}_{\tilde{\mathcal{O}}}(\boldsymbol{x}_{\mathrm{pa}_{\mathcal{G}}(\tilde{\mathcal{O}})\setminus\tilde{\mathcal{O}}},\widehat{\boldsymbol{e}}_{\tilde{\mathcal{O}}})\,.$$

This structure of a compatible system of solution functions is at the heart of the defnition of a modular SCM.

DEFINITION A.27 (Modular structural causal model (mSCM) [8]). *A* modular structural causal model (mSCM) *is a tuple*

$$\widehat{\mathcal{M}}:=\langle\mathcal{G},\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{E}},(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})},\mathbb{P}_{\boldsymbol{\mathcal{E}}}\rangle\,,$$

where

1. *$\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{H})$ is a HEDG,*
2. *$\boldsymbol{\mathcal{X}}=\prod_{v\in\mathcal{V}}\mathcal{X}_v$ is the product of standard measurable spaces $\mathcal{X}_v$,*
3. *$\boldsymbol{\mathcal{E}}=\prod_{\mathcal{F}\in\hat{\mathcal{H}}}\mathcal{E}_{\mathcal{F}}$ is the product of standard measurable spaces $\mathcal{E}_{\mathcal{F}}$,*
4. *$(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}$ is a compatible system of solution functions,*
5. *$\mathbb{P}_{\boldsymbol{\mathcal{E}}}=\prod_{\mathcal{F}\in\hat{\mathcal{H}}}\mathbb{P}_{\mathcal{E}_{\mathcal{F}}}$ is a product measure, where $\mathbb{P}_{\mathcal{E}_{\mathcal{F}}}$ is a probability measure on $\mathcal{E}_{\mathcal{F}}$ for each $\mathcal{F}\in\hat{\mathcal{H}}$.*

---

[3] We deviate from the terminology in [8] where this is called a "compatible system of structural equations".

[4] We use the "hat" notation $\widehat{\boldsymbol{\mathcal{E}}}_{\mathcal{O}}$ to distinguish it from the ordinary subscript convention that $\boldsymbol{\mathcal{E}}_{\mathcal{O}}=\prod_{\mathcal{F}\in\mathcal{O}}\mathcal{E}_{\mathcal{F}}$ for some subset $\mathcal{O}\subseteq\hat{\mathcal{H}}$.

Let $\widehat{\mathcal{M}} = \langle \mathcal{G}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}, \mathbb{P}_{\boldsymbol{\mathcal{E}}}\rangle$ be a modular SCM and $\mathcal{O}_1, \dots, \mathcal{O}_r \in \mathcal{L}(\mathcal{G})$ the strongly connected components of $\mathcal{G}$ ordered according to a topological order of the DAG of strongly connected components of $\mathcal{G}$. Then for any random variable $\boldsymbol{E} : \Omega \to \boldsymbol{\mathcal{E}}$ such that $\mathbb{P}^{\boldsymbol{E}} = \mathbb{P}_{\boldsymbol{\mathcal{E}}}$ one can inductively define the random variables $X_v := (\boldsymbol{g}_{\mathcal{O}_i})_v(\boldsymbol{X}_{\mathrm{pa}_{\mathcal{G}}(\mathcal{O}_i)\setminus\mathcal{O}_i}, \widehat{\boldsymbol{E}}_{\mathcal{O}_i})$ for all $v \in \mathcal{O}_i$ for all $i \geq 1$, starting at $X_v := (\boldsymbol{g}_{\mathcal{O}_1})_v(\widehat{\boldsymbol{E}}_{\mathcal{O}_1})$ for all $v \in \mathcal{O}_1$. Because $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}$ is a compatible system of solution functions, we have for every $\mathcal{O} \in \mathcal{L}(\mathcal{G})$

$$\boldsymbol{X}_{\mathcal{O}} = \boldsymbol{g}_{\mathcal{O}}(\boldsymbol{X}_{\mathrm{pa}_{\mathcal{G}}(\mathcal{O})\setminus\mathcal{O}}, \widehat{\boldsymbol{E}}_{\mathcal{O}}).$$

We call the random variable $\boldsymbol{X}$ a *solution* of the modular SCM $\widehat{\mathcal{M}}$. Note that the solution $\boldsymbol{X}$ depends on the choice of the random variable $\boldsymbol{E} : \Omega \to \boldsymbol{\mathcal{E}}$.

The causal semantics of modular SCMs can be defined in terms of perfect interventions, which is defined as follows.

DEFINITION A.28 (Perfect intervention on an mSCM). *Consider a modular SCM* $\widehat{\mathcal{M}} = \langle \mathcal{G}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}, \mathbb{P}_{\boldsymbol{\mathcal{E}}}\rangle$*, a subset* $I \subseteq \mathcal{V}$ *of endogenous variables and a value* $\boldsymbol{\xi}_I \in \boldsymbol{\mathcal{X}}_I$*. The* perfect intervention $\mathrm{do}(I, \boldsymbol{\xi}_I)$ *maps* $\widehat{\mathcal{M}}$ *to the modular SCM*

$$\widehat{\mathcal{M}}_{\mathrm{do}(I,\boldsymbol{\xi}_I)} := \langle \mathcal{G}^{\mathrm{do}}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}^{\mathrm{do}}, (\boldsymbol{g}^{\mathrm{do}}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G}^{\mathrm{do}})}, \mathbb{P}_{\boldsymbol{\mathcal{E}}^{\mathrm{do}}}\rangle,$$

where

1. $\mathcal{G}^{\mathrm{do}} = (\mathcal{V}, \mathcal{E}^{\mathrm{do}}, \mathcal{H}^{\mathrm{do}})$*, where*

$$\mathcal{E}^{\mathrm{do}} = \mathcal{E} \setminus \{v \to w : v \in \mathcal{V}, w \in I\}$$

$$\mathcal{H}^{\mathrm{do}} = \{\mathcal{F} \setminus I : \mathcal{F} \in \mathcal{H}\} \cup \{\{v\} : v \in I\},$$

2. $\phi : \{\mathcal{F} \in \hat{\mathcal{H}} : \mathcal{F} \setminus I \neq \emptyset\} \to \hat{\mathcal{H}}^{\mathrm{do}} \setminus \{\{v\} : v \in I\}$ *is a mapping such that* $\phi(\mathcal{F}) \supseteq \mathcal{F} \setminus I$ *for all* $\mathcal{F} \in \hat{\mathcal{H}}$ *for which* $\mathcal{F} \setminus I \neq \emptyset$*,*
3. $\boldsymbol{\mathcal{E}}^{\mathrm{do}} = \prod_{\tilde{\mathcal{F}}\in\hat{\mathcal{H}}^{\mathrm{do}}} \mathcal{E}^{\mathrm{do}}_{\tilde{\mathcal{F}}}$*, where*

$$\mathcal{E}^{\mathrm{do}}_{\tilde{\mathcal{F}}} = \begin{cases} \mathcal{X}_v & \text{if } \tilde{\mathcal{F}} = \{v\} \text{ for } v \in I \\ \prod_{\mathcal{F}=\phi^{-1}(\tilde{\mathcal{F}})} \mathcal{E}_{\mathcal{F}} & \text{if } \tilde{\mathcal{F}} \in \hat{\mathcal{H}}^{\mathrm{do}} \setminus \{\{v\} : v \in I\}, \end{cases}$$

4. *for every* $\mathcal{O} \in \mathcal{L}(\mathcal{G}^{\mathrm{do}})$

$$\boldsymbol{g}^{\mathrm{do}}_{\mathcal{O}} = \begin{cases} \mathbb{I}_{\{v\}} & \text{if } \mathcal{O} = \{v\} \text{ for } v \in I \\ \boldsymbol{g}_{\mathcal{O}} & \text{otherwise,} \end{cases}$$

   (note that if $\mathcal{O}$ *is a loop in* $\mathcal{G}^{\mathrm{do}}$*, then it is a loop in* $\mathcal{G}$*),*
5. $\mathbb{P}_{\boldsymbol{\mathcal{E}}^{\mathrm{do}}} = \prod_{\tilde{\mathcal{F}}\in\hat{\mathcal{H}}^{\mathrm{do}}} \mathbb{P}_{\mathcal{E}^{\mathrm{do}}_{\tilde{\mathcal{F}}}}$*, where*

$$\mathbb{P}_{\mathcal{E}^{\mathrm{do}}_{\tilde{\mathcal{F}}}} = \begin{cases} \delta_{\xi_v} & \text{if } \tilde{\mathcal{F}} = \{v\} \text{ for } v \in I \\ \prod_{\mathcal{F}=\phi^{-1}(\tilde{\mathcal{F}})} \mathbb{P}_{\mathcal{E}_{\mathcal{F}}} & \text{if } \tilde{\mathcal{F}} \in \hat{\mathcal{H}}^{\mathrm{do}} \setminus \{\{v\} : v \in I\}. \end{cases}$$

In contrast to SCMs, these perfect interventions on modular SCMs are directly defined on the underlying HEDG and depend on the choice of the mapping ϕ.

A.3.2. *Relation between SCMs and modular SCMs.* The solutions of a modular SCM can be described by an SCM that is loop-wisely solvable.

DEFINITION A.29 (Underlying SCM). *Let $\widehat{\mathcal{M}} = \langle \mathcal{G}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}, \mathbb{P}_{\boldsymbol{\mathcal{E}}}\rangle$ be a modular SCM. Then the mapping ι maps $\widehat{\mathcal{M}}$ to the* underlying SCM $\tilde{\mathcal{M}} := \langle \tilde{\mathcal{I}}, \tilde{\mathcal{J}}, \tilde{\boldsymbol{\mathcal{X}}}, \tilde{\boldsymbol{\mathcal{E}}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}\rangle$, *where*

1. $\tilde{\mathcal{I}} = \mathcal{V}$,
2. $\tilde{\mathcal{J}} = \hat{\mathcal{H}}$,
3. $\tilde{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{X}}$,
4. $\tilde{\boldsymbol{\mathcal{E}}} = \boldsymbol{\mathcal{E}}$,
5. $\tilde{\boldsymbol{f}}$ *is given by* $\tilde{f}_v = (\boldsymbol{g}_{\{v\}})_v$ *for all* $v \in \mathcal{V}$,
6. $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}} = \mathbb{P}_{\boldsymbol{\mathcal{E}}}$.

Every solution $\boldsymbol{X}$ of a modular SCM $\widehat{\mathcal{M}}$ is also a solution of the underlying SCM $\iota(\widehat{\mathcal{M}})$.

Observe that for the modular SCM $\widehat{\mathcal{M}}$ we have that the induced subgraph $\mathcal{G}^a(\iota(\widehat{\mathcal{M}}))_{\tilde{\mathcal{I}}}$, of the augmented graph of the underlying SCM $\mathcal{G}^a(\iota(\widehat{\mathcal{M}}))$ on $\tilde{\mathcal{I}}$, is a subgraph of the underlying HEDG $\mathcal{G}$, that is, $\mathcal{G}^a(\iota(\widehat{\mathcal{M}}))_{\tilde{\mathcal{I}}} \subseteq \mathcal{G}$. This implies that, in general, the underlying HEDG $\mathcal{G}$ of $\widehat{\mathcal{M}}$ may have more loops than the loops in $\mathcal{G}(\iota(\widehat{\mathcal{M}}))$. For a subset $\mathcal{O} \subseteq \tilde{\mathcal{I}}$, we have for the exogenous parents of the underlying SCM $\iota(\widehat{\mathcal{M}})$

$$\mathrm{pa}(\mathcal{O}) \cap \tilde{\mathcal{J}} \subseteq \{\mathcal{F} \in \tilde{\mathcal{J}} : \mathcal{F} \cap \mathcal{O} \neq \emptyset\},$$

where $\mathrm{pa}(\mathcal{O})$ denotes the set of parents of $\mathcal{O}$ in $\mathcal{G}^a(\iota(\widehat{\mathcal{M}}))$. Hence, in general, not all the hyperedges $\mathcal{F} \in \mathcal{H}$ such that $|\mathcal{F}| = 2$ (i.e., bidirected edges) are in the set of bidirected edges $\mathcal{B}$ of the graph of the underlying SCM $\mathcal{G}(\iota(\widehat{\mathcal{M}})) = (\mathcal{V}, \mathcal{E}, \mathcal{B})$. We conclude that the graph of the underlying SCM is, in general, a sparser graph than the HEDG of the modular SCM.

Next, we show that the compatible system of solution functions of a modular SCM induces a compatible system of solution functions on the underlying SCM. For this we need the notion of loop-wise solvability for SCMs.

DEFINITION A.30 (Loop-wise (unique) solvability for SCMs). *We call an SCM $\mathcal{M}$*

1. loop-wisely solvable, *if $\mathcal{M}$ is solvable w.r.t. every loop $\mathcal{O} \in \mathcal{L}(\mathcal{G}(\mathcal{M}))$, and*
2. loop-wisely uniquely solvable, *if $\mathcal{M}$ is uniquely solvable w.r.t. every loop $\mathcal{O} \in \mathcal{L}(\mathcal{G}(\mathcal{M}))$.*

DEFINITION A.31 (Compatible system of solution functions for SCMs). *For a loop-wisely solvable SCM $\mathcal{M}$, we call a family of measurable solution functions $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G}(\mathcal{M}))}$, where $\boldsymbol{g}_{\mathcal{O}}$ is a measurable solution function of $\mathcal{M}$ w.r.t. $\mathcal{O}$, a* compatible system of solution functions, *if for all $\mathcal{O}, \tilde{\mathcal{O}} \in \mathcal{L}(\mathcal{G}(\mathcal{M}))$ with $\tilde{\mathcal{O}} \subseteq \mathcal{O}$ and for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\mathbf{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ we have*

$$\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \quad \Longrightarrow \quad \boldsymbol{x}_{\tilde{\mathcal{O}}} = \boldsymbol{g}_{\tilde{\mathcal{O}}}(\boldsymbol{x}_{\mathrm{pa}(\tilde{\mathcal{O}})\setminus\tilde{\mathcal{O}}}, \boldsymbol{e}_{\mathrm{pa}(\tilde{\mathcal{O}})}).$$

The underlying SCM of a modular SCM always has a compatible system of solution functions, by construction.

PROPOSITION A.32. *Let $\widehat{\mathcal{M}} = \langle \mathcal{G}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}, \mathbb{P}_{\boldsymbol{\mathcal{E}}}\rangle$ be a modular SCM. Then the underlying SCM $\tilde{\mathcal{M}} := \iota(\widehat{\mathcal{M}})$ is loop-wisely solvable. Moreover, it has a compatible system of solution functions $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G}(\tilde{\mathcal{M}}))}$, where $\boldsymbol{g}_{\mathcal{O}}$ is a measurable solution function of $\tilde{\mathcal{M}}$ w.r.t. $\mathcal{O}$.*

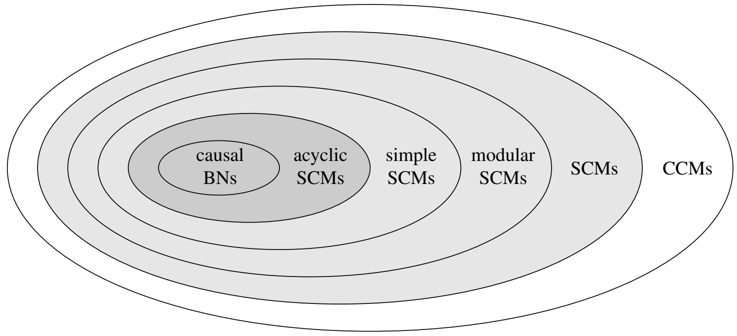

Fig 3: Overview of causal graphical models. The “gray” and “dark gray” areas contain all the causal graphical models that can be modeled by an SCM and an acyclic SCM, respectively.

This shows that a modular SCM can be seen as an SCM together with an additional structure of a compatible system of solution functions, and is, in particular, loop-wisely solvable.

Moreover, the class of simple SCMs corresponds exactly with those SCMs that are loop-wisely uniquely solvable.

LEMMA A.33. *An SCM $\mathcal{M}$ is simple if and only if it is loop-wisely uniquely solvable.*

In particular, for simple SCMs, or loop-wisely uniquely solvable SCMs, there always exists a compatible system of solution functions.

PROPOSITION A.34. *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ be a simple SCM. Then every family of measurable solution functions $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O} \in \mathcal{L}(\mathcal{G}(\mathcal{M}))}$, where $\boldsymbol{g}_{\mathcal{O}}$ is a measurable solution function of $\mathcal{M}$ w.r.t. $\mathcal{O}$, is a compatible system of solution functions.*

A.4. Overview of causal graphical models. Figure 3 gives an overview of the causal graphical models related to SCMs. The “gray” area contains all the causal graphical models that can be modeled by an SCM, by which we mean, that there exists an SCM that can describe all its observational and interventional distributions. The “dark gray” area contains all the causal graphical models which can be modeled by an acyclic SCM. Acyclic SCMs generalize causal Bayesian networks (causal BNs) [20] to allow for latent confounders and to derive counterfactuals. Simple SCMs form a subclass of SCMs that extends acyclic SCMs to the cyclic setting, while preserving many of their convenient properties. Modular SCMs [8] can be seen as SCMs that have an additional structure of compatible system of solution functions and contain, in particular, the class of simple SCMs. Forré and Mooij [8] showed that modular SCMs satisfy various convenient properties, like marginalization and the general directed global Markov property. We show that for SCMs in general various of those properties still hold under certain solvability conditions. A generalization of SCMs, known as *causal constraints models (CCMs)*, has been proposed [1] in order to completely model the causal semantics of the equilibrium solutions of a dynamical system given the initial conditions. This class of CCMs is rich enough to model the causal semantics of SCMs, but does not come with a single graphical representation that provides both a Markov property and a causal interpretation [2].

APPENDIX B: (UNIQUE) SOLVABILITY PROPERTIES

In this appendix, we provide additional (unique) solvability properties for SCMs. In Appendix B.1 we provide a sufficient condition of solvability w.r.t. (strict) subsets. In Appendix B.2 we discuss how (unique) solvability is preserved under strict super- and subsets.

In Appendix B.3 we discuss how (unique) solvability is preserved under unions and intersections. The proofs of the theoretical results in this appendix are given in Appendix E.

B.1. Sufficient condition for solvability w.r.t. subsets. For solvability w.r.t. a (strict) subset of $\mathcal{I}$ there exists a sufficient condition that is similar to the sufficient (and necessary) condition (2) in Theorem 3.2 in the sense that it is formulated in terms of the solutions of (a subset of) the structural equations, but no measurability is required.

PROPOSITION B.1 (Sufficient condition for solvability w.r.t. a subset). *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ be an SCM and $\mathcal{O} \subseteq \mathcal{I}$ a subset. If for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\mathbf{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}_{\setminus \mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\setminus \mathcal{O}}$ the topological space*

$$\boldsymbol{\mathcal{S}}_{(\boldsymbol{e}, \boldsymbol{x}_{\setminus \mathcal{O}})} := \{ \boldsymbol{x}_{\mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\mathcal{O}} : \boldsymbol{x}_{\mathcal{O}} = \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e}) \} ,$$

with the subspace topology induced by $\boldsymbol{\mathcal{X}}_{\mathcal{O}}$ is nonempty and σ-compact,[5] then $\mathcal{M}$ is solvable w.r.t. $\mathcal{O}$.

For many purposes, this condition of σ-compactness suffices since it contains for example all countable discrete spaces, every interval of the real line, and moreover all the Euclidean spaces. In particular, it suffices to prove a sufficient and necessary condition for unique solvability w.r.t. a subset, in terms of the solutions of a subset of the structural equations (see Theorem 3.6). For larger solution spaces, we refer the reader to [12]. For the class of linear SCMs (see Definition C.1), we provide in Proposition C.2 a sufficient and necessary condition for solvability w.r.t. a (strict) subset of $\mathcal{I}$.

B.2. (Unique) solvability w.r.t. strict super- and subsets. In general, (unique) solvability w.r.t. $\mathcal{O} \subseteq \mathcal{I}$ does not imply (unique) solvability w.r.t. a strict superset $\mathcal{O} \subsetneq \mathcal{V} \subseteq \mathcal{I}$ nor w.r.t. a strict subset $\mathcal{W} \subsetneq \mathcal{O}$, as can be seen in the following example.

EXAMPLE B.2 (Solvability is not preserved under strict sub- or supersets). *Consider the SCM $\mathcal{M} = \langle \mathbf{3}, \emptyset, \mathbb{R}^3, \mathbf{1}, \boldsymbol{f}, \mathbb{P}_{\mathbf{1}} \rangle$ where the causal mechanism is given by*

$$f_1(\boldsymbol{x}) = x_1 \cdot (1 - \mathbf{1}_{\{1\}}(x_2)) + 1 \,, \; f_2(\boldsymbol{x}) = x_2 \,, \; f_3(\boldsymbol{x}) = x_3 \cdot (1 - \mathbf{1}_{\{-1\}}(x_2)) + 1 \,.$$

This SCM is (uniquely) solvable w.r.t. the subsets $\{1,2\}$, $\{2,3\}$, however it is not (uniquely) solvable w.r.t. the subsets $\{1\}$, $\{3\}$ and $\{1,2,3\}$, and not uniquely solvable w.r.t. $\{2\}$.

However, in Proposition 3.10 we show that solvability w.r.t. $\mathcal{O}$ implies solvability w.r.t. every ancestral subset in $\mathcal{G}(\mathcal{M})_{\mathcal{O}}$.

B.3. (Unique) solvability w.r.t. unions and intersections. In general, (unique) solvability is not preserved under unions and intersections. The following example illustrates that (unique) solvability is in general not preserved under intersections.

EXAMPLE B.3 (Solvability is not preserved under intersections). *Consider the SCM $\mathcal{M} = \langle \mathbf{3}, \emptyset, \mathbb{R}^3, \mathbf{1}, \boldsymbol{f}, \mathbb{P}_{\mathbf{1}} \rangle$ where the causal mechanism is given by*

$$f_1(\boldsymbol{x}) = 0 \,, \; f_2(\boldsymbol{x}) = x_2 \cdot (1 - \mathbf{1}_{\{0\}}(x_1 \cdot x_3)) + 1 \,, \; f_3(\boldsymbol{x}) = 0 \,.$$

Then $\mathcal{M}$ is (uniquely) solvable w.r.t. $\{1,2\}$ and $\{2,3\}$, however it is not (uniquely) solvable w.r.t. their intersection.

[5] A topological space $\boldsymbol{\mathcal{X}}$ is called *σ-compact* if it is the union of a countable set of compact topological spaces.

Example B.2 gives an example where (unique) solvability is not preserved under unions. Even, if we take the union of disjoint subsets, (unique) solvability is not preserved (see Example 2.4). Although, in general, unique solvability is not preserved under unions, we show next that unique solvability is preserved under the union of ancestral subsets, under the following assumptions.

PROPOSITION B.4 (Combining measurable solution functions on different sets). *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ be an SCM, $\mathcal{O} \subseteq \mathcal{I}$ a subset and $\mathcal{A}, \tilde{\mathcal{A}} \subseteq \mathcal{O}$ two ancestral subsets in $\mathcal{G}(\mathcal{M})_{\mathcal{O}}$. If $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{A}$, $\tilde{\mathcal{A}}$ and $\mathcal{A} \cap \tilde{\mathcal{A}}$, then $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{A} \cup \tilde{\mathcal{A}}$.*

A consequence of this property is that in order to check whether an SCM is ancestrally uniquely solvable w.r.t. $\mathcal{O}$, it suffices to check that it is uniquely solvable w.r.t. the ancestral subsets for each node in $\mathcal{O}$.

COROLLARY B.5. *Let $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ be an SCM and $\mathcal{O} \subseteq \mathcal{I}$ a subset. Then $\mathcal{M}$ is ancestrally uniquely solvable w.r.t. $\mathcal{O}$ if and only if $\mathcal{M}$ is uniquely solvable w.r.t. $\mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(i)$ for every $i \in \mathcal{O}$.*

APPENDIX C: LINEAR SCMS

In this appendix, we provide some results about (unique) solvability and marginalization for linear SCMs. Linear SCMs form a special class of SCMs that has seen much attention in the literature [see, e.g., 3, 11]. The proofs of the theoretical results in this appendix are given in Appendix E.

DEFINITION C.1 (Linear SCM). *We call an SCM $\mathcal{M} = \langle \mathcal{I}, \mathcal{J}, \mathbb{R}^{\mathcal{I}}, \mathbb{R}^{\mathcal{J}}, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^{\mathcal{J}}} \rangle$* linear *if each component of the causal mechanism is a linear combination of the endogenous and exogenous variables, that is*

$$f_i(\boldsymbol{x}, \boldsymbol{e}) = \sum_{j \in \mathcal{I}} B_{ij} x_j + \sum_{k \in \mathcal{J}} \Gamma_{ik} e_k \,,$$

where $i \in \mathcal{I}$, $B \in \mathbb{R}^{\mathcal{I} \times \mathcal{I}}$ and $\Gamma \in \mathbb{R}^{\mathcal{I} \times \mathcal{J}}$ are matrices, and $\mathbb{P}_{\mathbb{R}^{\mathcal{J}}}$ is a product probability measure[6] on $\mathbb{R}^{\mathcal{J}}$.

For a subset $\mathcal{O} \subseteq \mathcal{I}$ we also use the shorthand vector-notation

$$\boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e}) = B_{\mathcal{O}\mathcal{I}} \boldsymbol{x} + \Gamma_{\mathcal{O}\mathcal{J}} \boldsymbol{e} \,.$$

A nonzero coefficient B_{ij} for $i, j \in \mathcal{I}$ such that $i \neq j$ corresponds with a directed edge $j \to i$ in the (augmented) graph, and a coefficient $B_{ii} = 1$ for $i \in \mathcal{I}$ corresponds with a self-cycle $i \to i$ in the (augmented) graph of the SCM. A nonzero coefficient Γ_{ij} for $i \in \mathcal{I}$, $j \in \mathcal{J}$ with $\mathbb{P}_{\mathcal{E}_j}$ a nondegenerate probability distribution over $\mathbb{R}$ corresponds with a directed edge $j \to i$ in the augmented graph. A nonzero entry $(\Gamma\Gamma^T)_{ij}$ for $i, j \in \mathcal{I}$ with $i \neq j$ such that there exists a $k \in \mathcal{J}$ for which $\Gamma_{ik}, \Gamma_{jk} \neq 0$ and $\mathbb{P}_{\mathcal{E}_k}$ a nondegenerate probability distribution over $\mathbb{R}$ corresponds with a bidirected edge $i \leftrightarrow j$ in the graph of the SCM.

For linear SCMs, the solvability condition w.r.t. a subset, Definition 3.1, translates into a matrix condition. In order to state this condition we need to define the pseudoinverse (or the Moore-Penrose inverse) A^+ of a real matrix A [10, 22]. The *pseudoinverse of the matrix A* is defined by $A^+ := V\Sigma^+ U^*$, where $A = U\Sigma V^*$ is the singular value decomposition of A and Σ^+ is obtained by replacing each nonzero entry on the diagonal of Σ by its reciprocal [10]. One of its useful properties is that $AA^+A = A$.

[6] Note that we do not assume that the probability measure $\mathbb{P}_{\mathbb{R}^{\mathcal{J}}}$ is Gaussian.

PROPOSITION C.2 (Sufficient and necessary condition for solvability w.r.t. a subset for linear SCMs). *Let $\mathcal{M}$ be a linear SCM and $\mathcal{L} \subseteq \mathcal{I}$ and $\mathcal{O} = \mathcal{I} \setminus \mathcal{L}$. Then $\mathcal{M}$ is solvable w.r.t. $\mathcal{L}$ if and only if for the matrix $A_{\mathcal{L}\mathcal{L}} = \mathbb{I}_{\mathcal{L}} - B_{\mathcal{L}\mathcal{L}}$, for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}_{\mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ the identity*

$$A_{\mathcal{L}\mathcal{L}} A_{\mathcal{L}\mathcal{L}}^{+}(B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}} + \Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e}) = B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}} + \Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e}$$

is satisfied, where $A_{\mathcal{L}\mathcal{L}}^{+}$ is the pseudoinverse of $A_{\mathcal{L}\mathcal{L}}$. Moreover, if $\mathcal{M}$ is solvable w.r.t. $\mathcal{L}$, then for every vector $\boldsymbol{v} \in \mathbb{R}^{\mathcal{L}}$ the mapping $\boldsymbol{g}_{\mathcal{L}}^{\boldsymbol{v}} : \mathbb{R}^{\mathcal{O}} \times \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{L}}$ given by

$$\boldsymbol{g}_{\mathcal{L}}^{\boldsymbol{v}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) = A_{\mathcal{L}\mathcal{L}}^{+}(B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}} + \Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e}) + [\mathbb{I}_{\mathcal{L}} - A_{\mathcal{L}\mathcal{L}}^{+}A_{\mathcal{L}\mathcal{L}}]\boldsymbol{v}\,,$$

is a measurable solution function for $\mathcal{M}$ w.r.t. $\mathcal{L}$.

For linear SCMs, the unique solvability condition w.r.t. a subset translates into a matrix invertibility condition, as was already shown in [11].

PROPOSITION C.3 (Sufficient and necessary condition for unique solvability w.r.t. a subset for linear SCMs). *Let $\mathcal{M}$ be a linear SCM, $\mathcal{L} \subseteq \mathcal{I}$ and $\mathcal{O} = \mathcal{I} \setminus \mathcal{L}$. Then $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$ if and only if the matrix $A_{\mathcal{L}\mathcal{L}} = \mathbb{I}_{\mathcal{L}} - B_{\mathcal{L}\mathcal{L}}$ is invertible. Moreover, if $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$, then the mapping $\boldsymbol{g}_{\mathcal{L}} : \mathbb{R}^{\mathcal{O}} \times \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{L}}$ given by*

$$\boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) = A_{\mathcal{L}\mathcal{L}}^{-1}(B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}} + \Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e})\,,$$

is a measurable solution function for $\mathcal{M}$ w.r.t. $\mathcal{L}$.

Note that if $A_{\mathcal{L}\mathcal{L}}$ is invertible, then $A_{\mathcal{L}\mathcal{L}}^{+} = A_{\mathcal{L}\mathcal{L}}^{-1}$ (see Lemma 1.3 in [22]), and the matrix condition of Proposition C.2 is always satisfied and all the measurable solution functions $\boldsymbol{g}_{\mathcal{L}}^{\boldsymbol{v}}$ of Proposition C.2 are (up to a $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-null set) equal to the solution function $\boldsymbol{g}_{\mathcal{L}}$ of Proposition C.3.

REMARK. *A sufficient condition for $A_{\mathcal{L}\mathcal{L}}$ to be invertible is that the spectral radius of $B_{\mathcal{L}\mathcal{L}}$ is less than one. If that is the case, then $A_{\mathcal{L}\mathcal{L}}^{-1} = \sum_{n=0}^{\infty}(B_{\mathcal{L}\mathcal{L}})^n$. Note that the nonzero nondiagonal entries of the matrix $B_{\mathcal{L}\mathcal{L}}$ represent the directed edges in the induced subgraph $\mathcal{G}(\mathcal{M})_{\mathcal{L}}$. In particular, if the diagonal entries of the matrix $B_{\mathcal{L}\mathcal{L}}$ are zero, then for $n \in \mathbb{N}$, the coefficients of the matrix $(B_{\mathcal{L}\mathcal{L}})^n$ in the sum represent the sum of the product of the edge weights B_{ij} over directed paths of length n in the induced subgraph $\mathcal{G}(\mathcal{M})_{\mathcal{L}}$.*

From Proposition 3.10 we know that an SCM is solvable w.r.t. $\mathcal{L}$ if and only if it is ancestrally solvable w.r.t. $\mathcal{L}$. In particular, this result also holds for linear SCMs. We saw in Example 3.11 that a similar result for unique solvability does not hold, that is, in general, it does not hold that unique solvability w.r.t. $\mathcal{L}$ implies ancestral unique solvability w.r.t. $\mathcal{L}$. For the class of linear SCMs we do have the following positive result.

PROPOSITION C.4 (Equivalent unique solvability conditions for linear SCMs). *For a linear SCM $\mathcal{M}$ and a subset $\mathcal{L} \subseteq \mathcal{I}$ the following are equivalent:*

1. *$\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$;*
2. *$\mathcal{M}$ is ancestrally uniquely solvable w.r.t. $\mathcal{L}$;*
3. *$\mathcal{M}$ is uniquely solvable w.r.t. each strongly connected component in $\mathcal{G}(\mathcal{M})_{\mathcal{L}}$.*

Under the condition of unique solvability w.r.t. a subset $\mathcal{L}$ we can define the marginalization w.r.t. $\mathcal{L}$ of a linear SCM by mere substitution.

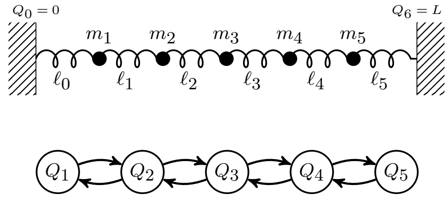

Fig 4: Damped coupled harmonic oscillator (top) and the graph of the SCM $\mathcal{M}$ that describes the positions of the masses at equilibrium (bottom) of Example D.1 for $d = 5$.

PROPOSITION C.5 (Marginalization of a linear SCM). *Let $\mathcal{M}$ be a linear SCM and $\mathcal{L} \subseteq \mathcal{I}$ a subset of endogenous variables such that $\mathbb{I}_{\mathcal{L}} - B_{\mathcal{L}\mathcal{L}}$ is invertible. Then there exists a marginalization $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ that is linear and with marginal causal mechanism $\tilde{\boldsymbol{f}} : \mathbb{R}^{\mathcal{O}} \times \mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{O}}$ given by*

$$\tilde{\boldsymbol{f}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) = [B_{\mathcal{O}\mathcal{O}} + B_{\mathcal{O}\mathcal{L}} A_{\mathcal{L}\mathcal{L}}^{-1} B_{\mathcal{L}\mathcal{O}}]\boldsymbol{x}_{\mathcal{O}} + [B_{\mathcal{O}\mathcal{L}} A_{\mathcal{L}\mathcal{L}}^{-1} \Gamma_{\mathcal{L}\mathcal{J}} + \Gamma_{\mathcal{O}\mathcal{J}}]\boldsymbol{e}\,,$$

where $A_{\mathcal{L}\mathcal{L}} = \mathbb{I}_{\mathcal{L}} - B_{\mathcal{L}\mathcal{L}}$. Moreover, this marginalization respects the latent projection, that is, $\big(\mathcal{G}^a \circ \mathrm{marg}(\mathcal{L})\big)(\mathcal{M}) \subseteq \big(\mathrm{marg}(\mathcal{L}) \circ \mathcal{G}^a\big)(\mathcal{M})$.

From Theorem 5.6 we know that $\mathcal{M}$ and its marginalization $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ over $\mathcal{L}$ are observationally, interventionally and counterfactually equivalent w.r.t. $\mathcal{O}$. A similar result can also be found in [11]. In contrast to nonlinear SCMs, this class of linear SCMs has the convenient property that every marginalization of a model of this class respects the latent projection. Moreover, the subclass of simple linear SCMs is even closed under marginalization.

APPENDIX D: EXAMPLES

In this appendix, we provide additional examples. In Appendix D.1 we provide some examples of SCMs that describe the equilibrium states of certain feedback systems governed by (random) differential equations [4] that motivated our study of cyclic SCMs. In Appendix D.2 we provide additional examples that support the main text.

D.1. SCMs as equilibrium models. In many systems occurring in the real world feedback loops between observed variables are present. For example, in economics, the price of a product may be a function of the demanded or supplied quantities, and vice versa; or in physics, two masses that are connected by a spring may exert forces on each other. Such systems are often described by a system of (random) differential equations. In [4] it was shown that SCMs are capable of modeling the causal semantics of the equilibrium states of such systems. For illustration purposes we provide the following toy example of interacting masses that are attached to springs.

EXAMPLE D.1 (Damped coupled harmonic oscillator). *Consider a one-dimensional system of d point masses $m_i \in \mathbb{R}$ ($i = 1, \ldots, d$) with positions Q_i, which are coupled by springs, with spring constants $k_i > 0$ and equilibrium lengths $\ell_i > 0$ ($i = 0, \ldots, d$), under influence of friction with friction coefficients $b_i \in \mathbb{R}$ ($i = 1, \ldots, d$) and with fixed endpoints $Q_0 = 0$ and $Q_{d+1} = L > 0$ (see Figure 4 (top)). The equations of motion of this system are provided by the following differential equations*

$$\frac{d^2 Q_i}{dt^2} = \frac{k_i}{m_i}(Q_{i+1} - Q_i - \ell_i) + \frac{k_{i-1}}{m_i}(Q_{i-1} - Q_i + \ell_{i-1}) - \frac{b_i}{m_i}\frac{dQ_i}{dt} \qquad (i = 1, \ldots, d)\,.$$

The dynamics of the masses, in terms of the position, velocity and acceleration, is described by a single and separate equation of motion for each mass. Under friction, that is, $b_i > 0$

($i = 1, \dots, d$), there is a unique equilibrium position, where the sum of forces vanishes for each mass. If one starts out of equilibrium, for example, by moving one or several masses out of equilibrium, then the masses will start to oscillate and converge to their unique equilibrium position. At equilibrium (i.e., for $t \to \infty$) the velocity $\frac{dQ_i}{dt}$ and acceleration $\frac{d^2Q_i}{dt^2}$ of the masses vanish (i.e., $\frac{dQ_i}{dt}, \frac{d^2Q_i}{dt^2} \to 0$), and thus the following equation holds at equilibrium

$$0 = \frac{k_i}{m_i}(Q_{i+1} - Q_i - \ell_i) + \frac{k_{i-1}}{m_i}(Q_{i-1} - Q_i + \ell_{i-1}),$$

for each mass ($i = 1, \dots, d$). Hence, for each mass $i = 1, \dots, d$ its equilibrium position Q_i is given by

$$Q_i = \frac{k_i(Q_{i+1} - \ell_i) + k_{i-1}(Q_{i-1} + \ell_{i-1})}{k_i + k_{i-1}}.$$

By considering the ℓ_i and k_i and L as fixed parameters, we arrive at a linear SCM (see [4] for more details about constructing an SCM from a dynamical system)

$$\mathcal{M} = \langle \{1, \dots, d\}, \emptyset, \mathbb{R}^d, \mathbf{1}, \boldsymbol{f}, \mathbb{P}_{\mathbf{1}} \rangle,$$

where the causal mechanism $\boldsymbol{f}$ is given by

$$f_i(\boldsymbol{q}) = \frac{k_i(q_{i+1} - \ell_i) + k_{i-1}(q_{i-1} + \ell_{i-1})}{k_i + k_{i-1}}.$$

Alternatively, (some of) the parameters could be treated as exogenous variables instead. Its graph is depicted in Figure 4 (bottom). This SCM allows us to describe the equilibrium behavior of the system under perfect intervention. For example, when forcing the mass j to a fixed position $Q_j = \xi_j$ with $0 \leq \xi_j \leq L$, the equilibrium positions of the masses correspond to the solutions of the intervened model $\mathcal{M}_{\mathrm{do}(\{j\},\xi_j)}$. It is an easy exercise to show that $\mathcal{M}$ is a simple SCM by using Proposition C.3.

Next, we show that the well known market equilibrium model from economics, which has been thoroughly discussed in the literature [see, e.g., 25], can be described by a (non-simple) SCM. This example illustrates how self-cycles enrich the class of SCMs.

EXAMPLE D.2 (Price, supply and demand). *Let X_D denote the demand and X_S the supply of a quantity of a product. The price of the product is denoted by X_P. The following system of differential equations describes how the demanded and supplied quantities are determined by the price, and how price adjustments occur in the market:*

$$\begin{aligned} X_D &= \beta_D X_P + E_D \\ X_S &= \beta_S X_P + E_S \\ \frac{dX_P}{dt} &= X_D - X_S, \end{aligned}$$

where E_D and E_S are exogenous random influences on the demand and supply, respectively, $\beta_D < 0$ is the reciprocal of the slope of the demand curve, and $\beta_S > 0$ is the reciprocal of the slope of the supply curve. At the situation known as a "market equilibrium", the price is determined implicitly by the condition that demanded and supplied quantities should be equal, since $\frac{dX_P}{dt} = 0$ at equilibrium. Applying the results in [4] gives rise to a linear SCM $\mathcal{M} = \langle \{P, S, D\}, \{S, D\}, \mathbb{R}^3, \mathbb{R}^2, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ at equilibrium with the causal mechanism defined by

$$\begin{aligned} f_D(\boldsymbol{x}, \boldsymbol{e}) &:= \beta_D x_P + e_D \\ f_S(\boldsymbol{x}, \boldsymbol{e}) &:= \beta_S x_P + e_S \\ f_P(\boldsymbol{x}, \boldsymbol{e}) &:= x_P + (x_D - x_S). \end{aligned}$$

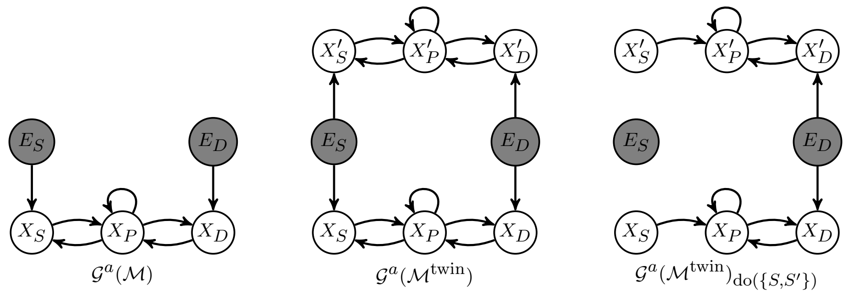

Fig 5: The augmented graph of the SCM $\mathcal{M}$ (left), its twin SCM $\mathcal{M}^{\text{twin}}$ (center) and the intervened twin SCM $(\mathcal{M}^{\text{twin}})_{\text{do}(\{S,S'\},(s,s'))}$ (right) of Examples D.2 and D.3.

Note how we use a self-cycle for P in order to implement the equilibrium equation $X_D = X_S$ as the causal mechanism for the price P.[7] *Moreover, $\mathcal{M}$ is uniquely solvable. Its augmented graph is depicted in Figure 5 (left).*

Next, we provide an example of how counterfactuals can be sensibly formulated for cyclic SCMs, namely for the price, supply and demand model at equilibrium.

EXAMPLE D.3 (Price, supply and demand at equilibrium). *Consider the price, supply and demand model at equilibrium of Example D.2 given by the SCM $\mathcal{M}$. As an example of a counterfactual query, consider*

$$\mathbb{P}(X'_P \,|\, \text{do}(X_S = s, X_{S'} = s'), X_P = p)\,,$$

which denotes the conditional distribution of X'_P given $X_P = p$ of a solution of the intervened twin model $\mathcal{M}^{\text{twin}}_{\text{do}(\{S,S'\},(s,s'))}$. In words: how would—ceteris paribus—*price have been distributed, had we intervened to set supplied quantities equal to s', given that actually we intervened to set supplied quantities equal to s and observed that this led to price p? A straightforward calculation shows that this counterfactual distribution of price is the Dirac measure on $x'_P = p + (s' - s)/\beta_D$. The augmented graphs of the SCM, its twin graph, and its intervened twin graph are depicted in Figure 5.*

D.2. Additional examples. In this subsection, we provide additional examples that support the main text.

Section 2.

EXAMPLE D.4 (Structural equations up to almost sure equality). *Consider the SCM $\mathcal{M} = \langle \mathbf{1}, \mathbf{1}, \mathcal{X}, \mathcal{E}, f, \mathbb{P}_{\mathcal{E}} \rangle$ with $\mathcal{X} = \mathcal{E} = \{-1, 0, 1\}$, $\mathbb{P}_{\mathcal{E}}(\{-1\}) = \mathbb{P}_{\mathcal{E}}(\{1\}) = \frac{1}{2}$ and $f(x,e) = e^2 + e - 1$. Let $\tilde{\mathcal{M}}$ be the SCM $\mathcal{M}$ but with a different causal mechanism $\tilde{f}(x,e) = e$. Then the sets of solutions of the structural equations agree for both SCMs for $e \in \{-1, +1\}$, while they differ only for $e = 0$, which occurs with probability zero. Hence, a pair of random variables (X, E) is a solution of $\mathcal{M}$ if and only if it is a solution of $\tilde{\mathcal{M}}$.*

[7] Richardson and Robins [25] argue that this market equilibrium model cannot be modeled as an SCM. We observe that it can, as long as one allows for self-cycles.

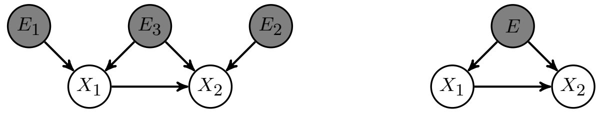

Fig 6: Augmented graphs of the SCMs $\mathcal{M}$ (left) and $\mathcal{M}^*$ (right) in Example D.6. For SCM $\mathcal{M}^*$, the exogenous variable E consists of two real-valued components; the structural equation for X_1 depends only on the first, while the structural equation for X_2 depends only on the second component.

EXAMPLE D.5 (The for-all and for-almost-every quantifier do not commute in general). *Consider the SCM $\mathcal{M} = \langle \mathbf{2}, \mathbf{1}, \boldsymbol{\mathcal{X}}, \mathcal{E}, \boldsymbol{f}, \mathbb{P}_{\mathcal{E}} \rangle$ with $\boldsymbol{\mathcal{X}} = (0,1)^2$, $\mathcal{E} = (0,1)$, the causal mechanism $\boldsymbol{f}$ given by*

$$f_1(\boldsymbol{x}, e) = x_1 \,, \quad f_2(\boldsymbol{x}, e) = \mathbf{1}_{\{0\}}(x_1 - e) \cdot (x_2 + 1) \,,$$

and $\mathbb{P}_{\mathcal{E}} = \mathbb{P}^E$ with $E \sim \mathcal{U}(0,1)$. Define the property

$$P(\boldsymbol{x}, e) := \begin{cases} 1 & \text{if } \boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, e) \text{ holds,} \\ 0 & \text{otherwise.} \end{cases}$$

Then, for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ and for $\mathbb{P}_{\mathcal{E}}$-almost every $e \in \mathcal{E}$ the property $P(\boldsymbol{x}, e)$ holds, however for $\mathbb{P}_{\mathcal{E}}$-almost every $e \in \mathcal{E}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ the property $P(\boldsymbol{x}, e)$ does not hold, since for $\mathbb{P}_{\mathcal{E}}$-almost every $e \in \mathcal{E}$ the equation $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, e)$ does not hold for $x_1 = e$. Hence, in general, for a property $P(\boldsymbol{x}, e)$ we have that for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ and for $\mathbb{P}_{\mathcal{E}}$-almost every $e \in \mathcal{E}$ $P(\boldsymbol{x}, e)$ does not imply for $\mathbb{P}_{\mathcal{E}}$-almost every $e \in \mathcal{E}$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ $P(\boldsymbol{x}, e)$ (see Lemma F.11 for additional properties of the for-almost-every quantifier).

EXAMPLE D.6 (Representation of latent confounders). *Consider the SCM $\mathcal{M} = \langle \mathbf{2}, \mathbf{3}, \mathbb{R}^2, \mathbb{R}^3, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}^3} \rangle$ with causal mechanism given by*

$$f_1(e_1, e_3) = e_1 + e_3$$
$$f_2(x_1, e_2, e_3) = x_1 e_3 + e_2$$

and $\mathbb{P}_{\mathbb{R}^3}$ the standard-normal distribution on $\mathbb{R}^3$; Figure 6 (left) shows the corresponding augmented graph. Then there exists no SCM $\mathcal{M}^* = \langle \mathbf{2}, \mathbf{1}, \mathbb{R}^2, \mathbb{R}^2, \boldsymbol{f}^*, \mathbb{P}^*_{\mathbb{R}^2} \rangle$ that satisfies the following conditions:

1. *$\mathcal{M}^*$ is interventionally equivalent to $\mathcal{M}$,*
2. *its structural equations have the form*

$$x_1 = f_1^*(e_1^*)$$
$$x_2 = f_2^*(x_1, e_2^*),$$

   where e_1^*, e_2^* are the two components of $e^* = (e_1^*, e_2^*) \in \mathbb{R}^2$,
3. *the function $e_2^* \mapsto f_2^*(x_1, e_2^*)$ is strictly monotonically increasing for all $x_1 \in \mathbb{R}$,*
4. *the cumulative distribution function F_2^* of the second component of $\mathbb{P}^*_{\mathbb{R}^2}$ is continuous and strictly monotonically increasing.*

The augmented graph of such an SCM is shown in Figure 6 (right).

The proof of this statement proceeds by contradiction. Assume that such an SCM $\mathcal{M}^*$ exists. For any uniquely solvable SCM $\bar{\mathcal{M}}$ and any endogenous variable i appearing in $\bar{\mathcal{M}}$, we denote with $F_{X_i}^{\bar{\mathcal{M}}}$ the marginal cumulative distribution function of the i^{th} component of the observational distribution of $\bar{\mathcal{M}}$. For all $\xi \in \mathbb{R}$, we have for all $x_2 \in \mathbb{R}$

$$F_{X_2}^{\mathcal{M}_{\mathrm{do}(\{1\},\xi)}}(x_2) = \mathbb{P}(\xi E_3 + E_2 \le x_2) = \Phi\left(x_2 / \sqrt{1 + \xi^2}\right), \tag{1}$$

where Φ denotes the (invertible) cdf of the standard-normal distribution. Now define $\phi : \mathbb{R} \to \mathbb{R}$ with $\phi(e_2) := \Phi^{-1}(F_2^*(e_2))$ and define the SCM $\tilde{\mathcal{M}} := \langle \mathbf{2}, \mathbf{1}, \mathbb{R}^2, \mathbb{R}^2, \tilde{\boldsymbol{f}}, \tilde{\mathbb{P}}_{\mathbb{R}^2} \rangle$ such that the causal mechanism $\tilde{\boldsymbol{f}}$ is given by

$$\tilde{f}_1(e_1) = f_1^*(e_1),$$

$$\tilde{f}_2(x_1, e_2) = f_2^*(x_1, \phi^{-1}(e_2)),$$

and $\tilde{\mathbb{P}}_{\mathbb{R}^2}$ is the push-forward measure of $\mathbb{P}^*_{\mathbb{R}^2}$ using $(id_{\mathbb{R}}, \phi)$. Then, $\tilde{\mathcal{M}}$ is interventionally equivalent to $\mathcal{M}^*$ by construction, and the second component of $\tilde{\mathbb{P}}_{\mathbb{R}^2}$ has a standard-normal distribution. Let $(\tilde{X}_1, \tilde{X}_2, \tilde{E})$ be a solution of $\tilde{\mathcal{M}}$ and let us write $\tilde{E} = (\tilde{E}_1, \tilde{E}_2)$. Then, for all $\xi \in \mathbb{R}$ and $\tilde{e}_2 \in \mathbb{R}$,

$$F_{X_2}^{\tilde{\mathcal{M}}_{\mathrm{do}(\{1\},\xi)}}(\tilde{f}_2(\xi, \tilde{e}_2)) = \mathbb{P}(\tilde{f}_2(\xi, \tilde{E}_2) \le \tilde{f}_2(\xi, \tilde{e}_2)) = \mathbb{P}(\tilde{E}_2 \le \tilde{e}_2) = \Phi(\tilde{e}_2),$$

using that $\tilde{e}_2 \mapsto \tilde{f}_2(\xi, \tilde{e}_2)$, too, is strictly monotonically increasing for all ξ. This implies that, for all $\xi \in \mathbb{R}$ and $\tilde{e}_2 \in \mathbb{R}$,

$$\tilde{f}_2(\xi, \tilde{e}_2) = (F_{X_2}^{\mathcal{M}_{\mathrm{do}(\{1\},\xi)}})^{-1}\big(\Phi(\tilde{e}_2)\big) = \sqrt{1 + \xi^2}\, \tilde{e}_2\,,$$

where we used interventional equivalence of $\mathcal{M}$ and $\tilde{\mathcal{M}}$, and (1) for the second equality. Furthermore, $\tilde{X}_2 = \tilde{f}_2(\tilde{X}_1, \tilde{E}_2) = \sqrt{1 + \tilde{X}_1^2}\; \tilde{E}_2$ a.s., so $\tilde{E}_2 = \tilde{X}_2 / \sqrt{1 + \tilde{X}_1^2}$ a.s.. Now let $(X_1, X_2, E_1, E_2, E_3)$ be a solution of $\mathcal{M}$. By observational equivalence, $(\tilde{X}_1, \tilde{X}_2)$ has the same distribution as (X_1, X_2), and thus $\tilde{E}_2$ is distributed as

$$\frac{X_2}{\sqrt{1 + X_1^2}} = \frac{(E_1 + E_3)E_3 + E_2}{\sqrt{1 + (E_1 + E_3)^2}} \text{ a.s.}.$$

This contradicts the fact that $\tilde{E}_2$ has a standard-normal distribution as, for example, the mean of the right-hand side is nonzero.

EXAMPLE D.7 (Counterfactual density unidentifiable from observational and interventional densities [6]). *Let $\rho \in \mathbb{R}$ and*

$$\mathcal{M}_\rho = \langle \mathbf{2}, \mathbf{2}, \{0,1\} \times \mathbb{R}, \{0,1\} \times \mathbb{R}^2, \boldsymbol{f}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$$

be the SCM with causal mechanism given by

$$f_1(\boldsymbol{x}, \boldsymbol{e}) = e_1\,, \quad f_2(\boldsymbol{x}, \boldsymbol{e}) = e_{21}(1 - x_1) + e_{22}x_1$$

and $\mathbb{P}_{\boldsymbol{\mathcal{E}}} = \mathbb{P}^{(E_1, \boldsymbol{E}_2)}$ with $E_1 \sim \mathrm{Bernoulli}(1/2)$,

$$\boldsymbol{E}_2 := \begin{pmatrix} E_{21} \\ E_{22} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

normally distributed and $E_1 \perp\!\!\!\perp \boldsymbol{E}_2$. In an epidemiological setting, this SCM could be used to model whether a patient was treated or not (X_1) and the corresponding outcome for that patient (X_2).

Suppose in the actual world we did not assign treatment to a patient ($X_1 = 0$) and the outcome was $X_2 = c \in \mathbb{R}$. Consider the counterfactual query "What would the outcome have been, if we had assigned treatment to this patient?". We can answer this question by introducing a parallel counterfactual world that is modeled by the twin SCM $\mathcal{M}_\rho^{\mathrm{twin}}$, as depicted in Figure 7. The counterfactual query then asks for $p(X_{2'} = x_{2'} \mid \mathrm{do}(X_{1'} = 1, X_1 = 0), X_2 = c)$. One can calculate that

$$\begin{pmatrix} X_{2'} \\ X_2 \end{pmatrix} \mid \mathrm{do}(X_{1'} = 1, X_1 = 0) \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

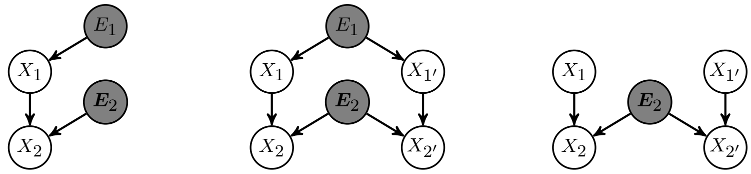

Fig 7: The augmented graph of the SCM $\mathcal{M}_\rho$ (left), its twin SCM $\mathcal{M}_\rho^{\text{twin}}$ (center) and the intervened twin SCM $(\mathcal{M}_\rho^{\text{twin}})_{\text{do}(\{1',1\},(1,0))}$ (right) of Example D.7.

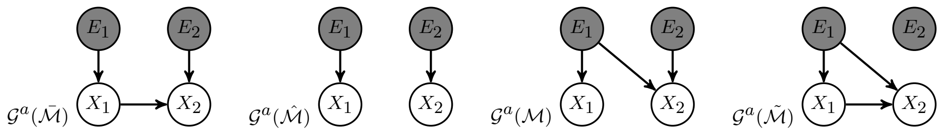

Fig 8: The augmented graphs of SCMs $\bar{\mathcal{M}}$, $\hat{\mathcal{M}}$, $\mathcal{M}$, and $\tilde{\mathcal{M}}$ that appear in Examples 4.4, D.10, and D.13.

and hence $X_{2'} \mid \text{do}(X_{1'} = 1, X_1 = 0), X_2 = c \sim \mathcal{N}(\rho c, 1 - \rho^2)$. Note that the answer to the counterfactual query depends on a quantity ρ that we cannot identify from the observational density $p(X_1, X_2)$ or the interventional densities $p(X_2 \mid \text{do}(X_1 = 0))$ and $p(X_2 \mid \text{do}(X_1 = 1))$, none of which depends on ρ. Therefore, even data from randomized controlled trials combined with observational data would not suffice to determine the value of this particular counterfactual query. Indeed, SCMs $\mathcal{M}_\rho$ and $\mathcal{M}_{\rho'}$ with $\rho \neq \rho'$ are interventionally equivalent, but not counterfactually equivalent.

Section 3.

EXAMPLE D.8 (Mixtures of solutions are solutions). *Let $\mathcal{M} = \langle \mathbf{1}, \emptyset, \mathbb{R}, \mathbf{1}, f, \mathbb{P}_{\mathbf{1}} \rangle$ be an SCM with causal mechanism $f : \mathcal{X} \times \mathcal{E} \to \mathcal{X}$ defined by $f(x, e) = x - x^2 + 1$. There exist only two measurable solution functions $g_\pm : \mathcal{E} \to \mathcal{X}$ for $\mathcal{M}$, defined by $g_\pm(e) = \pm 1$. Let $X : \Omega \to \mathbb{R}$ be a random variable that is a nontrivial mixture of point masses on $\{-1, +1\}$. Then X is a solution of $\mathcal{M}$, however neither $g_+(E) = X$ a.s., nor $g_-(E) = X$ a.s., for any random variable E such that $\mathbb{P}^E = \mathbb{P}_{\mathcal{E}}$.*

EXAMPLE D.9 (Solvability is not preserved under perfect intervention). *Consider the SCM $\mathcal{M} = \langle \mathbf{2}, \emptyset, \mathbb{R}^2, \mathbf{1}, \boldsymbol{f}, \mathbb{P}_{\mathbf{1}} \rangle$ with the following causal mechanism*

$$f_1(\boldsymbol{x}) = x_1 + x_1^2 - x_2 + 1\,, \quad f_2(\boldsymbol{x}) = x_2(1 - \mathbf{1}_{\{0\}}(x_1)) + 1\,.$$

This SCM has a unique solution $(0, 1)$. Doing a perfect intervention $\text{do}(\{1\}, \xi_1)$ for some $\xi_1 \neq 0$, however, leads to an intervened model $\mathcal{M}_{\text{do}(\{1\},\xi_1)}$ that is not solvable. Performing instead the perfect intervention $\text{do}(\{2\}, \xi_2)$ for some $\xi_2 > 1$ leads also to a nonuniquely solvable SCM $\mathcal{M}_{\text{do}(\{2\},\xi_2)}$ which has solutions with multiple induced distributions, for example, $(X_1, X_2) = (\phi(\xi_2)\sqrt{\xi_2 - 1}, \xi_2)$ with some measurable $\phi : \mathbb{R} \to \{-1, +1\}$, but also mixtures of those.

Section 4.

EXAMPLE D.10 (Counterfactually equivalent SCMs with different graphs). *Consider the SCM $\hat{\mathcal{M}} = \langle \mathbf{2}, \mathbf{2}, \{-1, 1\}^2, \{-1, 1\}^2, \hat{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ with causal mechanism given by $\hat{f}_1(\boldsymbol{x}, \boldsymbol{e}) =$*

e_1 and $\hat{f}_2(\boldsymbol{x},\boldsymbol{e}) = e_2$, and $\mathbb{P}_{\boldsymbol{\mathcal{E}}} = \mathbb{P}^{\boldsymbol{E}}$ with $E_1, E_2 \sim \mathcal{U}(\{-1,1\})$ uniformly distributed and $E_1 \perp\!\!\!\perp E_2$. Consider also the SCM $\mathcal{M}$ that is the same as $\hat{\mathcal{M}}$ except for its causal mechanism, which is given by $f_1(\boldsymbol{x},\boldsymbol{e}) = e_1$ and $f_2(\boldsymbol{x},\boldsymbol{e}) = e_1 e_2$. Then $\mathcal{M}$ and $\hat{\mathcal{M}}$ are counterfactually equivalent although $\mathcal{G}(\mathcal{M})$ is not equal to $\mathcal{G}(\hat{\mathcal{M}})$ (see Figure 8).

Section 5.

EXAMPLE D.11 (Marginalization condition of an SCM is not a necessary condition). *Consider the SCM $\mathcal{M} = \langle \mathbf{4}, \mathbf{1}, \mathbb{R}^4, \mathbb{R}, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}} \rangle$ with causal mechanism given by*

$$f_1(\boldsymbol{x}, e) = e\,, \quad f_2(\boldsymbol{x}, e) = x_1\,, \quad f_3(\boldsymbol{x}, e) = x_2\,, \quad f_4(\boldsymbol{x}, e) = x_4$$

and $\mathbb{P}_{\mathbb{R}}$ is the standard-normal measure on $\mathbb{R}$. This SCM is solvable w.r.t. $\mathcal{L} = \{2,4\}$, but not uniquely solvable w.r.t. $\mathcal{L}$, and hence we cannot apply Definition 5.3 to $\mathcal{L}$. However, the SCM $\tilde{\mathcal{M}}$ on the endogenous variables $\{1,3\}$ with the causal mechanism $\tilde{\boldsymbol{f}}$ given by $\tilde{f}_1(\boldsymbol{x}, e) = e$ and $\tilde{f}_3(\boldsymbol{x}, e) = x_1$ is counterfactually equivalent to $\mathcal{M}$ w.r.t. $\{1,3\}$, which can be checked easily.

EXAMPLE D.12 (Graph of the marginal SCM is a strict subgraph of the latent projection). *Consider the SCM $\mathcal{M} = \langle \mathbf{3}, \mathbf{1}, \mathbb{R}^3, \mathbb{R}, \boldsymbol{f}, \mathbb{P}_{\mathbb{R}} \rangle$ with causal mechanism given by*

$$f_1(\boldsymbol{x}, \boldsymbol{e}) = e_1\,, \quad f_2(\boldsymbol{x}, \boldsymbol{e}) = x_1 - x_3\,, \quad f_3(\boldsymbol{x}, \boldsymbol{e}) = x_1$$

and take for $\mathbb{P}_{\mathbb{R}}$ the standard-normal measure on $\mathbb{R}$. In contrast, to the (augmented) graph of $\mathcal{M}$, there is no directed path in the (augmented) graph of the marginal SCM $\mathcal{M}_{\mathrm{marg}(\{3\})}$.

Section 7.

EXAMPLE D.13 (Detecting a bidirected edge in the graph of an SCM). *Consider the SCM $\bar{\mathcal{M}} = \langle \mathbf{2}, \mathbf{2}, \{-1,1\}^2, \{-1,1\}^2, \bar{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ with causal mechanism given by $\bar{f}_1(\boldsymbol{x},\boldsymbol{e}) = e_1$ and $\bar{f}_2(\boldsymbol{x},\boldsymbol{e}) = x_1 e_2$, and $\mathbb{P}_{\boldsymbol{\mathcal{E}}} = \mathbb{P}^{\boldsymbol{E}}$ with $E_1, E_2 \sim \mathcal{U}(\{-1,1\})$ uniformly distributed and $E_1 \perp\!\!\!\perp E_2$. Consider also the SCM $\tilde{\mathcal{M}}$ that is the same as $\bar{\mathcal{M}}$ except for its causal mechanism, which is given by $\tilde{f}_1(\boldsymbol{x},\boldsymbol{e}) = e_1$ and $\tilde{f}_2(\boldsymbol{x},\boldsymbol{e}) = x_1 e_1$. See Figure 8 for their augmented graphs. For the SCM $\tilde{\mathcal{M}}$ we observe that the marginal interventional distribution $\mathbb{P}_{\tilde{\mathcal{M}}_{\mathrm{do}(\{1\},\xi_1)}}(X_2 = -1)$ is not equal to the conditional distribution $\mathbb{P}_{\tilde{\mathcal{M}}}(X_2 = -1 \,|\, X_1 = \xi_1)$ for both $\xi_1 = -1$ and $\xi_1 = 1$. This observation suffices to identify the presence of the bidirected edge $1 \leftrightarrow 2$ in the graph $\mathcal{G}(\tilde{\mathcal{M}})$. For the SCM $\bar{\mathcal{M}}$, whose graph does not contain the bidirected edge $1 \leftrightarrow 2$, the marginal interventional distribution and conditional distribution coincide.*

APPENDIX E: PROOFS

This appendix contains the proofs of all the theoretical results in the appendices A, B and C, and the main text. Some of the proofs will rely on the measure theoretic terminology and results of Appendix F.

E.1. Proofs of the appendices.

PROOF OF LEMMA A.5. It suffices to show that for every C-d-open walk between i and j in $\mathcal{G}$, there exists a C-d-open path between i and j in $\mathcal{G}$. Take a C-d-open walk $\pi = (i = i_0, \dots, i_n = j)$. If a node ℓ occurs more than once in π, let i_j be the first occurrence of ℓ in π and i_k the last occurrence of ℓ in π. We now construct a new walk π' from π by removing the subwalk between i_j and i_k of π from π. It is easy to check that the new walk π' is still C-d-open. If ℓ is an endpoint on π', then i_j or i_k must be endpoint of π, and hence $\ell \notin C$. If ℓ is a non-endpoint non-collider on π', then also i_j or i_k must have been a non-endpoint non-collider on π, and hence $\ell \notin C$. If ℓ is a collider on π', then either (i) i_j or i_k are both colliders on π, and hence ℓ is ancestor of C in $\mathcal{G}$, or (ii) on the subwalk between i_j and i_k that was removed, there must be a directed path in $\mathcal{G}$ from i_j or i_k to a collider in $\mathrm{an}_{\mathcal{G}}(C)$, and hence, ℓ is in $\mathrm{an}_{\mathcal{G}}(C)$. The other nodes on π' cannot be responsible for C-d-blocking the walk, since they also occur (together with their adjacent edges) on π and they do not C-d-block π.

In π', the number of nodes that occur multiple times is at least one less than in π. Repeat this procedure until no repeated nodes are left. □

PROOF OF THEOREM A.7. The first case is a well known result. An elementary proof is obtained by noting that an acyclic system of structural equations trivially satisfies the local directed Markov property, and then apply [16, Proposition 4], followed by applying the stability of d-separation with respect to (graphical) marginalization [8, Lemma 2.2.15]. Alternatively, the result also follows from sequential application of Theorems 3.8.2, 3.8.11, 3.7.7, 3.7.2 and 3.3.3 (using Remark 3.3.4) in [8].

The discrete case is proved by the series of results Theorem 3.8.12, Remark 3.7.2, Theorem 3.6.6 and 3.5.2 in [8].

The linear case is proved in Example 3.8.17 in [8]. To connect the assumptions made there with the ones we state here, observe that under the linear transformation rule for Lebesgue measures, the image measure of $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$ under the linear mapping $\mathbb{R}^{\mathcal{J}} \to \mathbb{R}^{\mathcal{I}} : \boldsymbol{e} \mapsto \Gamma_{\mathcal{I}\mathcal{J}}\boldsymbol{e}$ gives a measure on $\boldsymbol{\mathcal{X}} = \mathbb{R}^{\mathcal{I}}$ with a density w.r.t. the Lebesgue measure on $\mathbb{R}^{\mathcal{I}}$, as long as the image of the linear mapping is the entire $\mathbb{R}^{\mathcal{I}}$. This is guaranteed if each causal mechanism has a nontrivial dependence on some exogenous variable(s), that is, for each $i \in \mathcal{I}$ there is some $j \in \mathcal{J}$ with $\Gamma_{ij} \neq 0$. □

PROOF OF PROPOSITION A.12. This follows directly from the fact that the strongly connected components of $\mathcal{G}^a(\mathcal{M})$ form a DAG by Lemma A.2 and that the directed edges in $\mathcal{G}^a(\mathrm{acy}(\mathcal{M}))$ by construction respect every topological ordering of that DAG. Both SCMs are observationally equivalent by construction. □

PROOF OF PROPOSITION A.14. This follows immediately from the Definitions A.11 and A.13. □

PROOF OF LEMMA A.17. It suffices to show that for every C-σ-open walk between i and j in $\mathcal{G}$, there exists a C-σ-open path between i and j in $\mathcal{G}$. Let $\pi = (i = i_0, \dots, i_n = j)$ be a C-σ-open walk in $\mathcal{G}$. If a node ℓ occurs more than once in π, let i_j be the first node in π and i_k the last node in π that are in the same strongly connected component as ℓ. Since i_j and i_k are in the same strongly connected component, there are directed paths $i_j \to \cdots \to i_k$ and $i_k \to \cdots \to i_j$ in $\mathcal{G}$. We now construct a new walk π' from π by replacing the subwalk between i_j and i_k of π by a particular directed path between i_j and i_k: (i) If $k = n$, or if $k < n$ and $i_k \to i_{k+1}$ on π, we replace it by a shortest directed path $i_j \to \cdots \to i_k$, otherwise

(ii) we replace it by a shortest directed path $i_j \leftarrow \cdots \leftarrow i_k$. We now show that the new walk π' is still C-σ-open.

π' cannot become C-σ-blocked through one of the initial nodes $i_0 \dots i_{j-1}$ or one of the final nodes $i_{k+1} \dots i_n$ on π', since these nodes occur in the same local configuration on π and do not C-σ-block π by assumption. Furthermore, π' cannot become C-σ-blocked through one of the nodes strictly between i_j and i_k on π' (if there are any), since these nodes are all non-endpoint non-colliders that only point to nodes in the same strongly connected component on π'. Because π is C-σ-open, $i_k \notin C$ if $k = n$ or if $i_k \to i_{k+1}$ on π. This holds in particular in case (i). Similarly, $i_j \notin C$ if $j = 0$ or $i_{j-1} \leftarrow i_j$ on π.

In case (i), π' is not C-σ-blocked by i_k because i_k is a non-collider on π' but $i_k \notin C$. Also i_j does not C-σ-block π'. Assume $i_j \neq i_k$ (otherwise there is nothing to prove). If $j = 0$, or if $j > 0$ and $i_{j-1} \leftarrow i_j$ on π', then the same holds for π and hence $i_j \notin C$; i_j is then a non-collider on π', but $i_j \notin C$. If $j > 0$ and $i_{j-1} \leftrightarrow i_j$ or $i_{j-1} \to i_j$ on π' then i_j is a non-endpoint non-collider on π' that does not point to a node in another strongly connected component.

Now consider case (ii). If $j = 0$ or $i_{j-1} \leftarrow i_j$ on π' then this case is analogous to case (i). So assume $j > 0$ and $i_{j-1} \to i_j$ or $i_{j-1} \leftrightarrow i_j$ on π'. If i_j is an endpoint of π', then $i_j = i_k$ and $k = n$ and therefore $i_k \notin C$, and hence i_j and i_k do not C-σ-block π'. Otherwise, i_j must be a collider on π' (whether $i_j = i_k$ or not). Then on the subwalk of π between i_j and i_k there must be a directed path from i_j to a collider that is ancestor of C, which implies that i_j is itself ancestor of C, and hence i_j does not C-σ-block π'. Also i_k cannot C-σ-block π'. Assume $i_j \neq i_k$ (otherwise there is nothing to prove). Since $i_k \leftarrow i_{k+1}$ or $i_k \leftrightarrow i_{k+1}$ on π', i_k is a non-endpoint non-collider on π' that does not point to a node in another strongly connected component.

Now in π', the number of nodes that occurs more than once is at least one less than in π. Repeat this procedure until no nodes occur more than once. □

PROOF OF PROPOSITION A.19. This follows directly as a special case of Corollary 2.8.4 in [8]. □

PROOF OF THEOREM A.21. An SCM $\mathcal{M}$ that is uniquely solvable w.r.t. each strongly connected component is uniquely solvable and hence, by Theorem 3.6, all its solutions have the same observational distribution. The last statement follows from the series of results Theorem 3.8.2, 3.8.11, Lemma 3.7.7 and Remark 3.7.2 in [8]. Alternatively, we give here a shorter proof: Under the stated conditions one can always construct the acyclification $\mathrm{acy}(\mathcal{M})$ which is observationally equivalent to $\mathcal{M}$ and is acyclic (see Proposition A.12) and hence we can apply Theorem A.7 to $\mathrm{acy}(\mathcal{M})$. Together with Proposition A.14 and A.19 this gives

$$A \underset{\mathcal{G}(\mathcal{M})}{\overset{\sigma}{\perp}} B \,|\, C \iff A \underset{\mathrm{acy}(\mathcal{G}(\mathcal{M}))}{\overset{d}{\perp}} B \,|\, C \implies A \underset{\mathcal{G}(\mathrm{acy}(\mathcal{M}))}{\overset{d}{\perp}} B \,|\, C \implies \boldsymbol{X}_A \underset{\mathbb{P}_{\mathcal{M}}^{\boldsymbol{X}}}{\perp\!\!\!\perp} \boldsymbol{X}_B \,|\, \boldsymbol{X}_C\,,$$

for $A, B, C \subseteq \mathcal{I}$ and $\boldsymbol{X}$ a solution of $\mathcal{M}$. □

PROOF OF COROLLARY A.22. First observe that simplicity is preserved under both perfect intervention and the twin operation (see Proposition 8.2). Now the first statement follows from Theorem A.21 if one takes into account the identities of Proposition 2.14 and 2.19. Similarly, the last statement follows from Theorem A.7. □

PROOF OF PROPOSITION A.32. Let $\tilde{\mathcal{M}} =: \langle \mathcal{V}, \hat{\mathcal{H}}, \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{E}}, \tilde{\boldsymbol{f}}, \mathbb{P}_{\boldsymbol{\mathcal{E}}} \rangle$ be the induced SCM. Observe that every loop $\mathcal{O} \in \mathcal{L}(\mathcal{G}(\tilde{\mathcal{M}}))$ is a loop in $\mathcal{L}(\mathcal{G})$. Fix $\check{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}}$ and $\check{\boldsymbol{e}} \in \boldsymbol{\mathcal{E}}$. For every $\mathcal{O} \in \mathcal{L}(\mathcal{G}(\tilde{\mathcal{M}}))$, define

$$I_{\mathcal{O}} := (\mathrm{pa}_{\mathcal{G}}(\mathcal{O}) \setminus \mathcal{O}) \setminus (\mathrm{pa}(\mathcal{O}) \setminus \mathcal{O}) \subseteq \tilde{\mathcal{I}}$$

and

$$J_{\mathcal{O}} := \{\mathcal{F} \in \tilde{\mathcal{J}} : \mathcal{F} \cap \mathcal{O} \neq \emptyset\} \setminus \mathrm{pa}(\mathcal{O}) \subseteq \tilde{\mathcal{J}}\,.$$

Now, define the family of measurable mappings $(\tilde{\boldsymbol{g}}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G}(\tilde{\mathcal{M}}))}$, where the mapping $\tilde{\boldsymbol{g}}_{\mathcal{O}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ is given by

$$\tilde{\boldsymbol{g}}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) := \boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \check{\boldsymbol{x}}_{I_{\mathcal{O}}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}, \check{\boldsymbol{e}}_{J_{\mathcal{O}}})$$

where $\boldsymbol{x}_{\mathrm{pa}_{\mathcal{G}}(\mathcal{O})\setminus\mathcal{O}} = (\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \check{\boldsymbol{x}}_{I_{\mathcal{O}}})$ and $\widehat{\boldsymbol{e}}_{\mathcal{O}} = (\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}, \check{\boldsymbol{e}}_{J_{\mathcal{O}}})$. Observe that from the definition of the parents (see Definition 2.6) it follows that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ we have

$$\boldsymbol{x}_{\mathcal{O}} = \tilde{\boldsymbol{f}}_{\mathcal{O}}(\boldsymbol{x}_{\setminus I_{\mathcal{O}}}, \check{\boldsymbol{x}}_{I_{\mathcal{O}}}, \boldsymbol{e}_{\setminus J_{\mathcal{O}}}, \check{\boldsymbol{e}}_{J_{\mathcal{O}}}) \quad \Longleftrightarrow \quad \boldsymbol{x}_{\mathcal{O}} = \tilde{\boldsymbol{f}}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e})\,.$$

This, together with the fact that the family of mappings $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G})}$ is a compatible system of solution functions, implies that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ we have

$$\boldsymbol{x}_{\mathcal{O}} = \tilde{\boldsymbol{g}}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \quad \Longrightarrow \quad \boldsymbol{x}_{\mathcal{O}} = \tilde{\boldsymbol{f}}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e})\,.$$

Hence, $\iota(\widehat{\mathcal{M}})$ is loop-wisely solvable and thus $(\tilde{\boldsymbol{g}}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G}(\tilde{\mathcal{M}}))}$ is a family of measurable solution functions. In particular, for all $\mathcal{O}, \tilde{\mathcal{O}} \in \mathcal{L}(\mathcal{G}(\tilde{\mathcal{M}}))$ with $\tilde{\mathcal{O}} \subseteq \mathcal{O}$ and for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ we have

$$\boldsymbol{x}_{\mathcal{O}} = \tilde{\boldsymbol{g}}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \quad \Longrightarrow \quad \boldsymbol{x}_{\tilde{\mathcal{O}}} = \tilde{\boldsymbol{g}}_{\tilde{\mathcal{O}}}(\boldsymbol{x}_{\mathrm{pa}(\tilde{\mathcal{O}})\setminus\tilde{\mathcal{O}}}, \boldsymbol{e}_{\mathrm{pa}(\tilde{\mathcal{O}})})\,.$$

From this we conclude that $(\tilde{\boldsymbol{g}}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G}(\tilde{\mathcal{M}}))}$ is a compatible system of solution functions. □

PROOF OF LEMMA A.33. Suppose $\mathcal{M}$ is loop-wisely uniquely solvable and consider a subset $\mathcal{O} \subseteq \mathcal{I}$. Consider the induced subgraph $\mathcal{G}^a(\mathcal{M})_{\mathcal{O}}$ of $\mathcal{G}^a(\mathcal{M})$ on the nodes $\mathcal{O}$. Then every strongly connected component of $\mathcal{G}^a(\mathcal{M})_{\mathcal{O}}$ is an element of $\mathcal{L}(\mathcal{G}(\mathcal{M}))$. Let $\mathcal{C}$ be such a strongly connected component in $\mathcal{G}^a(\mathcal{M})_{\mathcal{O}}$, and let $\boldsymbol{g}_{\mathcal{C}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{C})\setminus\mathcal{C}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{C})} \to \boldsymbol{\mathcal{X}}_{\mathcal{C}}$ be a measurable solution function for $\mathcal{M}$ w.r.t. $\mathcal{C}$. Since $\mathcal{G}^a(\mathcal{M})_{\mathcal{O}}$ partitions into strongly connected components, we can recursively (by following a topological ordering of the DAG $\mathcal{G}^a(\mathcal{M})^{\mathrm{sc}}_{\mathcal{O}}$ from Lemma A.2) insert these mappings into each other to obtain a mapping $\boldsymbol{g}_{\mathcal{O}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ that makes $\mathcal{M}$ uniquely solvable w.r.t. $\mathcal{O}$. □

PROOF OF PROPOSITION A.34. Let $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\in\mathcal{L}(\mathcal{G}(\mathcal{M}))}$ be any family of measurable solution functions, where $\boldsymbol{g}_{\mathcal{O}}$ is measurable solution function of $\mathcal{M}$ w.r.t. $\mathcal{O}$. Then, for $\mathcal{O}, \tilde{\mathcal{O}} \in \mathcal{L}(\mathcal{G}(\mathcal{M}))$ such that $\tilde{\mathcal{O}} \subseteq \mathcal{O}$, we have that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e}) \quad \Longrightarrow \quad \boldsymbol{x}_{\tilde{\mathcal{O}}} = \boldsymbol{f}_{\tilde{\mathcal{O}}}(\boldsymbol{x}, \boldsymbol{e})\,.$$

This implies that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \quad \Longrightarrow \quad \boldsymbol{x}_{\tilde{\mathcal{O}}} = \boldsymbol{g}_{\tilde{\mathcal{O}}}(\boldsymbol{x}_{\mathrm{pa}(\tilde{\mathcal{O}})\setminus\tilde{\mathcal{O}}}, \boldsymbol{e}_{\mathrm{pa}(\tilde{\mathcal{O}})})\,.$$

□

PROOF OF COROLLARY 8.5. This follows directly from Proposition 7.1 and 7.2. □

Appendix B.

PROOF OF PROPOSITION B.1. Let $\tilde{\boldsymbol{f}}:\boldsymbol{\mathcal{E}}\times\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{X}}$ be the causal mechanism of a structurally minimal SCM that is equivalent to $\mathcal{M}$ (see Proposition 2.11). In particular, for any $\boldsymbol{\epsilon}_{\setminus\mathrm{pa}(\mathcal{O})}\in\boldsymbol{\mathcal{E}}_{\setminus\mathrm{pa}(\mathcal{O})}$ and $\boldsymbol{\xi}_{\setminus\mathrm{pa}(\mathcal{O})}\in\boldsymbol{\mathcal{X}}_{\setminus\mathrm{pa}(\mathcal{O})}$, we have that for all $\boldsymbol{x}\in\boldsymbol{\mathcal{X}}$ and all $\boldsymbol{e}\in\boldsymbol{\mathcal{E}}$, $\tilde{\boldsymbol{f}}(\boldsymbol{x},\boldsymbol{e})=\tilde{\boldsymbol{f}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})},\boldsymbol{\xi}_{\setminus\mathrm{pa}(\mathcal{O})},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})},\boldsymbol{\epsilon}_{\setminus\mathrm{pa}(\mathcal{O})})$. This means that we may also consider $\tilde{\boldsymbol{f}}$ as a mapping $\tilde{\boldsymbol{f}}:\boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})}\times\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}\to\boldsymbol{\mathcal{X}}$.

Consider the set

$$\tilde{\boldsymbol{\mathcal{S}}}:=\{(\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})},\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}},\boldsymbol{x}_{\mathcal{O}})\in\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}\times\boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}\times\boldsymbol{\mathcal{X}}_{\mathcal{O}}:\boldsymbol{x}_{\mathcal{O}}=\tilde{\boldsymbol{f}}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})})\}\,.$$

By similar reasoning as in the proof of Theorem 3.2, $\tilde{\boldsymbol{\mathcal{S}}}$ is measurable.

By assumption, for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}\in\boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}_{\setminus\mathcal{O}}\in\boldsymbol{\mathcal{X}}_{\setminus\mathcal{O}}$ the space $\{\boldsymbol{x}_{\mathcal{O}}\in\boldsymbol{\mathcal{X}}_{\mathcal{O}}:\boldsymbol{x}_{\mathcal{O}}=\boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x},\boldsymbol{e})\}$ is nonempty and σ-compact. By applying Lemma F.10 to the canonical projection $\boldsymbol{pr}_{\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}}:\boldsymbol{\mathcal{E}}\to\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}$ and using the equivalence of $\boldsymbol{f}$ and $\tilde{\boldsymbol{f}}$, we obtain that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}}$-almost every $\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}\in\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}$ and for all $\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}\in\boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}$ the space

$$\tilde{\boldsymbol{\mathcal{S}}}_{(\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})},\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}})}:=\{\boldsymbol{x}_{\mathcal{O}}\in\boldsymbol{\mathcal{X}}_{\mathcal{O}}:\boldsymbol{x}_{\mathcal{O}}=\tilde{\boldsymbol{f}}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})})\}$$

is nonempty and σ-compact.

The second measurable selection theorem, Theorem F.9, now implies that there exists a measurable $\boldsymbol{g}_{\mathcal{O}}:\boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}\times\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}\to\boldsymbol{\mathcal{X}}_{\mathcal{O}}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}}$-almost every $\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}\in\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}$ and for all $\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}\in\boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}$

$$\boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})})=\tilde{\boldsymbol{f}}_{\mathcal{O}}\big(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}},\boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}),\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}\big).$$

Once more applying Lemma F.10, we obtain that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}\in\boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}\in\boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{O}}=\boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})})\implies\boldsymbol{x}_{\mathcal{O}}=\boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x},\boldsymbol{e}).$$

Hence $\mathcal{M}$ is solvable w.r.t. $\mathcal{O}$. □

PROOF OF PROPOSITION B.4. Without loss of generality, we assume that $\mathcal{M}$ is structurally minimal (see Proposition 2.11). Define $\mathcal{C}:=\mathcal{A}\cap\tilde{\mathcal{A}}$ and $\mathcal{D}:=\mathcal{A}\cup\tilde{\mathcal{A}}$. Let $\boldsymbol{g}_{\mathcal{A}}$, $\boldsymbol{g}_{\tilde{\mathcal{A}}}$ be measurable solution functions for $\mathcal{M}$ w.r.t. $\mathcal{A}$ and $\tilde{\mathcal{A}}$, respectively. Note that $\mathrm{pa}(\mathcal{C})\setminus\mathcal{C}\subseteq\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}$ and similarly $\mathrm{pa}(\mathcal{C})\setminus\mathcal{C}\subseteq\mathrm{pa}(\tilde{\mathcal{A}})\setminus\tilde{\mathcal{A}}$. Indeed, for $c\in\mathrm{pa}(\mathcal{C})$: if $c\in\mathcal{O}$ then $c\in\mathcal{C}$ because $\mathcal{A}$ and $\tilde{\mathcal{A}}$ are both ancestral in $\mathcal{G}(\mathcal{M})_{\mathcal{O}}$, while if $c\notin\mathcal{O}$ then $c\notin\mathcal{A}$ and $c\notin\tilde{\mathcal{A}}$. Hence by Lemma E.1, for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost all $\boldsymbol{e}\in\boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}\in\boldsymbol{\mathcal{X}}$

$$(\boldsymbol{g}_{\mathcal{A}})_{\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{A})})=(\boldsymbol{g}_{\tilde{\mathcal{A}}})_{\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\tilde{\mathcal{A}})\setminus\tilde{\mathcal{A}}},\boldsymbol{e}_{\mathrm{pa}(\tilde{\mathcal{A}})})\,.$$

Hence for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}\in\boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}\in\boldsymbol{\mathcal{X}}$

$$\begin{aligned}
&\boldsymbol{x}_{\mathcal{D}}=\boldsymbol{f}_{\mathcal{D}}(\boldsymbol{x},\boldsymbol{e})\\
\iff&\begin{cases}\boldsymbol{x}_{\mathcal{A}\setminus\mathcal{C}} &=\boldsymbol{f}_{\mathcal{A}\setminus\mathcal{C}}(\boldsymbol{x},\boldsymbol{e})\\ \boldsymbol{x}_{\mathcal{C}} &=\boldsymbol{f}_{\mathcal{C}}(\boldsymbol{x},\boldsymbol{e})\\ \boldsymbol{x}_{\mathcal{C}} &=\boldsymbol{f}_{\mathcal{C}}(\boldsymbol{x},\boldsymbol{e})\\ \boldsymbol{x}_{\tilde{\mathcal{A}}\setminus\mathcal{C}} &=\boldsymbol{f}_{\tilde{\mathcal{A}}\setminus\mathcal{C}}(\boldsymbol{x},\boldsymbol{e})\end{cases}\\
\iff&\begin{cases}\boldsymbol{x}_{\mathcal{A}\setminus\mathcal{C}} &=(\boldsymbol{g}_{\mathcal{A}})_{\mathcal{A}\setminus\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{A})})\\ \boldsymbol{x}_{\mathcal{C}} &=(\boldsymbol{g}_{\mathcal{A}})_{\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{A})})\\ \boldsymbol{x}_{\mathcal{C}} &=(\boldsymbol{g}_{\tilde{\mathcal{A}}})_{\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\tilde{\mathcal{A}})\setminus\tilde{\mathcal{A}}},\boldsymbol{e}_{\mathrm{pa}(\tilde{\mathcal{A}})})\\ \boldsymbol{x}_{\tilde{\mathcal{A}}\setminus\mathcal{C}} &=(\boldsymbol{g}_{\tilde{\mathcal{A}}})_{\tilde{\mathcal{A}}\setminus\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\tilde{\mathcal{A}})\setminus\tilde{\mathcal{A}}},\boldsymbol{e}_{\mathrm{pa}(\tilde{\mathcal{A}})})\end{cases}\\
\iff&\begin{cases}\boldsymbol{x}_{\mathcal{A}} &=\boldsymbol{g}_{\mathcal{A}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{A})})\\ \boldsymbol{x}_{\tilde{\mathcal{A}}} &=\boldsymbol{g}_{\tilde{\mathcal{A}}}(\boldsymbol{x}_{\mathrm{pa}(\tilde{\mathcal{A}})\setminus\tilde{\mathcal{A}}},\boldsymbol{e}_{\mathrm{pa}(\tilde{\mathcal{A}})})\,.\end{cases}
\end{aligned}$$

Now $\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}\subseteq\mathrm{pa}(\mathcal{D})\setminus\mathcal{D}$, and similarly, $\mathrm{pa}(\tilde{\mathcal{A}})\setminus\tilde{\mathcal{A}}\subseteq\mathrm{pa}(\mathcal{D})\setminus\mathcal{D}$. Hence, we conclude that the mapping $\boldsymbol{h}_{\mathcal{D}}:\boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{D})\setminus\mathcal{D}}\times\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{D})}\to\boldsymbol{\mathcal{X}}_{\mathcal{D}}$ defined by

$$\begin{aligned}&\boldsymbol{h}_{\mathcal{D}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{D})\setminus\mathcal{D}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{D})}):=\\ &\quad\big((\boldsymbol{g}_{\mathcal{A}})_{\mathcal{A}\setminus\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{A})}),(\boldsymbol{g}_{\mathcal{A}})_{\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{A})\setminus\mathcal{A}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{A})}),(\boldsymbol{g}_{\tilde{\mathcal{A}}})_{\tilde{\mathcal{A}}\setminus\mathcal{C}}(\boldsymbol{x}_{\mathrm{pa}(\tilde{\mathcal{A}})\setminus\tilde{\mathcal{A}}},\boldsymbol{e}_{\mathrm{pa}(\tilde{\mathcal{A}})})\big)\end{aligned}$$

is a measurable solution function for $\mathcal{M}$ w.r.t. $\mathcal{D}$, and that $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{D}$. □

PROOF OF COROLLARY B.5. It suffices to show the implication to the left. We have to show that $\mathcal{M}$ is uniquely solvable w.r.t. each ancestral subset of $\mathcal{G}(\mathcal{M})_{\mathcal{O}}$. The proof proceeds via induction with respect to the size of the ancestral subset. For ancestral subsets of size 0, the claim is trivially true. Ancestral subsets of size 1 must be of the form $\{i\}=\mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(i)$ for $i\in\mathcal{O}$ and hence the claim is true by assumption. Assume that the claim holds for all ancestral subsets of size $\leq n$. Let $\mathcal{A}$ be an ancestral subset of $\mathcal{G}(\mathcal{M})_{\mathcal{O}}$ of size $n+1$. If $\mathcal{A}=\mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(i)$ for some $i\in\mathcal{O}$ then the claim holds for $\mathcal{A}$ by assumption. Otherwise, $\mathcal{A}=\bigcup_{i\in\mathcal{A}}\mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(i)$ is a union of ancestral subsets of size $\leq n$. Choose distinct elements $\{i_1,\dots,i_k\}\subseteq\mathcal{A}$ where k is the smallest integer such that $\bigcup_{j=1}^{k}\mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(i_j)=\mathcal{A}$. By applying Proposition B.4 to $\bigcup_{j=1}^{k-1}\mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(i_j)$ and $\mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(i_k)$, thereby noting that the intersection of these two sets is an ancestral subset of size $\leq n$ and making use of the induction hypothesis, we arrive at the conclusion that $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{A}$. □

Appendix C.

PROOF OF PROPOSITION C.2. Let $\boldsymbol{e}\in\boldsymbol{\mathcal{E}}$ and $\boldsymbol{x}_{\mathcal{O}}\in\boldsymbol{\mathcal{X}}_{\mathcal{O}}$. For $\boldsymbol{x}_{\mathcal{L}}\in\boldsymbol{\mathcal{X}}$,

$$\begin{aligned}
&\boldsymbol{x}_{\mathcal{L}}=\boldsymbol{f}_{\mathcal{L}}(\boldsymbol{x},\boldsymbol{e})\\
&\iff \boldsymbol{x}_{\mathcal{L}}=B_{\mathcal{L}\mathcal{L}}\boldsymbol{x}_{\mathcal{L}}+B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}}+\Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e}\\
&\iff A_{\mathcal{L}\mathcal{L}}\boldsymbol{x}_{\mathcal{L}}=B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}}+\Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e}\\
&\iff \begin{cases}A_{\mathcal{L}\mathcal{L}}A^{+}_{\mathcal{L}\mathcal{L}}(B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}}+\Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e})=B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}}+\Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e}\\ \exists_{\boldsymbol{v}\in\boldsymbol{\mathcal{X}}_{\mathcal{L}}}:\boldsymbol{x}_{\mathcal{L}}=A^{+}_{\mathcal{L}\mathcal{L}}(B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}}+\Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e})+[\mathbb{I}_{\mathcal{L}}-A^{+}_{\mathcal{L}\mathcal{L}}A_{\mathcal{L}\mathcal{L}}]\boldsymbol{v}\,,\end{cases}
\end{aligned}$$

where the last equivalence follows from [Theorem 2, 22]. □

PROOF OF PROPOSITION C.3. $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$ if and only if for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}_{\mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ the linear system of equations

$$\begin{aligned} &\boldsymbol{x}_{\mathcal{L}} = \boldsymbol{f}_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{e}) \\ \iff &\boldsymbol{x}_{\mathcal{L}} = B_{\mathcal{L}\mathcal{L}}\boldsymbol{x}_{\mathcal{L}} + B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}} + \Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e} \\ \iff &A_{\mathcal{L}\mathcal{L}}\boldsymbol{x}_{\mathcal{L}} = B_{\mathcal{L}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}} + \Gamma_{\mathcal{L}\mathcal{J}}\boldsymbol{e} \end{aligned}$$

has a unique solution $\boldsymbol{x}_{\mathcal{L}} \in \mathcal{X}_{\mathcal{L}}$. Hence, $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$ if and only if $A_{\mathcal{L}\mathcal{L}}$ is invertible. □

PROOF OF PROPOSITION C.4. It suffices to show $(1) \implies (2)$ and $(1) \iff (3)$. We start by showing that $(1) \implies (2)$. Let $\mathcal{V} \subseteq \mathcal{L}$ and denote $\mathcal{U} := \mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{L}}}(\mathcal{V})$, then we need to show that $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{U}$. From Proposition C.3 we know that $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$ if and only if the matrix $A_{\mathcal{L}\mathcal{L}} = \mathbb{I}_{\mathcal{L}} - B_{\mathcal{L}\mathcal{L}}$ is invertible. The matrix $A_{\mathcal{L}\mathcal{L}}$ is invertible if and only if the rows of $A_{\mathcal{L}\mathcal{L}}$ are all linearly independent. In particular, the rows of $A_{\mathcal{U}\mathcal{L}}$ are all linearly independent. Because $A_{\mathcal{U}\mathcal{L}} = [A_{\mathcal{U}\mathcal{U}}\, Z_{\mathcal{U}\mathcal{L}}]$, where $Z_{\mathcal{U}\mathcal{L}}$ is the zero matrix, we know that the rows of $A_{\mathcal{U}\mathcal{U}} = \mathbb{I}_{\mathcal{U}} - B_{\mathcal{U}\mathcal{U}}$ are also all linearly independent, and hence $A_{\mathcal{U}\mathcal{U}}$ is invertible.

Next, we show that $(1) \iff (3)$. Observe that the strongly connected components of $\mathcal{G}(\mathcal{M})_{\mathcal{L}}$ form a partition of the set $\mathcal{L}$ and that the directed mixed graph $\mathcal{G}(\mathcal{M})_{\mathcal{L}}$ and the directed graph $\mathcal{G}^a(\mathcal{M})_{\mathcal{L}}$ have the same strongly connected components. Because, by Lemma A.2, the graph of strongly connected components $\mathcal{G}^{\mathrm{sc}}$ of the directed graph $\mathcal{G}^a(\mathcal{M})_{\mathcal{L}}$ is a DAG, the square matrix $B_{\mathcal{L}\mathcal{L}}$ can be permuted to an upper triangular block matrix $\tilde{B}_{\mathcal{L}\mathcal{L}}$, where for each diagonal block $\tilde{B}_{\mathcal{V}\mathcal{V}}$ of $\tilde{B}_{\mathcal{L}\mathcal{L}}$ the set of nodes $\mathcal{V}$ is a strongly connected component in $\mathcal{G}(\mathcal{M})_{\mathcal{L}}$.

Without loss of generality we assume now that $B_{\mathcal{L}\mathcal{L}}$ is an upper triangular block matrix. From Proposition C.3 it follows that $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$ if and only if the matrix $A_{\mathcal{L}\mathcal{L}} = \mathbb{I}_{\mathcal{L}} - B_{\mathcal{L}\mathcal{L}}$ is invertible. Because $B_{\mathcal{L}\mathcal{L}}$ is an upper triangular block matrix, we know that $A_{\mathcal{L}\mathcal{L}}$ is an upper triangular block matrix, where for each diagonal block $A_{\mathcal{V}\mathcal{V}}$ of $A_{\mathcal{L}\mathcal{L}}$ the set of nodes $\mathcal{V}$ is a strongly connected component in $\mathcal{G}(\mathcal{M})_{\mathcal{L}}$. Since an upper triangular block matrix $A_{\mathcal{L}\mathcal{L}}$ is invertible if and only if every diagonal block in $A_{\mathcal{L}\mathcal{L}}$ is invertible, we have that $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$ if and only if $\mathcal{M}$ is uniquely solvable w.r.t. each strongly connected component in $\mathcal{G}(\mathcal{M})_{\mathcal{L}}$. □

PROOF OF PROPOSITION C.5. By the definition of marginalization and Proposition C.3 the marginal causal mechanism $\tilde{\boldsymbol{f}}$ is given by

$$\begin{aligned} \tilde{\boldsymbol{f}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) &:= \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}), \boldsymbol{e}) \\ &= B_{\mathcal{O}\mathcal{O}}\boldsymbol{x}_{\mathcal{O}} + B_{\mathcal{O}\mathcal{L}}\boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) + \Gamma_{\mathcal{O}\mathcal{J}}\boldsymbol{e} \\ &= [B_{\mathcal{O}\mathcal{O}} + B_{\mathcal{O}\mathcal{L}}A_{\mathcal{L}\mathcal{L}}^{-1}B_{\mathcal{L}\mathcal{O}}]\boldsymbol{x}_{\mathcal{O}} + [B_{\mathcal{O}\mathcal{L}}A_{\mathcal{L}\mathcal{L}}^{-1}\Gamma_{\mathcal{L}\mathcal{J}} + \Gamma_{\mathcal{O}\mathcal{J}}]\boldsymbol{e}\,. \end{aligned}$$

From Propositions C.4 and 5.11 it follows that the marginalization respects the latent projection. □

E.2. Proofs of the main text.

Section 2.

PROOF OF PROPOSITION 2.11. Let $i \in \mathcal{I}$. Note that Definition 2.6 can alternatively be formulated as follows: for $k \in \mathcal{I} \cup \mathcal{J}$, $k \notin \mathrm{pa}(i)$ if and only if there exists a measurable mapping $\hat{f}_i : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \to \mathcal{X}_i$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$,

$$x_i = f_i(\boldsymbol{x}, \boldsymbol{e}) \iff x_i = \hat{f}_i(\boldsymbol{x}, \boldsymbol{e})$$

and either $k \in \mathcal{I}$ and there exists $\hat{x}_k \in \mathcal{X}_k$ such that $\hat{f}_i(\boldsymbol{x}, \boldsymbol{e}) = \hat{f}_i(\boldsymbol{x}_{\setminus k}, \hat{x}_k, \boldsymbol{e})$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}, \boldsymbol{e} \in \boldsymbol{\mathcal{E}}$, or $k \in \mathcal{J}$ and there exists $\hat{e}_k \in \mathcal{E}_k$ such that $\hat{f}_i(\boldsymbol{x}, \boldsymbol{e}) = \hat{f}_i(\boldsymbol{x}, \boldsymbol{e}_{\setminus k}, \hat{e}_k)$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}, \boldsymbol{e} \in \boldsymbol{\mathcal{E}}$. By repeatedly applying (this formulation of) Definition 2.6 to all $k \notin \mathrm{pa}(i)$, we obtain the existence of a measurable mapping $\tilde{f}_i : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \to \mathcal{X}_i$ and $\hat{\boldsymbol{x}}_{\setminus \mathrm{pa}(i)} \in \boldsymbol{\mathcal{X}}_{\setminus \mathrm{pa}(i)}$, $\hat{\boldsymbol{e}}_{\setminus \mathrm{pa}(i)} \in \boldsymbol{\mathcal{E}}_{\setminus \mathrm{pa}(i)}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$,

$$x_i = f_i(\boldsymbol{x}, \boldsymbol{e}) \iff x_i = \tilde{f}_i(\boldsymbol{x}, \boldsymbol{e}),$$

and for all $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$,

$$\tilde{f}_i(\boldsymbol{x}, \boldsymbol{e}) = \tilde{f}_i(\boldsymbol{x}_{\mathrm{pa}(i)}, \hat{\boldsymbol{x}}_{\setminus \mathrm{pa}(i)}, \boldsymbol{e}_{\mathrm{pa}(i)}, \hat{\boldsymbol{e}}_{\setminus \mathrm{pa}(i)}).$$

Define the SCM $\tilde{\mathcal{M}}$ as $\mathcal{M}$ except that its causal mechanism is $\tilde{\boldsymbol{f}}$ instead of $\boldsymbol{f}$. Then $\tilde{\mathcal{M}}$ is structurally minimal and equivalent to $\mathcal{M}$. □

PROOF OF PROPOSITION 2.14. The $\mathrm{do}(I, \boldsymbol{\xi}_I)$ operation on $\mathcal{M}$ completely removes the functional dependence on $\boldsymbol{x}$ and $\boldsymbol{e}$ from the f_i components for $i \in I$ and hence the corresponding incoming directed and bidirected edges on nodes in I from the (augmented) graph. □

PROOF OF PROPOSITION 2.15. The first statement follows from Definitions 2.12 and 2.13. For the second statement, note that a perfect intervention can only remove parental relations, and therefore will never introduce a cycle. □

PROOF OF PROPOSITION 2.19. This follows directly from Definitions 2.17 and 2.18. □

PROOF OF PROPOSITION 2.20. The additional edges introduced by the twin operation cannot lead to a directed cycle involving both copied and original nodes, because there are no edges pointing from copied nodes to original nodes (i.e., of the form $i' \to v$ with $i' \in I'$ and $v \in \mathcal{V}$). Directed cycles involving only original nodes are absent by assumption, and directed cycles involving only copied nodes as well since they would correspond with a directed cycle in the original directed graph. □

PROOF OF PROPOSITION 2.21. It suffices to prove the property for directed graphs, since the property for SCMs follows directly from Definitions 2.12 and 2.17.

Applying the intervention $\mathrm{do}(I)$ on the graph $\mathcal{G}$ removes all the incoming edges from the nodes in I. Now, if we perform the twin operation w.r.t. $\mathcal{I}$ on this graph $\mathrm{do}(I)(\mathcal{G})$, then we copy the same edges as if we had twinned the graph $\mathcal{G}$ w.r.t. $\mathcal{I}$, except those edges that do point to one of the nodes in I. Hence, if we apply the intervention $\mathrm{do}(I \cup I')$ on the graph $\mathrm{twin}(\mathcal{I})(\mathcal{G})$, which removes all incoming edges of both I and its copy I', then we clearly obtain the same graph. □

Section 3.

PROOF OF THEOREM 3.2. First we define the solution space $\boldsymbol{\mathcal{S}}(\mathcal{M})$ of $\mathcal{M}$ by

$$\boldsymbol{\mathcal{S}}(\mathcal{M}) := \{(\boldsymbol{e}, \boldsymbol{x}) \in \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{X}} : \boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e})\}.$$

This is a measurable set, since $\boldsymbol{\mathcal{S}}(\mathcal{M}) = \boldsymbol{h}^{-1}(\Delta)$, where $\boldsymbol{h} : \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{X}}$ is the measurable mapping defined by $\boldsymbol{h}(\boldsymbol{e}, \boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e}))$ and Δ is the set defined by $\{(\boldsymbol{x}, \boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{\mathcal{X}}\}$, which is measurable since $\boldsymbol{\mathcal{X}}$ is Hausdorff. Note that

$$\boldsymbol{\mathcal{A}} := \boldsymbol{pr}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{S}}(\mathcal{M})) = \{\boldsymbol{e} \in \boldsymbol{\mathcal{E}} : \exists \boldsymbol{x} \in \boldsymbol{\mathcal{X}} \text{ s.t. } \boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e})\},$$

is an analytic set because the projection $\boldsymbol{pr}_{\boldsymbol{\mathcal{E}}} : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{E}}$ is a measurable mapping between standard measurable spaces (Lemma F.3).

Suppose that (1) holds, that is, $\mathcal{M}$ has a solution. Then there exists a pair of random variables $(\boldsymbol{E}, \boldsymbol{X}) : \Omega \to \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{X}}$ such that $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{E})$ $\mathbb{P}$-a.s.. Note that

$$\begin{aligned}\{\omega \in \Omega : \boldsymbol{X}(\omega) = \boldsymbol{f}\big(\boldsymbol{X}(\omega), \boldsymbol{E}(\omega)\big)\} &\subseteq \{\omega \in \Omega : \exists \boldsymbol{x} \in \boldsymbol{\mathcal{X}} \text{ s.t. } \boldsymbol{x} = \boldsymbol{f}\big(\boldsymbol{x}, \boldsymbol{E}(\omega)\big)\} \\ &\subseteq \boldsymbol{E}^{-1}\Big(\{\boldsymbol{e} \in \boldsymbol{\mathcal{E}} : \exists \boldsymbol{x} \in \boldsymbol{\mathcal{X}} \text{ s.t. } \boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e})\}\Big) \\ &= \boldsymbol{E}^{-1}(\boldsymbol{\mathcal{A}}).\end{aligned}$$

By Lemma F.6, $\boldsymbol{\mathcal{A}}$ is $\mathbb{P}^{\boldsymbol{E}}$-measurable because it is analytic, and we can write $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{B}} \dot\cup \boldsymbol{\mathcal{N}}$ with $\boldsymbol{\mathcal{B}} \subseteq \boldsymbol{\mathcal{E}}$ measurable and $\boldsymbol{\mathcal{N}}$ a $\mathbb{P}^{\boldsymbol{E}}$-null set. Hence $\boldsymbol{E}^{-1}(\boldsymbol{\mathcal{A}}) = \boldsymbol{E}^{-1}(\boldsymbol{\mathcal{B}}) \cup \boldsymbol{E}^{-1}(\boldsymbol{\mathcal{N}})$ where $\boldsymbol{E}^{-1}(\boldsymbol{\mathcal{N}})$ is a $\mathbb{P}$-null set. Therefore,

$$\boldsymbol{E}^{-1}(\boldsymbol{\mathcal{B}}) \supseteq \{\omega \in \Omega : \boldsymbol{X}(\omega) = \boldsymbol{f}\big(\boldsymbol{X}(\omega), \boldsymbol{E}(\omega)\big)\} \setminus \boldsymbol{E}^{-1}(\boldsymbol{\mathcal{N}})$$

which implies that $\mathbb{P}(\boldsymbol{E}^{-1}(\boldsymbol{\mathcal{B}})) = 1$. Hence, $\boldsymbol{\mathcal{E}} \setminus \boldsymbol{\mathcal{A}}$ is a $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-null set. In other words, for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ the structural equations $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e})$ have a solution $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$, that is, (2) holds.

Suppose that (2) holds. Then $\boldsymbol{\mathcal{E}} \setminus \boldsymbol{pr}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{S}}(\mathcal{M}))$ is a $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-null set. By application of the measurable selection theorem F.8, there exists a measurable $\boldsymbol{g} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost all $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$, $\boldsymbol{g}(\boldsymbol{e}) = \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{e}), \boldsymbol{e})$. Hence, there exists a measurable mapping $\boldsymbol{g} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x} = \boldsymbol{g}(\boldsymbol{e}) \quad \Longrightarrow \quad \boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e}),$$

which we call property (A). Let $\tilde{\boldsymbol{f}} : \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}$ be the causal mechanism of a structurally minimal SCM that is equivalent to $\mathcal{M}$ (see Proposition 2.11). In particular, for any $\boldsymbol{\epsilon}_{\setminus \text{pa}(\mathcal{I})} \in \boldsymbol{\mathcal{E}}_{\setminus \text{pa}(\mathcal{I})}$, we have that $\tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{e}) = \tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{e}_{\text{pa}(\mathcal{I})}, \boldsymbol{\epsilon}_{\setminus \text{pa}(\mathcal{I})})$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ and all $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$. This means that we may also consider $\tilde{\boldsymbol{f}}$ as a mapping $\tilde{\boldsymbol{f}} : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})} \to \boldsymbol{\mathcal{X}}$. By applying Lemma F.10 to the canonical projection $\boldsymbol{pr}_{\boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})}} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})}$ and using the equivalence of $\boldsymbol{f}$ and $\tilde{\boldsymbol{f}}$, we obtain that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})}}$-almost all $\boldsymbol{e}_{\text{pa}(\mathcal{I})} \in \boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})}$ there exists $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ with $\boldsymbol{x} = \tilde{\boldsymbol{f}}(\boldsymbol{x}, \boldsymbol{e}_{\text{pa}(\mathcal{I})})$. By applying the implication (2) $\Longrightarrow$ (A) to $\boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})}$ and $\tilde{\boldsymbol{f}}$, we conclude the existence of a measurable $\boldsymbol{g} : \boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})} \to \boldsymbol{\mathcal{X}}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})}}$-almost all $\boldsymbol{e}_{\text{pa}(\mathcal{I})} \in \boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})}$, $\boldsymbol{g}(\boldsymbol{e}_{\text{pa}(\mathcal{I})}) = \tilde{\boldsymbol{f}}(\boldsymbol{g}(\boldsymbol{e}_{\text{pa}(\mathcal{I})}), \boldsymbol{e}_{\text{pa}(\mathcal{I})})$. Once more using Lemma F.10, we obtain that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost all $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$, $\boldsymbol{g}(\boldsymbol{e}_{\text{pa}(\mathcal{I})}) = \boldsymbol{f}(\boldsymbol{g}(\boldsymbol{e}_{\text{pa}(\mathcal{I})}), \boldsymbol{e})$. In other words, (3) holds.

Lastly, suppose that (3) holds, that is there exists a measurable solution function $\boldsymbol{g} : \boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{I})} \to \boldsymbol{\mathcal{X}}$. Then the measurable mappings $\boldsymbol{E} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{E}}$ and $\boldsymbol{X} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$, defined by $\boldsymbol{E}(\boldsymbol{e}) := \boldsymbol{e}$ and $\boldsymbol{X}(\boldsymbol{e}) := \boldsymbol{g}(\boldsymbol{e}_{\text{pa}(\mathcal{I})})$, respectively, define a pair of random variables $(\boldsymbol{X}, \boldsymbol{E})$ such that $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{E})$ holds a.s. and hence $(\boldsymbol{X}, \boldsymbol{E})$ is a solution. Hence (1) holds. □

PROOF OF PROPOSITION 3.4. Let $\tilde{\boldsymbol{f}} : \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}$ be the causal mechanism of a structurally minimal SCM $\tilde{\mathcal{M}}$ that is equivalent to $\mathcal{M}$ (see Proposition 2.11). For a subset $\mathcal{O} \subseteq \mathcal{I}$ consider the induced subgraph $\mathcal{G}^a(\mathcal{M})_{\mathcal{O}}$ of the augmented graph $\mathcal{G}^a(\mathcal{M})$ on $\mathcal{O}$. Then the acyclicity of $\mathcal{G}^a(\mathcal{M})$ implies that the induced subgraph $\mathcal{G}^a(\mathcal{M})_{\mathcal{O}}$ is acyclic, and hence there exists a topological ordering on the nodes $\mathcal{O}$. We can substitute the components $\tilde{f}_i$ of the causal mechanism $\tilde{\boldsymbol{f}}$ for $i \in \mathcal{O}$ into each other along this topological ordering. This gives a measurable solution function $\boldsymbol{g}_{\mathcal{O}} : \boldsymbol{\mathcal{X}}_{\text{pa}(\mathcal{O}) \setminus \mathcal{O}} \times \boldsymbol{\mathcal{E}}_{\text{pa}(\mathcal{O})} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ for $\tilde{\mathcal{M}}$, and hence for $\mathcal{M}$. It is clear from the acyclic structure that this mapping $\boldsymbol{g}_{\mathcal{O}}$ is independent of the choice of the topological ordering and is the only solution function for $\mathcal{M}$. Therefore, $\tilde{\mathcal{M}}$ is uniquely solvable w.r.t. $\mathcal{O}$, and so is $\mathcal{M}$. □

PROOF OF PROPOSITION 3.7. This follows immediately from Definitions 2.7 and 3.3. □

PROOF OF THEOREM 3.6. Suppose that (1) holds. By Proposition B.1 there exists a measurable solution function $\boldsymbol{g}_{\mathcal{O}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ for $\mathcal{M}$ w.r.t. $\mathcal{O}$. Then for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}_{\setminus\mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\setminus\mathcal{O}}$ we have that $\boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})})$ is a solution of $\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e})$. Hence, because of (1), for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}_{\setminus\mathcal{O}} \in \boldsymbol{\mathcal{X}}_{\setminus\mathcal{O}}$ we have that $\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e})$ implies $\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})})$. Thus, $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{O}$, that is, (2) holds.

Suppose that (2) holds. Let $\boldsymbol{g}_{\mathcal{O}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ be a measurable solution function for $\mathcal{M}$ w.r.t. $\mathcal{O}$. Then, for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \quad \Longleftrightarrow \quad \boldsymbol{x}_{\mathcal{O}} = \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e})\,.$$

This implies (1).

For the last statement, assume that $\mathcal{M}$ is uniquely solvable. Let $\boldsymbol{g} : \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{I})} \to \boldsymbol{\mathcal{X}}$ be a measurable solution function. Then there exists a measurable set $\boldsymbol{B} \subseteq \boldsymbol{\mathcal{E}}$ with $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{B}) = 1$ and for all $\boldsymbol{e} \in \boldsymbol{B}$,

$$\forall \boldsymbol{x} \in \boldsymbol{\mathcal{X}} : \boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{e}) \implies \boldsymbol{x} = \boldsymbol{g}(\boldsymbol{e}_{\mathrm{pa}(\mathcal{I})}).$$

The existence of a solution for $\mathcal{M}$ follows directly from Theorem 3.2. Each solution $(\boldsymbol{X}, \boldsymbol{E}) : \Omega \to \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}}$ of $\mathcal{M}$ satisfies $\boldsymbol{X}(\omega) = \boldsymbol{f}(\boldsymbol{X}(\omega), \boldsymbol{E}(\omega))$ $\mathbb{P}$-a.s.. In addition, it satisfies $\boldsymbol{E}(\omega) \in \boldsymbol{B}$ $\mathbb{P}$-a.s., since $\mathbb{P} \circ \boldsymbol{E}^{-1} = \mathbb{P}_{\boldsymbol{\mathcal{E}}}$. Hence, it satisfies $\boldsymbol{X}(\omega) = \boldsymbol{g}(\boldsymbol{E}(\omega)_{\mathrm{pa}(\mathcal{I})})$ $\mathbb{P}$-a.s.. Thus for every solution $(\boldsymbol{X}, \boldsymbol{E})$ the associated observational distribution is the push-forward of $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$ under $\boldsymbol{g} \circ \boldsymbol{pr}_{\mathrm{pa}(\mathcal{I})}$. □

PROOF OF PROPOSITION 3.8. Let $\boldsymbol{g}_{\mathcal{O}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ be a measurable solution function for $\mathcal{M}$ w.r.t. $\mathcal{O}$. Then the mapping $\tilde{\boldsymbol{g}}_{\mathcal{O}\cup I} : \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}\cup I}$ defined by $\tilde{\boldsymbol{g}}_{\mathcal{O}\cup I}(\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) := (\boldsymbol{g}_{\mathcal{O}}(\boldsymbol{\xi}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}), \boldsymbol{\xi}_I)$ is a measurable solution function for the SCM $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ w.r.t. $\mathcal{O} \cup I$. If $\mathcal{M}$ is (uniquely) solvable w.r.t. $\mathcal{O}$, then it follows that $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ is (uniquely) solvable w.r.t. $\mathcal{O} \cup I$. □

PROOF OF PROPOSITION 3.10. It suffices to show that solvability of $\mathcal{M}$ w.r.t. $\mathcal{O}$ implies ancestral solvability w.r.t. $\mathcal{O}$. Solvability of $\mathcal{M}$ w.r.t. $\mathcal{O}$ implies that there exists a measurable mapping $\boldsymbol{g}_{\mathcal{O}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{O}} = \boldsymbol{g}_{\mathcal{O}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \quad \Longrightarrow \quad \boldsymbol{x}_{\mathcal{O}} = \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e})\,.$$

Let $\tilde{\boldsymbol{f}} : \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}$ be the causal mechanism of a structurally minimal SCM $\tilde{\mathcal{M}}$ that is equivalent to $\mathcal{M}$ (see Proposition 2.11). Let $\mathcal{P} := \mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(\mathcal{A})$ for some $\mathcal{A} \subseteq \mathcal{O}$. Then for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\begin{cases} \boldsymbol{x}_{\mathcal{P}} & = (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \\ \boldsymbol{x}_{\mathcal{O}\setminus\mathcal{P}} & = (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{O}\setminus\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \end{cases} \Longrightarrow \begin{cases} \boldsymbol{x}_{\mathcal{P}} & = \tilde{\boldsymbol{f}}_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{P})}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{P})}) \\ \boldsymbol{x}_{\mathcal{O}\setminus\mathcal{P}} & = \tilde{\boldsymbol{f}}_{\mathcal{O}\setminus\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O}\setminus\mathcal{P})}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O}\setminus\mathcal{P})})\,. \end{cases}$$

Since $\mathrm{pa}(\mathcal{P}) \setminus \mathcal{P} \subseteq \mathrm{pa}(\mathcal{O}) \setminus \mathcal{O}$, we have that in particular for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{P}} = (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) \quad \Longrightarrow \quad \boldsymbol{x}_{\mathcal{P}} = \tilde{\boldsymbol{f}}_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{P})}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{P})})\,.$$

This implies that the mapping $(\boldsymbol{g}_{\mathcal{O}})_{\mathcal{P}}$ cannot depend on elements different from $\mathrm{pa}(\mathcal{P})$. Moreover, it follows from the definition of $\mathcal{P}$ that $(\mathrm{pa}(\mathcal{O}) \setminus \mathcal{O}) \cap \mathrm{pa}(\mathcal{P}) = \mathrm{pa}(\mathcal{P}) \setminus \mathcal{P}$ and thus we have $\mathrm{pa}(\mathcal{O}) \setminus \mathcal{O} = (\mathrm{pa}(\mathcal{P}) \setminus \mathcal{P}) \cup (\mathrm{pa}(\mathcal{O}) \setminus (\mathcal{O} \cup \mathrm{pa}(\mathcal{P})))$. Now, pick an element $\hat{\boldsymbol{x}}_{\mathrm{pa}(\mathcal{O})\setminus(\mathcal{O}\cup\mathrm{pa}(\mathcal{P}))} \in \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus(\mathcal{O}\cup\mathrm{pa}(\mathcal{P}))}$ and define the mapping $\tilde{\boldsymbol{g}}_{\mathcal{P}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{P})\setminus\mathcal{P}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{P})} \to \boldsymbol{\mathcal{X}}_{\mathcal{P}}$ by

$$\tilde{\boldsymbol{g}}_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{P})\setminus\mathcal{P}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{P})}) := (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{P})\setminus\mathcal{P}}, \hat{\boldsymbol{x}}_{\mathrm{pa}(\mathcal{O})\setminus(\mathcal{O}\cup\mathrm{pa}(\mathcal{P}))}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) .$$

Then, for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{P}} = \tilde{\boldsymbol{g}}_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{P})\setminus\mathcal{P}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{P})}) \quad \iff \quad \boldsymbol{x}_{\mathcal{P}} = (\boldsymbol{g}_{\mathcal{O}})_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}) .$$

Together this gives that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{P}} = \tilde{\boldsymbol{g}}_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{P})\setminus\mathcal{P}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{P})}) \quad \implies \quad \boldsymbol{x}_{\mathcal{P}} = \tilde{\boldsymbol{f}}_{\mathcal{P}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{P})}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{P})}) .$$

which is equivalent to the statement that $\mathcal{M}$ is solvable w.r.t. $\mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{O}}}(\mathcal{A})$. □

Section 4.

LEMMA E.1. *Let $\mathcal{M}$ be an SCM that is uniquely solvable w.r.t. two subsets $A, B \subseteq \mathcal{I}$ that satisfy $A \subseteq B$ and $\mathrm{pa}(A) \setminus A \subseteq \mathrm{pa}(B) \setminus B$. Let $\boldsymbol{g}_A : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(A)\setminus A} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(A)} \to \boldsymbol{\mathcal{X}}_A$ and $\boldsymbol{g}_B : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(B)\setminus B} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(B)} \to \boldsymbol{\mathcal{X}}_B$ be measurable solution functions for $\mathcal{M}$ w.r.t. A and B, respectively. Then for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$*

$$\boldsymbol{g}_A(\boldsymbol{x}_{\mathrm{pa}(A)\setminus A}, \boldsymbol{e}_{\mathrm{pa}(A)}) = (\boldsymbol{g}_B)_A(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) .$$

PROOF. Without loss of generality, we assume that $\mathcal{M}$ is structurally minimal (see Proposition 2.11). Let $\bar{\boldsymbol{\mathcal{E}}} \subseteq \boldsymbol{\mathcal{E}}$ be a measurable set with $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\bar{\boldsymbol{\mathcal{E}}}) = 1$ such that for all $\boldsymbol{e} \in \bar{\boldsymbol{\mathcal{E}}}$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$:

$$\boldsymbol{x}_A = \boldsymbol{g}_A(\boldsymbol{x}_{\mathrm{pa}(A)\setminus A}, \boldsymbol{e}_{\mathrm{pa}(A)}) \iff \boldsymbol{x}_A = \boldsymbol{f}_A(\boldsymbol{x}_{\mathrm{pa}(A)}, \boldsymbol{e}_{\mathrm{pa}(A)})$$

and

$$\boldsymbol{x}_B = \boldsymbol{g}_B(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) \iff \boldsymbol{x}_B = \boldsymbol{f}_B(\boldsymbol{x}_{\mathrm{pa}(B)}, \boldsymbol{e}_{\mathrm{pa}(B)}) .$$

Now let $\boldsymbol{e} \in \bar{\boldsymbol{\mathcal{E}}}$ and let $\boldsymbol{x}_{A\cup\mathrm{pa}(B)\setminus B} \in \boldsymbol{\mathcal{X}}_{A\cup\mathrm{pa}(B)\setminus B}$. Then

$$\begin{aligned}
&\boldsymbol{x}_A = (\boldsymbol{g}_B)_A(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) \\
&\implies \begin{cases} & \boldsymbol{x}_A = (\boldsymbol{g}_B)_A(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) \\ \exists \boldsymbol{x}_{B\setminus A} \in \boldsymbol{\mathcal{X}}_{B\setminus A} : & \boldsymbol{x}_{B\setminus A} = (\boldsymbol{g}_B)_{B\setminus A}(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) \end{cases} \\
&\implies \exists \boldsymbol{x}_{B\setminus A} \in \boldsymbol{\mathcal{X}}_{B\setminus A} : \quad \boldsymbol{x}_B = \boldsymbol{g}_B(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) \\
&\implies \exists \boldsymbol{x}_{B\setminus A} \in \boldsymbol{\mathcal{X}}_{B\setminus A} : \quad \boldsymbol{x}_B = \boldsymbol{f}_B(\boldsymbol{x}_{\mathrm{pa}(B)}, \boldsymbol{e}_{\mathrm{pa}(B)}) \\
&\implies \exists \boldsymbol{x}_{B\setminus A} \in \boldsymbol{\mathcal{X}}_{B\setminus A} : \quad \boldsymbol{x}_A = \boldsymbol{f}_A(\boldsymbol{x}_{\mathrm{pa}(A)}, \boldsymbol{e}_{\mathrm{pa}(A)}) \\
&\implies \boldsymbol{x}_A = \boldsymbol{f}_A(\boldsymbol{x}_{\mathrm{pa}(A)}, \boldsymbol{e}_{\mathrm{pa}(A)}) \\
&\implies \boldsymbol{x}_A = \boldsymbol{g}_A(\boldsymbol{x}_{\mathrm{pa}(A)\setminus A}, \boldsymbol{e}_{\mathrm{pa}(A)}) ,
\end{aligned}$$

where the exists-quantifier could be omitted because the expression it binds to does not depend on $\boldsymbol{x}_{B\setminus A}$ (from the assumptions it follows that $(A \cup \mathrm{pa}(A)) \cap (B \setminus A) = \emptyset$). Hence, for all $\boldsymbol{e} \in \bar{\boldsymbol{\mathcal{E}}}$ and all $\boldsymbol{x}_{A\cup\mathrm{pa}(B)\setminus B} \in \boldsymbol{\mathcal{X}}_{A\cup\mathrm{pa}(B)\setminus B}$

$$\boldsymbol{x}_A = (\boldsymbol{g}_B)_A(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) \implies \boldsymbol{x}_A = \boldsymbol{g}_A(\boldsymbol{x}_{\mathrm{pa}(A)\setminus A}, \boldsymbol{e}_{\mathrm{pa}(A)}) .$$

Hence, for all $\boldsymbol{e} \in \bar{\boldsymbol{\mathcal{E}}}$ and all $\boldsymbol{x}_{A\cup \mathrm{pa}(B)\setminus B} \in \boldsymbol{\mathcal{X}}_{A\cup \mathrm{pa}(B)\setminus B}$

$$(\boldsymbol{g}_B)_A(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) = \boldsymbol{g}_A(\boldsymbol{x}_{\mathrm{pa}(A)\setminus A}, \boldsymbol{e}_{\mathrm{pa}(A)})\,.$$

Since this expression does not depend on $\boldsymbol{x}_{(B\setminus A)\cup\mathcal{I}\setminus(B\cup \mathrm{pa}(B))}$, from Lemma F.11.(2) we conclude that for all $\boldsymbol{e} \in \bar{\boldsymbol{\mathcal{E}}}$ and all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$(\boldsymbol{g}_B)_A(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) = \boldsymbol{g}_A(\boldsymbol{x}_{\mathrm{pa}(A)\setminus A}, \boldsymbol{e}_{\mathrm{pa}(A)})\,.$$

□

LEMMA E.2. *An SCM $\mathcal{M}$ is observationally equivalent to $\mathcal{M}^{\mathrm{twin}}$ w.r.t. $\mathcal{O} \subseteq \mathcal{I}$.*

PROOF. Let $(\boldsymbol{X}, \boldsymbol{E})$ be a solution of $\mathcal{M}$, then $((\boldsymbol{X}, \boldsymbol{X}), \boldsymbol{E})$ is a solution of $\mathcal{M}^{\mathrm{twin}}$. Conversely, let $((\boldsymbol{X}, \boldsymbol{X}'), \boldsymbol{E})$ be a solution of $\mathcal{M}^{\mathrm{twin}}$, then $(\boldsymbol{X}, \boldsymbol{E})$ is a solution of $\mathcal{M}$. □

PROOF OF PROPOSITION 4.6. First we show that equivalence implies counterfactual equivalence w.r.t. $\mathcal{O}$. The twin operation preserves the equivalence relation on SCMs and since equivalent SCMs are interventionally equivalent w.r.t. every subset, the two equivalent twin SCMs have to be interventionally equivalent w.r.t. $\mathcal{O} \cup \mathcal{O}'$ for every $\mathcal{O} \subseteq \mathcal{I}$ with $\mathcal{O}'$ the copy of $\mathcal{O}$ in $\mathcal{I}'$.

Now, let $\mathcal{M}$ and $\tilde{\mathcal{M}}$ be counterfactually equivalent w.r.t. $\mathcal{O}$. Then $\mathcal{M}^{\mathrm{twin}}$ and $\tilde{\mathcal{M}}^{\mathrm{twin}}$ are interventionally equivalent w.r.t. $\mathcal{O} \cup \mathcal{O}'$. Thus for $I \subseteq \mathcal{O}$, $I' \subseteq \mathcal{O}'$ the copy of I and $\boldsymbol{\xi}_{I'} = \boldsymbol{\xi}_I \in \boldsymbol{\mathcal{X}}_I$, $\mathcal{M}^{\mathrm{twin}}_{\mathrm{do}(I\cup I', \boldsymbol{\xi}_{I\cup I'})}$ and $\tilde{\mathcal{M}}^{\mathrm{twin}}_{\mathrm{do}(I\cup I', \boldsymbol{\xi}_{I\cup I'})}$ are observationally equivalent w.r.t. $\mathcal{O} \cup \mathcal{O}'$. In particular, they are observationally equivalent w.r.t. $\mathcal{O}$. From Proposition 2.21 we have that $\mathcal{M}^{\mathrm{twin}}_{\mathrm{do}(I\cup I', \boldsymbol{\xi}_{I\cup I'})} = (\mathcal{M}_{\mathrm{do}(I, \boldsymbol{\xi}_I)})^{\mathrm{twin}}$ and $\tilde{\mathcal{M}}^{\mathrm{twin}}_{\mathrm{do}(I\cup I', \boldsymbol{\xi}_{I\cup I'})} = (\tilde{\mathcal{M}}_{\mathrm{do}(I, \boldsymbol{\xi}_I)})^{\mathrm{twin}}$, and together with Lemma E.2 this gives that $\mathcal{M}_{\mathrm{do}(I, \boldsymbol{\xi}_I)}$ and $\tilde{\mathcal{M}}_{\mathrm{do}(I, \boldsymbol{\xi}_I)}$ are observationally equivalent w.r.t. $\mathcal{O}$. □

Section 5.

LEMMA E.3. *Let $\mathcal{M}$ be an SCM. Let $B \subseteq \mathcal{I}$ and $A \subseteq \mathcal{I} \cup \mathcal{J}$ such that $(\mathrm{pa}(B) \setminus B) \subseteq A$ and $B \cap A = \emptyset$. Assume that $\boldsymbol{g}_B : \boldsymbol{\mathcal{X}}_A \times \boldsymbol{\mathcal{E}}_A \to \boldsymbol{\mathcal{X}}_B$ is a measurable function such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$*

$$\boldsymbol{x}_B = \boldsymbol{f}_B(\boldsymbol{x}_{\mathrm{pa}(B)}, \boldsymbol{e}_{\mathrm{pa}(B)}) \iff \boldsymbol{x}_B = \boldsymbol{g}_B(\boldsymbol{x}_A, \boldsymbol{e}_A)\,.$$

Then $\mathcal{M}$ is uniquely solvable w.r.t. B.

PROOF. Assume that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_B = \boldsymbol{f}_B(\boldsymbol{x}_{\mathrm{pa}(B)}, \boldsymbol{e}_{\mathrm{pa}(B)}) \iff \boldsymbol{x}_B = \boldsymbol{g}_B(\boldsymbol{x}_A, \boldsymbol{e}_A)\,.$$

Let $C := A \setminus (\mathrm{pa}(B) \setminus B)$, then by Lemma F.11.(7) we have that there exists $\hat{\boldsymbol{e}}_C \in \boldsymbol{\mathcal{E}}_C$ and $\hat{\boldsymbol{x}}_C \in \boldsymbol{\mathcal{X}}_C$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}_{\mathcal{J}\setminus C}}$-almost every $\boldsymbol{e}_{\mathcal{J}\setminus C} \in \boldsymbol{\mathcal{E}}_{\mathcal{J}\setminus C}$ and for all $\boldsymbol{x}_{\mathcal{I}\setminus C} \in \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus C}$

$$\boldsymbol{x}_B = \boldsymbol{f}_B(\boldsymbol{x}_{\mathrm{pa}(B)}, \boldsymbol{e}_{\mathrm{pa}(B)}) \iff \boldsymbol{x}_B = \boldsymbol{g}_B(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \hat{\boldsymbol{x}}_C, \boldsymbol{e}_{\mathrm{pa}(B)}, \hat{\boldsymbol{e}}_C)\,.$$

Defining the mapping $\boldsymbol{h}_B : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(B)\setminus B} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(B)} \to \boldsymbol{\mathcal{X}}_B$ by

$$\boldsymbol{h}_B(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)}) := \boldsymbol{g}_B(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \hat{\boldsymbol{x}}_C, \boldsymbol{e}_{\mathrm{pa}(B)}, \hat{\boldsymbol{e}}_C)\,,$$

where we picked $\hat{\boldsymbol{e}}_C \in \boldsymbol{\mathcal{E}}_C$ and $\hat{\boldsymbol{x}}_C \in \boldsymbol{\mathcal{X}}_C$ such that the above equivalence holds, and applying Lemma F.11.(6) we get that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_B = \boldsymbol{f}_B(\boldsymbol{x}_{\mathrm{pa}(B)}, \boldsymbol{e}_{\mathrm{pa}(B)}) \iff \boldsymbol{x}_B = \boldsymbol{h}_B(\boldsymbol{x}_{\mathrm{pa}(B)\setminus B}, \boldsymbol{e}_{\mathrm{pa}(B)})$$

holds. Thus, $\mathcal{M}$ is uniquely solvable w.r.t. B. □

PROOF OF PROPOSITION 5.4. From unique solvability of $\mathcal{M}$ w.r.t. $\mathcal{L}_1$ it follows that there exists a mapping $\boldsymbol{g}_{\mathcal{L}_1} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{L}_1)\setminus(\mathcal{L}_1)} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{L}_1)} \to \boldsymbol{\mathcal{X}}_{\mathcal{L}_1}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{L}_1} = \boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1)\setminus\mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}) \quad \Longleftrightarrow \quad \boldsymbol{x}_{\mathcal{L}_1} = \boldsymbol{f}_{\mathcal{L}_1}(\boldsymbol{x}, \boldsymbol{e}) \,.$$

Let $\widehat{\mathrm{pa}}$ denotes the parents in $\mathcal{G}^a(\mathcal{M}_{\mathrm{marg}(\mathcal{L}_1)})$. Note that $\widehat{\mathrm{pa}}(\mathcal{L}_2) \setminus \mathcal{L}_2 \subseteq \mathrm{pa}(\mathcal{L}_1 \cup \mathcal{L}_2) \setminus (\mathcal{L}_1 \cup \mathcal{L}_2)$. Let $\tilde{\boldsymbol{f}}$ denote the marginal causal mechanism of a structurally minimal SCM that is equivalent to the marginalization $\mathcal{M}_{\mathrm{marg}(\mathcal{L}_1)}$ constructed from $\boldsymbol{g}_{\mathcal{L}_1}$ (see Proposition 2.11).

$\Longrightarrow$: If $\mathcal{M}_{\mathrm{marg}(\mathcal{L}_1)}$ is uniquely solvable w.r.t. $\mathcal{L}_2$, then there exists a mapping $\tilde{\boldsymbol{g}}_{\mathcal{L}_2} : \boldsymbol{\mathcal{X}}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)\setminus\mathcal{L}_2} \times \boldsymbol{\mathcal{E}}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)} \to \boldsymbol{\mathcal{X}}_{\mathcal{L}_2}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}_{\mathcal{I}\setminus\mathcal{L}_1} \in \boldsymbol{\mathcal{X}}_{\mathcal{I}\setminus\mathcal{L}_1}$

$$\boldsymbol{x}_{\mathcal{L}_2} = \tilde{\boldsymbol{g}}_{\mathcal{L}_2}(\boldsymbol{x}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)\setminus\mathcal{L}_2}, \boldsymbol{e}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)}) \iff \boldsymbol{x}_{\mathcal{L}_2} = \boldsymbol{f}_{\mathcal{L}_2}(\boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1)\setminus\mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}), \boldsymbol{x}_{\mathcal{I}\setminus\mathcal{L}_1}, \boldsymbol{e}) \,.$$

Define the mapping $\boldsymbol{h} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)\setminus(\mathcal{L}_1\cup\mathcal{L}_2)} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)} \to \boldsymbol{\mathcal{X}}_{\mathcal{L}_1\cup\mathcal{L}_2}$ by

$$(\boldsymbol{h}_{\mathcal{L}_1}, \boldsymbol{h}_{\mathcal{L}_2})(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)\setminus(\mathcal{L}_1\cup\mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)}) :=$$
$$\Big(\boldsymbol{g}_{\mathcal{L}_1}\big((\tilde{\boldsymbol{g}}_{\mathcal{L}_2})_{\mathrm{pa}(\mathcal{L}_1)}(\boldsymbol{x}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)\setminus\mathcal{L}_2}, \boldsymbol{e}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)}), \boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1)\setminus(\mathcal{L}_1\cup\mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}\big), \tilde{\boldsymbol{g}}_{\mathcal{L}_2}(\boldsymbol{x}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)\setminus\mathcal{L}_2}, \boldsymbol{e}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)})\Big) \,.$$

Then for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\begin{aligned}
&\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{f}_{\mathcal{L}_1}(\boldsymbol{x}, \boldsymbol{e}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \boldsymbol{f}_{\mathcal{L}_2}(\boldsymbol{x}, \boldsymbol{e}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1)\setminus\mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \boldsymbol{f}_{\mathcal{L}_2}(\boldsymbol{x}, \boldsymbol{e}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1)\setminus\mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \boldsymbol{f}_{\mathcal{L}_2}(\boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1)\setminus\mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}), \boldsymbol{x}_{\mathcal{I}\setminus\mathcal{L}_1}, \boldsymbol{e}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1)\setminus\mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \tilde{\boldsymbol{g}}_{\mathcal{L}_2}(\boldsymbol{x}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)\setminus\mathcal{L}_2}, \boldsymbol{e}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{g}_{\mathcal{L}_1}\big((\tilde{\boldsymbol{g}}_{\mathcal{L}_2})_{\mathrm{pa}(\mathcal{L}_1)}(\boldsymbol{x}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)\setminus\mathcal{L}_2}, \boldsymbol{e}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)}), \boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1)\setminus(\mathcal{L}_1\cup\mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}\big) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \tilde{\boldsymbol{g}}_{\mathcal{L}_2}(\boldsymbol{x}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)\setminus\mathcal{L}_2}, \boldsymbol{e}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{h}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)\setminus(\mathcal{L}_1\cup\mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \boldsymbol{h}_{\mathcal{L}_2}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)\setminus(\mathcal{L}_1\cup\mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)}) \,, \end{cases}
\end{aligned}$$

where in the first equivalence we used unique solvability w.r.t. $\mathcal{L}_1$ of $\mathcal{M}$, in the second we used substitution, in the third we used unique solvability w.r.t. $\mathcal{L}_2$ of $\mathcal{M}_{\mathrm{marg}(\mathcal{L}_1)}$, in the fourth we used again substitution and in the last equivalence we used the definition of $\boldsymbol{h}$. From this we conclude that $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}_1 \cup \mathcal{L}_2$. Hence, by definition it follows that $\mathrm{marg}(\mathcal{L}_2) \circ \mathrm{marg}(\mathcal{L}_1)(\mathcal{M}) = \mathrm{marg}(\mathcal{L}_1 \cup \mathcal{L}_2)(\mathcal{M})$.

$\Longleftarrow$: If $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}_1 \cup \mathcal{L}_2$, then there exists a mapping $\boldsymbol{h} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)\setminus(\mathcal{L}_1\cup\mathcal{L}_2)} \times \boldsymbol{\mathcal{E}}_{\mathcal{L}_1\cup\mathcal{L}_2} \to \boldsymbol{\mathcal{X}}_{\mathcal{L}_1\cup\mathcal{L}_2}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{L}_1\cup\mathcal{L}_2} = \boldsymbol{h}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)\setminus(\mathcal{L}_1\cup\mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1\cup\mathcal{L}_2)}) \quad \Longleftrightarrow \quad \boldsymbol{x}_{\mathcal{L}_1\cup\mathcal{L}_2} = \boldsymbol{f}_{\mathcal{L}_1\cup\mathcal{L}_2}(\boldsymbol{x}, \boldsymbol{e}) \,.$$

Then, for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$
\begin{aligned}
&\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{h}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1 \cup \mathcal{L}_2) \setminus (\mathcal{L}_1 \cup \mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1 \cup \mathcal{L}_2)}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \boldsymbol{h}_{\mathcal{L}_2}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1 \cup \mathcal{L}_2) \setminus (\mathcal{L}_1 \cup \mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1 \cup \mathcal{L}_2)}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{f}_{\mathcal{L}_1}(\boldsymbol{x}, \boldsymbol{e}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \boldsymbol{f}_{\mathcal{L}_2}(\boldsymbol{x}, \boldsymbol{e}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1) \setminus \mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \boldsymbol{f}_{\mathcal{L}_2}(\boldsymbol{x}, \boldsymbol{e}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1) \setminus \mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \boldsymbol{f}_{\mathcal{L}_2}(\boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1) \setminus \mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}), \boldsymbol{x}_{\mathcal{I} \setminus \mathcal{L}_1}, \boldsymbol{e}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}_1} &= \boldsymbol{g}_{\mathcal{L}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1) \setminus \mathcal{L}_1}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1)}) \\ \boldsymbol{x}_{\mathcal{L}_2} &= \tilde{\boldsymbol{f}}_{\mathcal{L}_2}(\boldsymbol{x}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)}, \boldsymbol{e}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)})\,. \end{cases}
\end{aligned}
$$

This gives for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ for all $\boldsymbol{x}_{\mathcal{I} \setminus \mathcal{L}_1} \in \boldsymbol{\mathcal{X}}_{\mathcal{I} \setminus \mathcal{L}_1}$

$$
\begin{aligned}
&\boldsymbol{x}_{\mathcal{L}_2} = \boldsymbol{h}_{\mathcal{L}_2}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}_1 \cup \mathcal{L}_2) \setminus (\mathcal{L}_1 \cup \mathcal{L}_2)}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L}_1 \cup \mathcal{L}_2)}) \\
\iff &\boldsymbol{x}_{\mathcal{L}_2} = \tilde{\boldsymbol{f}}_{\mathcal{L}_2}(\boldsymbol{x}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)}, \boldsymbol{e}_{\widehat{\mathrm{pa}}(\mathcal{L}_2)})\,.
\end{aligned}
$$

Now apply Lemma E.3 to conclude that $\mathcal{M}_{\mathrm{marg}(\mathcal{L}_1)}$ is uniquely solvable w.r.t. $\mathcal{L}_2$. □

PROOF OF PROPOSITION 5.5. The commutation relation with the perfect intervention follows straightforwardly from the definitions of perfect intervention and marginalization and the fact that if $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$, then $\mathcal{M}_{\mathrm{do}(I, \boldsymbol{\xi}_I)}$ is also uniquely solvable w.r.t. $\mathcal{L}$, since the structural equations for the variables $\mathcal{L}$ are the same for $\mathcal{M}$ and $\mathcal{M}_{\mathrm{do}(I, \boldsymbol{\xi}_I)}$.

The commutation relation with the twin operation follows straightforwardly from the definition of the twin operation and marginalization and the fact that if $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$, then $\mathrm{twin}(\mathcal{M})$ is uniquely solvable w.r.t. $\mathcal{L} \cup \mathcal{L}'$, where $\mathcal{L}'$ is the copy of $\mathcal{L}$ in $\mathcal{I}'$. □

LEMMA E.4. *Given an SCM $\mathcal{M}$ and a subset $\mathcal{L} \subseteq \mathcal{I}$ such that $\mathcal{M}$ is uniquely solvable w.r.t. $\mathcal{L}$. Then $\mathcal{M}$ and $\mathrm{marg}(\mathcal{L})(\mathcal{M})$ are observationally equivalent w.r.t. $\mathcal{I} \setminus \mathcal{L}$.*

PROOF. Let $\mathcal{O} := \mathcal{I} \setminus \mathcal{L}$. From unique solvability w.r.t. $\mathcal{L}$ it follows that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$
\begin{aligned}
&\begin{cases} \boldsymbol{x}_{\mathcal{L}} &= \boldsymbol{f}_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{e}) \\ \boldsymbol{x}_{\mathcal{O}} &= \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{x}, \boldsymbol{e}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}} &= \boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}) \\ \boldsymbol{x}_{\mathcal{O}} &= \boldsymbol{f}_{\mathcal{O}}(\boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}), \boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e}) \end{cases} \\
\iff &\begin{cases} \boldsymbol{x}_{\mathcal{L}} &= \boldsymbol{g}_{\mathcal{L}}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}) \\ \boldsymbol{x}_{\mathcal{O}} &= \tilde{\boldsymbol{f}}(\boldsymbol{x}_{\mathcal{O}}, \boldsymbol{e})\,, \end{cases}
\end{aligned}
$$

where $\tilde{\boldsymbol{f}}$ is the marginal causal mechanism of $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ constructed from a measurable solution function $\boldsymbol{g}_{\mathcal{L}} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{L})} \to \boldsymbol{\mathcal{X}}_{\mathcal{L}}$ for $\mathcal{M}$ w.r.t. $\mathcal{L}$. Hence, a solution $(\boldsymbol{X}, \boldsymbol{E})$ of $\mathcal{M}$ satisfies $\boldsymbol{X}_{\mathcal{O}} = \tilde{\boldsymbol{f}}(\boldsymbol{X}_{\mathcal{O}}, \boldsymbol{E})$ a.s.. Conversely, if $(\tilde{\boldsymbol{X}}_{\mathcal{O}}, \boldsymbol{E})$ is a solution of the marginal SCM $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ then with $\tilde{\boldsymbol{X}}_{\mathcal{L}} := \boldsymbol{g}_{\mathcal{L}}(\tilde{\boldsymbol{X}}_{\mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}}, \boldsymbol{E}_{\mathrm{pa}(\mathcal{L})})$, the random variables $(\boldsymbol{X}, \boldsymbol{E}) := (\tilde{\boldsymbol{X}}_{\mathcal{O}}, \tilde{\boldsymbol{X}}_{\mathcal{L}}, \boldsymbol{E})$ are a solution of $\mathcal{M}$. □

PROOF OF THEOREM 5.6. The observational equivalence follows from Lemma E.4. Using both Lemma E.4 and Proposition 5.5 we can prove the interventional equivalence. Observe that from Proposition 5.5 we know that for a subset $I \subseteq \mathcal{I} \setminus \mathcal{L}$ and a value $\boldsymbol{\xi}_I \in \boldsymbol{\mathcal{X}}_I$, $(\text{marg}(\mathcal{L}) \circ \text{do}(I, \boldsymbol{\xi}_I))(\mathcal{M})$ exists. By Lemma E.4 we know that $\text{do}(I, \boldsymbol{\xi}_I)(\mathcal{M})$ and $(\text{marg}(\mathcal{L}) \circ \text{do}(I, \boldsymbol{\xi}_I))(\mathcal{M})$ are observationally equivalent w.r.t. $\mathcal{O}$ and hence by applying again Proposition 5.5, $\text{do}(I, \boldsymbol{\xi}_I)(\mathcal{M})$ and $(\text{do}(I, \boldsymbol{\xi}) \circ \text{marg}(\mathcal{L}))(\mathcal{M})$ are observationally equivalent w.r.t. $\mathcal{O}$. This implies that $\mathcal{M}$ and $\text{marg}(\mathcal{L})(\mathcal{M})$ are interventionally equivalent w.r.t. $\mathcal{O}$. Lastly, we need to show that $\text{twin}(\mathcal{M})$ and $(\text{twin} \circ \text{marg}(\mathcal{L}))(\mathcal{M})$ are interventionally equivalent w.r.t. $(\mathcal{I} \cup \mathcal{I}') \setminus (\mathcal{L} \cup \mathcal{L}')$, where $\mathcal{L}'$ is the copy of $\mathcal{L}$ in $\mathcal{I}'$. From Proposition 5.5 $(\text{twin} \circ \text{marg}(\mathcal{L}))(\mathcal{M})$ is equivalent to $(\text{marg}(\mathcal{L} \cup \mathcal{L}') \circ \text{twin})(\mathcal{M})$ and since we proved that $(\text{marg}(\mathcal{L} \cup \mathcal{L}') \circ \text{twin})(\mathcal{M})$ and $\text{twin}(\mathcal{M})$ are interventionally equivalent w.r.t. $(\mathcal{I} \cup \mathcal{I}') \setminus (\mathcal{L} \cup \mathcal{L}')$ the result follows. □

PROOF OF PROPOSITION 5.8. A similar proof as for Theorem 1 in [7] works. □

PROOF OF PROPOSITION 5.9. First we prove the commutation relation of the perfect intervention. Observe that applying the $\text{do}(I)$ operation to the latent projection $\text{marg}(\mathcal{L})(\mathcal{G})$ removes all the incoming edges on the nodes I. Such an incoming edge at a node in I in $\text{marg}(\mathcal{L})(\mathcal{G})$ corresponds to a path in $\mathcal{G}$ that points to that node. But since $\text{do}(I)(\mathcal{G})$ is just $\mathcal{G}$ with all the incoming edges on I removed, the graph $(\text{marg}(\mathcal{L}) \circ \text{do}(I))(\mathcal{G})$ also has all the incoming edges on the nodes I removed.

Next, we will prove the commutation relation of the twin operation. We will denote the copy in $\mathcal{I}'$ of any node $i \in \mathcal{I}$ by i', that is, $\mathcal{I}' = \{i' : i \in \mathcal{I}\}$. The edges in $(\text{twin}(\mathcal{I} \setminus \mathcal{L}) \circ \text{marg}(\mathcal{L}))(\mathcal{G})$ can be partitioned into three cases:

$$\begin{cases} v \to w & v \in \mathcal{J} \cup \mathcal{I} \setminus \mathcal{L}, w \in \mathcal{J} \cup \mathcal{I} \setminus \mathcal{L}, v \to w \in \text{marg}(\mathcal{L})(\mathcal{G})\,, \\ v \to w' & v \in \mathcal{J}, w \in \mathcal{I} \setminus \mathcal{L}, v \to w \in \text{marg}(\mathcal{L})(\mathcal{G})\,, \\ v' \to w' & v \in \mathcal{I} \setminus \mathcal{L}, w \in \mathcal{I} \setminus \mathcal{L}, v \to w \in \text{marg}(\mathcal{L})(\mathcal{G})\,, \end{cases}$$

where $\mathcal{J} := \mathcal{V} \setminus \mathcal{I}$.

Note that in $\text{twin}(\mathcal{I})(\mathcal{G})$, there are no directed edges of the form $v' \to w$ by definition. Therefore, the edges in $(\text{marg}(\mathcal{L} \cup \mathcal{L}') \circ \text{twin}(\mathcal{I}))(\mathcal{G})$ can be partitioned into three cases:

$$\begin{cases} v \to w & v \in \mathcal{J} \cup \mathcal{I} \setminus \mathcal{L}, w \in \mathcal{J} \cup \mathcal{I} \setminus \mathcal{L}, v \to \ell_1 \to \cdots \to \ell_n \to w \in \text{twin}(\mathcal{I})(\mathcal{G})\,, \\ v \to w' & v \in \mathcal{J}, w \in \mathcal{I} \setminus \mathcal{L}, v \to \ell_1' \to \cdots \to \ell_n' \to w' \in \text{twin}(\mathcal{I})(\mathcal{G})\,, \\ v' \to w' & v \in \mathcal{I} \setminus \mathcal{L}, w \in \mathcal{I} \setminus \mathcal{L}, v' \to \ell_1' \to \cdots \to \ell_n' \to w' \in \text{twin}(\mathcal{I})(\mathcal{G})\,, \end{cases}$$

where all $\ell_1, \ldots, \ell_n \in \mathcal{L}$ and $\ell_1', \ldots, \ell_n' \in \mathcal{L}'$. Thus, the non-endpoint nodes on the directed paths in $\text{twin}(\mathcal{I})(\mathcal{G})$ must either all lie in $\mathcal{L}$ or in $\mathcal{L}'$. With the definition of $\text{twin}(\mathcal{I})(\mathcal{G})$ we can rewrite this as follows:

$$\begin{cases} v \to w & v \in \mathcal{J} \cup \mathcal{I} \setminus \mathcal{L}, w \in \mathcal{J} \cup \mathcal{I} \setminus \mathcal{L}, v \to \ell_1 \to \cdots \to \ell_n \to w \in \mathcal{G}\,, \\ v \to w' & v \in \mathcal{J}, w \in \mathcal{I} \setminus \mathcal{L}, v \to \ell_1 \to \cdots \to \ell_n \to w \in \mathcal{G}\,, \\ v' \to w' & v \in \mathcal{I} \setminus \mathcal{L}, w \in \mathcal{I} \setminus \mathcal{L}, v \to \ell_1 \to \cdots \to \ell_n \to w \in \mathcal{G}\,, \end{cases}$$

where all intermediate $\ell_1, \ldots, \ell_n$ must lie in $\mathcal{L}$. This corresponds exactly with the edges in $(\text{twin}(\mathcal{I} \setminus \mathcal{L}) \circ \text{marg}(\mathcal{L}))(\mathcal{G})$.

□

PROOF OF PROPOSITION 5.11. Without loss of generality, we assume that $\mathcal{M}$ is structurally minimal (see Proposition 2.11). Let $\boldsymbol{g}_{\mathcal{L}}$ be a measurable solution function for $\mathcal{M}$ w.r.t.

$\mathcal{L}$ and denote by $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ the marginal SCM constructed from $\boldsymbol{g}_{\mathcal{L}}$. For $j \in \mathcal{I} \setminus \mathcal{L}$, define $A_j := \mathrm{an}_{\mathcal{G}(\mathcal{M})_{\mathcal{L}}}(\mathrm{pa}(j) \cap \mathcal{L}) \subseteq \mathcal{L}$ and let $\tilde{\boldsymbol{g}}_{A_j}$ be a measurable solution function for $\mathcal{M}$ w.r.t. A_j. Because $A_j \subseteq \mathcal{L}$ and $\mathrm{pa}(A_j) \setminus A_j \subseteq \mathrm{pa}(\mathcal{L}) \setminus \mathcal{L}$, by Lemma E.1, for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$(\boldsymbol{g}_{\mathcal{L}})_{A_j}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L})\setminus\mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}) = \tilde{\boldsymbol{g}}_{A_j}(\boldsymbol{x}_{\mathrm{pa}(A_j)\setminus A_j}, \boldsymbol{e}_{\mathrm{pa}(A_j)}) \,.$$

Therefore, the component $\tilde{f}_j$ of the marginal causal mechanism $\tilde{\boldsymbol{f}}$ of $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ satisfies for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\begin{aligned}\tilde{f}_j(\boldsymbol{x}_{\mathcal{I}\setminus\mathcal{L}}, \boldsymbol{e}) &:= f_j\big((\boldsymbol{g}_{\mathcal{L}})_{\mathrm{pa}(j)}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{L})\setminus\mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(\mathcal{L})}), \boldsymbol{x}_{\mathrm{pa}(j)\setminus\mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(j)}\big) \\ &= f_j\big((\tilde{\boldsymbol{g}}_{A_j})_{\mathrm{pa}(j)\cap\mathcal{L}}(\boldsymbol{x}_{\mathrm{pa}(A_j)\setminus A_j}, \boldsymbol{e}_{\mathrm{pa}(A_j)}), \boldsymbol{x}_{\mathrm{pa}(j)\setminus\mathcal{L}}, \boldsymbol{e}_{\mathrm{pa}(j)}\big) \,.\end{aligned}$$

Hence, the endogenous parents of j in $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ are a subset of $\big((\mathrm{pa}(A_j) \setminus A_j) \cup (\mathrm{pa}(j) \setminus \mathcal{L})\big) \cap \mathcal{I}$ and the exogenous parents of j in $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ are a subset of $(\mathrm{pa}(A_j) \cup \mathrm{pa}(j)) \cap \mathcal{J}$. Hence, all parents of j in $\mathcal{M}_{\mathrm{marg}(\mathcal{L})}$ are a subset of those $k \in (\mathcal{I} \setminus \mathcal{L}) \cup \mathcal{J}$ such that there exists a path $k \to \ell_1 \to \cdots \to \ell_n \to j \in \mathcal{G}^a(\mathcal{M})$ for $n \geq 0$ and $\ell_1, \ldots, \ell_n \in \mathcal{L}$. Therefore, the augmented graph $\mathcal{G}^a\big(\mathrm{marg}(\mathcal{L})(\mathcal{M})\big)$ is a subgraph of the latent projection $\mathrm{marg}(\mathcal{L})\big(\mathcal{G}^a(\mathcal{M})\big)$. Hence,

$$\begin{aligned}\mathcal{G}\big(\mathrm{marg}(\mathcal{L})(\mathcal{M})\big) &= \mathrm{marg}(\mathcal{J})\Big(\mathcal{G}^a\big(\mathrm{marg}(\mathcal{L})(\mathcal{M})\big)\Big) \\ &\subseteq \mathrm{marg}(\mathcal{J})\Big(\mathrm{marg}(\mathcal{L})\big(\mathcal{G}^a(\mathcal{M})\big)\Big) \\ &= \mathrm{marg}(\mathcal{L})\Big(\mathrm{marg}(\mathcal{J})\big(\mathcal{G}^a(\mathcal{M})\big)\Big) \\ &= \mathrm{marg}(\mathcal{L})\big(\mathcal{G}(\mathcal{M})\big)\end{aligned}$$

and we conclude that also the graph $\mathcal{G}\big(\mathrm{marg}(\mathcal{L})(\mathcal{M})\big)$ is a subgraph of the latent projection $\mathrm{marg}(\mathcal{L})\big(\mathcal{G}(\mathcal{M})\big)$. □

Section 6.

PROOF OF THEOREM 6.3. This follows directly from Theorems A.7 and A.21. □

Section 7.

PROOF OF PROPOSITION 7.1. We define $\tilde{\mathcal{M}} := \mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$, $\widetilde{\mathrm{pa}} := \mathrm{pa}_{\mathcal{G}^a(\tilde{\mathcal{M}})}$ and $\mathcal{A} := \mathrm{an}_{\mathcal{G}(\tilde{\mathcal{M}})_{\setminus i}}(j)$. Suppose that $i \to j \notin \mathrm{marg}(\mathcal{I} \setminus \mathcal{O})(\mathcal{G}(\mathcal{M}))$ and assume that the two induced distributions do not coincide. Because $i \to j \notin \mathrm{marg}(\mathcal{I} \setminus \mathcal{O})(\mathcal{G}(\mathcal{M}))$ it follows that $(\widetilde{\mathrm{pa}}(\mathcal{A}) \setminus \mathcal{A}) \cap \mathcal{I} = \emptyset$. Let now $\tilde{\boldsymbol{g}}_{\mathcal{A}} : \boldsymbol{\mathcal{E}}_{\widetilde{\mathrm{pa}}(\mathcal{A})} \to \boldsymbol{\mathcal{X}}_{\mathcal{A}}$ be a measurable solution function for $\tilde{\mathcal{M}}$ w.r.t. $\mathcal{A}$, that is, we have for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{A}} = \tilde{\boldsymbol{f}}_{\mathcal{A}}(\boldsymbol{x}, \boldsymbol{e}) \quad \Longleftrightarrow \quad \boldsymbol{x}_{\mathcal{A}} = \tilde{\boldsymbol{g}}_{\mathcal{A}}(\boldsymbol{e}_{\widetilde{\mathrm{pa}}(\mathcal{A})}) \,,$$

where $\tilde{\boldsymbol{f}}$ is the ausal mechanism of $\tilde{\mathcal{M}}$. Because $i \notin \mathcal{A}$ and $j \in \mathcal{A}$, it follows that for the intervened model $(\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)})_{\mathrm{do}(\{i\},\xi_i)}$ the marginal solution X_j is also a marginal solution of $(\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)})_{\mathrm{do}(\{i\},\tilde{\xi}_i)}$ and vice versa, which is in contradiction with the assumption. □

PROOF OF PROPOSITION 7.2. Let's define $\tilde{\mathcal{M}} := \mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$, $\widetilde{\mathrm{pa}} := \mathrm{pa}_{\mathcal{G}^a(\tilde{\mathcal{M}})}$, $\mathcal{A}_i := \mathrm{an}_{\mathcal{G}(\tilde{\mathcal{M}})}(i)$ and $\mathcal{A}_j^{\setminus i} := \mathrm{an}_{\mathcal{G}(\tilde{\mathcal{M}})_{\setminus i}}(j)$. Suppose that there does not exist a bidirected edge $i \leftrightarrow j$

in the latent projection $\mathrm{marg}(\mathcal{I}\setminus\mathcal{O})(\mathcal{G}(\mathcal{M}))$. Because $i\leftrightarrow j\notin \mathrm{marg}(\mathcal{I}\setminus\mathcal{O})(\mathcal{G}(\tilde{\mathcal{M}}))$, where here $\tilde{\mathcal{M}}$ is the intervened model $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$, we have that $\mathrm{an}_{\mathcal{G}^a(\tilde{\mathcal{M}})_{\setminus j}}(i)\cap \mathrm{an}_{\mathcal{G}^a(\tilde{\mathcal{M}})_{\setminus i}}(j)\cap\mathcal{J}=\emptyset$. From $j\notin \mathrm{an}_{\mathcal{G}(\tilde{\mathcal{M}})}(i)$ it follows that $\mathrm{an}_{\mathcal{G}(\tilde{\mathcal{M}})_{\setminus j}}(i)=\mathrm{an}_{\mathcal{G}(\tilde{\mathcal{M}})}(i)$, and hence $\mathrm{an}_{\mathcal{G}^a(\tilde{\mathcal{M}})}(i)\cap \mathrm{an}_{\mathcal{G}^a(\tilde{\mathcal{M}})_{\setminus i}}(j)\cap\mathcal{J}=\emptyset$. Observe that $\widetilde{\mathrm{pa}}(\mathcal{A}_i)\subseteq \mathrm{an}_{\mathcal{G}^a(\tilde{\mathcal{M}})}(i)$ and $\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})\subseteq \mathrm{an}_{\mathcal{G}^a(\tilde{\mathcal{M}})_{\setminus i}}(j)\cup\{i\}$, and thus $\widetilde{\mathrm{pa}}(\mathcal{A}_i)\cap\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})\cap\mathcal{J}=\emptyset$. Let $\boldsymbol{g}_{\mathcal{A}_i}:\boldsymbol{\mathcal{E}}_{\widetilde{\mathrm{pa}}(\mathcal{A}_i)}\to\boldsymbol{\mathcal{X}}_{\mathcal{A}_i}$ be a measurable solution function for $\tilde{\mathcal{M}}$ w.r.t. $\mathcal{A}_i$, that is, we have for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}\in\boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}\in\boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{A}_i}=\tilde{\boldsymbol{f}}_{\mathcal{A}_i}(\boldsymbol{x},\boldsymbol{e})\quad\iff\quad \boldsymbol{x}_{\mathcal{A}_i}=\boldsymbol{g}_{\mathcal{A}_i}(\boldsymbol{e}_{\widetilde{\mathrm{pa}}(\mathcal{A}_i)})\,,$$

where $\tilde{\boldsymbol{f}}$ is the intervened causal mechanism of $\tilde{\mathcal{M}}$. Because $\widetilde{\mathrm{pa}}(\mathcal{A}_i)\cap\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})\cap\mathcal{J}=\emptyset$ and $i\in\mathcal{A}_i$, we have that $X_i\perp\!\!\!\perp \boldsymbol{E}_{\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})}$ for every solution $(\boldsymbol{X},\boldsymbol{E})$ of $\tilde{\mathcal{M}}$.

Assume for the moment that $i\in\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})\setminus\mathcal{A}_j^{\setminus i}$, then $(\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})\setminus\mathcal{A}_j^{\setminus i})\cap\mathcal{I}=\{i\}$. Let $\boldsymbol{g}_{\mathcal{A}_j^{\setminus i}}:\mathcal{X}_i\times\boldsymbol{\mathcal{E}}_{\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})}\to\boldsymbol{\mathcal{X}}_{\mathcal{A}_j^{\setminus i}}$ be a measurable solution function for $\tilde{\mathcal{M}}$ w.r.t. $\mathcal{A}_j^{\setminus i}$, that is, we have for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e}\in\boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x}\in\boldsymbol{\mathcal{X}}$

$$\boldsymbol{x}_{\mathcal{A}_j^{\setminus i}}=\tilde{\boldsymbol{f}}_{\mathcal{A}_j^{\setminus i}}(\boldsymbol{x},\boldsymbol{e})\iff \boldsymbol{x}_{\mathcal{A}_j^{\setminus i}}=\boldsymbol{g}_{\mathcal{A}_j^{\setminus i}}(x_i,\boldsymbol{e}_{\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})})\,.$$

For every measurable set $\mathcal{B}_j\subseteq\mathcal{X}_j$ there exists a version of the regular conditional probability $\mathbb{P}_{\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}}(X_j\in\mathcal{B}\,|\,X_i=\xi_i)$ such that for every value $\xi_i\in\mathcal{X}_i$ it satisfies

$$\begin{aligned}
\mathbb{P}_{\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}}\big(X_j\in\mathcal{B}_j\,|\,X_i=\xi_i\big)&=\mathbb{P}_{\tilde{\mathcal{M}}}\big(X_j\in\mathcal{B}_j\,|\,X_i=\xi_i\big)\\
&=\mathbb{P}_{\tilde{\mathcal{M}}}\big((\boldsymbol{g}_{\mathcal{A}_j^{\setminus i}})_j(X_i,\boldsymbol{E}_{\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})})\in\mathcal{B}_j\,|\,X_i=\xi_i\big)\\
&=\mathbb{P}_{\tilde{\mathcal{M}}}\big((\boldsymbol{g}_{\mathcal{A}_j^{\setminus i}})_j(\xi_i,\boldsymbol{E}_{\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})})\in\mathcal{B}_j\,|\,X_i=\xi_i\big)\\
&=\mathbb{P}_{\tilde{\mathcal{M}}}\big((\boldsymbol{g}_{\mathcal{A}_j^{\setminus i}})_j(\xi_i,\boldsymbol{E}_{\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})})\in\mathcal{B}_j\big)\\
&=\mathbb{P}_{\tilde{\mathcal{M}}_{\mathrm{do}(\{i\},\xi_i)}}\big((\boldsymbol{g}_{\mathcal{A}_j^{\setminus i}})_j(X_i,\boldsymbol{E}_{\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})})\in\mathcal{B}_j\big)\\
&=\mathbb{P}_{\tilde{\mathcal{M}}_{\mathrm{do}(\{i\},\xi_i)}}\big(X_j\in\mathcal{B}_j\big)\\
&=\mathbb{P}_{\big(\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}\big)_{\mathrm{do}(\{i\},\xi_i)}}\big(X_j\in\mathcal{B}_j\big)\,,
\end{aligned}$$

where we used $X_i\perp\!\!\!\perp \boldsymbol{E}_{\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})}$ in the fourth equality.

If we assume $i\notin\widetilde{\mathrm{pa}}(\mathcal{A}_j^{\setminus i})\setminus\mathcal{A}_j^{\setminus i}$ instead of $i\in\mathrm{pa}(\mathcal{A}_j^{\setminus i})\setminus\mathcal{A}_j^{\setminus i}$, then we similarly arrive at the same conclusion. □

Section 8.

PROOF OF PROPOSITION 8.2. We first show that the class of simple SCMs is closed under marginalization. Take two disjoint subsets $\mathcal{L}_1$ and $\mathcal{L}_2$ in $\mathcal{I}$. Then, it suffices to show that $\mathcal{M}_{\mathrm{marg}(\mathcal{L}_1)}$ is uniquely solvable w.r.t. $\mathcal{L}_2$. This follows directly from Proposition 5.4.

To show that the class of simple SCMs is closed under perfect intervention. Let $\mathcal{M}$ be a simple SCM, $\mathcal{O}\subseteq\mathcal{I}$, $I\subseteq\mathcal{I}$ and $\boldsymbol{\xi}_I\in\boldsymbol{\mathcal{X}}_I$. Define $\mathcal{O}_1:=\mathcal{O}\cap I$ and $\mathcal{O}_2:=\mathcal{O}\setminus I$, then $\mathcal{O}=\mathcal{O}_1\cup\mathcal{O}_2$. Note that $\mathrm{pa}(\mathcal{O}_2)\setminus\mathcal{O}_2=(\mathrm{pa}(\mathcal{O}_2)\setminus(\mathcal{O}_2\cup I))\cup(\mathrm{pa}(\mathcal{O}_2)\cap I)$ and $\mathrm{pa}(\mathcal{O}_2)\setminus(\mathcal{O}_2\cup I)\subseteq\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}$. Let $\boldsymbol{g}_{\mathcal{O}_2}:\boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O}_2)\setminus\mathcal{O}_2}\times\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O}_2)}\to\boldsymbol{\mathcal{X}}_{\mathcal{O}_2}$ be a measurable solution function for $\mathcal{M}$ w.r.t. $\mathcal{O}_2$. The mapping $\tilde{\boldsymbol{g}}_{\mathcal{O}}:\boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}}\times\boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O})}\to\boldsymbol{\mathcal{X}}_{\mathcal{O}}$ defined by

$$\begin{cases}(\tilde{\boldsymbol{g}}_{\mathcal{O}})_{\mathcal{O}_1}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}):=\boldsymbol{\xi}_{\mathcal{O}_1}\\
(\tilde{\boldsymbol{g}}_{\mathcal{O}})_{\mathcal{O}_2}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O})\setminus\mathcal{O}},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O})}):=\boldsymbol{g}_{\mathcal{O}_2}(\boldsymbol{x}_{\mathrm{pa}(\mathcal{O}_2)\setminus(\mathcal{O}_2\cup I)},\boldsymbol{\xi}_{\mathrm{pa}(\mathcal{O}_2)\cap I},\boldsymbol{e}_{\mathrm{pa}(\mathcal{O}_2)})\end{cases}$$

is a measurable solution function for $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ w.r.t. $\mathcal{O}$, and it is clear that $\mathcal{M}_{\mathrm{do}(I,\boldsymbol{\xi}_I)}$ is uniquely solvable w.r.t. $\mathcal{O}$.

Next, we show that the class of simple SCMs is closed under the twin operation. Let $\tilde{\mathcal{O}} \subseteq \mathcal{I} \cup \mathcal{I}'$. Take $\mathcal{O}_1 = \tilde{\mathcal{O}} \cap \mathcal{I}$, $\mathcal{O}_2' = \tilde{\mathcal{O}} \cap \mathcal{I}'$ and $\mathcal{O}_2$ the original copy of $\mathcal{O}_2'$ in $\mathcal{I}$. Let $\boldsymbol{g}_{\mathcal{O}_1} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O}_1)\setminus\mathcal{O}_1} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O}_1)} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}_1}$ and $\boldsymbol{g}_{\mathcal{O}_2} : \boldsymbol{\mathcal{X}}_{\mathrm{pa}(\mathcal{O}_2)\setminus\mathcal{O}_2} \times \boldsymbol{\mathcal{E}}_{\mathrm{pa}(\mathcal{O}_2)} \to \boldsymbol{\mathcal{X}}_{\mathcal{O}_2}$ be measurable solution functions for $\mathcal{M}$ w.r.t. $\mathcal{O}_1$ and $\mathcal{O}_2$, respectively. Define now the mapping $\boldsymbol{h}_{\tilde{\mathcal{O}}} : \boldsymbol{\mathcal{X}}_{\widetilde{\mathrm{pa}}(\tilde{\mathcal{O}})\setminus\tilde{\mathcal{O}}} \times \boldsymbol{\mathcal{E}}_{\widetilde{\mathrm{pa}}(\tilde{\mathcal{O}})} \to \boldsymbol{\mathcal{X}}_{\tilde{\mathcal{O}}}$ by

$$(\boldsymbol{h}_{\tilde{\mathcal{O}}})_{\tilde{\mathcal{O}}\cap\mathcal{I}}(\boldsymbol{x}_{\widetilde{\mathrm{pa}}(\tilde{\mathcal{O}})\setminus\tilde{\mathcal{O}}}, \boldsymbol{e}_{\widetilde{\mathrm{pa}}(\tilde{\mathcal{O}})}) := \boldsymbol{g}_{\mathcal{O}_1}(\boldsymbol{x}_{\widetilde{\mathrm{pa}}(\mathcal{O}_1)\setminus\mathcal{O}_1}, \boldsymbol{e}_{\widetilde{\mathrm{pa}}(\mathcal{O}_1)})$$
$$(\boldsymbol{h}_{\tilde{\mathcal{O}}})_{\tilde{\mathcal{O}}\cap\mathcal{I}'}(\boldsymbol{x}_{\widetilde{\mathrm{pa}}(\tilde{\mathcal{O}})\setminus\tilde{\mathcal{O}}}, \boldsymbol{e}_{\widetilde{\mathrm{pa}}(\tilde{\mathcal{O}})}) := \boldsymbol{g}_{\mathcal{O}_2}(\boldsymbol{x}_{\widetilde{\mathrm{pa}}(\mathcal{O}_2')\setminus\mathcal{O}_2'}, \boldsymbol{e}_{\widetilde{\mathrm{pa}}(\mathcal{O}_2')}),$$

where we define $\widetilde{\mathrm{pa}} := \mathrm{pa}_{\mathcal{G}^a(\mathcal{M}^{\mathrm{twin}})}$ as the parents w.r.t. the twin graph $\mathcal{G}^a(\mathcal{M}^{\mathrm{twin}})$. Then by construction this mapping $\boldsymbol{h}_{\tilde{\mathcal{O}}}$ is a measurable solution function for $\mathcal{M}^{\mathrm{twin}}$ w.r.t. $\tilde{\mathcal{O}}$, and it is clear that $\mathcal{M}^{\mathrm{twin}}$ is uniquely solvable w.r.t. $\tilde{\mathcal{O}}$.

Lastly, it follows that the observational and all the intervened models of $\mathcal{M}$ and $\mathcal{M}^{\mathrm{twin}}$ are uniquely solvable. From Theorem 3.6 we conclude that $\mathcal{M}$ induces unique observational, interventional and counterfactual distributions. □

PROOF OF COROLLARY 8.3. This follows from Corollary A.22. □

APPENDIX F: MEASURABLE SELECTION THEOREMS

In this appendix, we derive some lemmas and state two measurable selection theorems that are used in several proofs in Appendix E. First, we introduce the measure theoretic notation and terminology needed to understand the results (see [12] for more details).

DEFINITION F.1 (Standard measurable space). *A measurable space* $(\boldsymbol{\mathcal{X}}, \boldsymbol{\Sigma})$ *is a* standard measurable space *if it is isomorphic to* $(\boldsymbol{\mathcal{Y}}, \mathcal{B}(\boldsymbol{\mathcal{Y}}))$*, where* $\boldsymbol{\mathcal{Y}}$ *is a Polish space, that is, a separable completely metrizable space,*[8] *and* $\mathcal{B}(\boldsymbol{\mathcal{Y}})$ *are the Borel subsets of* $\boldsymbol{\mathcal{Y}}$*, that is, the* σ*-algebra generated by the open sets in* $\boldsymbol{\mathcal{Y}}$*. A measure space* $(\boldsymbol{\mathcal{X}}, \boldsymbol{\Sigma}, \boldsymbol{\mu})$ *is a* standard probability space *if* $(\boldsymbol{\mathcal{X}}, \boldsymbol{\Sigma})$ *is a standard measurable space and* $\boldsymbol{\mu}$ *is a probability measure.*

Examples of standard measurable spaces are the open and closed subsets of $\mathbb{R}^d$, and the finite sets with the usual complete metric. If we say that $\boldsymbol{\mathcal{X}}$ is a standard measurable space, then we implicitly assume that there exists a σ-algebra $\boldsymbol{\Sigma}$ such that $(\boldsymbol{\mathcal{X}}, \boldsymbol{\Sigma})$ is a standard measurable space. Similarly, if we say that $\boldsymbol{\mathcal{X}}$ is a standard probability space with probability measure $\mathbb{P}_{\boldsymbol{\mathcal{X}}}$, then we implicitly assume that there exists a σ-algebra $\boldsymbol{\Sigma}$ such that $(\boldsymbol{\mathcal{X}}, \boldsymbol{\Sigma}, \mathbb{P}_{\boldsymbol{\mathcal{X}}})$ is a standard probability space.

DEFINITION F.2 (Analytic set). *Let* $\boldsymbol{\mathcal{X}}$ *be a Polish space. A set* $\boldsymbol{\mathcal{A}} \subseteq \boldsymbol{\mathcal{X}}$ *is called* analytic *if there exist a Polish space* $\boldsymbol{\mathcal{Y}}$ *and a continuous mapping* $\boldsymbol{f} : \boldsymbol{\mathcal{Y}} \to \boldsymbol{\mathcal{X}}$ *with* $\boldsymbol{f}(\boldsymbol{\mathcal{Y}}) = \boldsymbol{\mathcal{A}}$*.*

---

[8] A *metrizable space* is a topological space $\boldsymbol{\mathcal{X}}$ for which there exists a metric d such that $(\boldsymbol{\mathcal{X}}, d)$ is a metric space and induces the topology on $\boldsymbol{\mathcal{X}}$. For a metric space $(\boldsymbol{\mathcal{X}}, d)$, a *Cauchy sequence* is a sequence $(x_n)_{n\in\mathbb{N}}$ of elements of $\boldsymbol{\mathcal{X}}$ such that for every $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all natural numbers $p, q > N$ we have $d(x_n, x_m) < \epsilon$. We call $(\boldsymbol{\mathcal{X}}, d)$ *complete* if every Cauchy sequence has a limit in $\boldsymbol{\mathcal{X}}$. A *completely metrizable space* is a topological space $\boldsymbol{\mathcal{X}}$ for which there exists a metric d such that $(\boldsymbol{\mathcal{X}}, d)$ is a complete metric space that induces the topology on $\boldsymbol{\mathcal{X}}$. A topological space $\boldsymbol{\mathcal{X}}$ is called *separable* if it contains a countable dense subset, that is, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ of elements in $\boldsymbol{\mathcal{X}}$ such that every nonempty open subset of $\boldsymbol{\mathcal{X}}$ contains at least one element of the sequence. A separable completely metrizable space is called a *Polish space* (see [5] and [12] for more details).

LEMMA F.3. *Let $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ be standard measurable spaces and $\boldsymbol{f}:\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{Y}}$ a measurable mapping. Then*

1. *every measurable set $\boldsymbol{\mathcal{A}}\subseteq\boldsymbol{\mathcal{X}}$ is analytic;*
2. *if the subsets $\boldsymbol{\mathcal{A}}\subseteq\boldsymbol{\mathcal{X}}$ and $\tilde{\boldsymbol{\mathcal{A}}}\subseteq\boldsymbol{\mathcal{Y}}$ are analytic, then the sets $\boldsymbol{f}(\boldsymbol{\mathcal{A}})$ and $\boldsymbol{f}^{-1}(\tilde{\boldsymbol{\mathcal{A}}})$ are analytic.*

PROOF. From Proposition 13.7 in [12] it follows that every measurable set $\boldsymbol{\mathcal{A}}\subseteq\boldsymbol{\mathcal{X}}$ is analytic. From Proposition 14.4.(ii) in [12] it follows that the image and the preimage of an analytic set is an analytic set. □

DEFINITION F.4 ($\boldsymbol{\mu}$-measurability). *Let $(\boldsymbol{\mathcal{X}},\boldsymbol{\Sigma},\boldsymbol{\mu})$ be a measure space. A set $\boldsymbol{\mathcal{E}}\subseteq\boldsymbol{\mathcal{X}}$ is called a* $\boldsymbol{\mu}$-null set *if there exists a $\boldsymbol{\mathcal{A}}\in\boldsymbol{\Sigma}$ with $\boldsymbol{\mathcal{E}}\subseteq\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mu}(\boldsymbol{\mathcal{A}})=0$. We denote the class of $\boldsymbol{\mu}$-null sets by $\boldsymbol{\mathcal{N}}$, and we denote the σ-algebra generated by $\boldsymbol{\Sigma}\cup\boldsymbol{\mathcal{N}}$ by $\bar{\boldsymbol{\Sigma}}$, and its members are called the* $\boldsymbol{\mu}$-measurable *sets. Note that each member of $\bar{\boldsymbol{\Sigma}}$ is of the form $\boldsymbol{\mathcal{A}}\cup\boldsymbol{\mathcal{E}}$ with $\boldsymbol{\mathcal{A}}\in\boldsymbol{\Sigma}$ and $\boldsymbol{\mathcal{E}}\in\boldsymbol{\mathcal{N}}$. The measure $\boldsymbol{\mu}$ is extended to a measure $\bar{\boldsymbol{\mu}}$ on $\bar{\boldsymbol{\Sigma}}$, by $\bar{\boldsymbol{\mu}}(\boldsymbol{\mathcal{A}}\cup\boldsymbol{\mathcal{E}})=\boldsymbol{\mu}(\boldsymbol{\mathcal{A}})$ for every $\boldsymbol{\mathcal{A}}\in\boldsymbol{\Sigma}$ and $\boldsymbol{\mathcal{E}}\in\boldsymbol{\mathcal{N}}$, and is called its* completion*. A mapping $\boldsymbol{f}:\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{Y}}$ between measurable spaces is called* $\boldsymbol{\mu}$-measurable *if the inverse image $\boldsymbol{f}^{-1}(\boldsymbol{\mathcal{C}})$ of every measurable set $\boldsymbol{\mathcal{C}}\subseteq\boldsymbol{\mathcal{Y}}$ is $\boldsymbol{\mu}$-measurable.*

DEFINITION F.5 (Universal measurability). *Let $(\boldsymbol{\mathcal{X}},\boldsymbol{\Sigma})$ be a standard measurable space. A set $\boldsymbol{\mathcal{A}}\subseteq\boldsymbol{\mathcal{X}}$ is called* universally measurable *if it is $\boldsymbol{\mu}$-measurable for every σ-finite measure[9] $\boldsymbol{\mu}$ on $\boldsymbol{\mathcal{X}}$ (i.e., in particular every probability measure). A mapping $\boldsymbol{f}:\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{Y}}$ between standard measurable spaces is* universally measurable *if it is $\boldsymbol{\mu}$-measurable for every σ-finite measure $\boldsymbol{\mu}$.*

LEMMA F.6. *Let $\boldsymbol{\mathcal{E}}$ be a standard probability space with probability measure $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$ and $\boldsymbol{\mathcal{A}}\subseteq\boldsymbol{\mathcal{E}}$ an analytic set. Then $\boldsymbol{\mathcal{A}}$ is $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-measurable and there exist measurable sets $\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{T}}\subseteq\boldsymbol{\mathcal{E}}$ such that $\boldsymbol{\mathcal{S}}\subseteq\boldsymbol{\mathcal{A}}\subseteq\boldsymbol{\mathcal{T}}$ and $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{S}})=\bar{\mathbb{P}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{A}})=\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{T}})$, where $\bar{\mathbb{P}}_{\boldsymbol{\mathcal{E}}}$ is the completion of $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$.*

PROOF. Let $\boldsymbol{\mathcal{A}}\subseteq\boldsymbol{\mathcal{E}}$ be an analytic set. Since every analytic set in a standard measurable space is a universally measurable set (see Theorem 21.10 in [12]), we know that $\boldsymbol{\mathcal{A}}$ is a universally measurable set, and hence it is in particular a $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-measurable set. Thus, there exist a measurable set $\boldsymbol{\mathcal{S}}\subseteq\boldsymbol{\mathcal{E}}$ and a $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-null set $\boldsymbol{\mathcal{C}}\subseteq\boldsymbol{\mathcal{E}}$ such that $\boldsymbol{\mathcal{A}}=\boldsymbol{\mathcal{S}}\cup\boldsymbol{\mathcal{C}}$ and $\bar{\mathbb{P}}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{A}})=\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{S}})$, where $\bar{\mathbb{P}}_{\boldsymbol{\mathcal{E}}}$ is the completion of $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$. Moreover, there exists a measurable set $\tilde{\boldsymbol{\mathcal{C}}}\subseteq\boldsymbol{\mathcal{E}}$ such that $\boldsymbol{\mathcal{C}}\subseteq\tilde{\boldsymbol{\mathcal{C}}}$ and $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\tilde{\boldsymbol{\mathcal{C}}})=0$. Let $\boldsymbol{\mathcal{T}}:=\boldsymbol{\mathcal{S}}\cup\tilde{\boldsymbol{\mathcal{C}}}$, then $\boldsymbol{\mathcal{A}}\subseteq\boldsymbol{\mathcal{T}}$ and $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{T}})=\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{S}})$. □

LEMMA F.7. *Let $\boldsymbol{f}:\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{Y}}$ be a $\boldsymbol{\mu}$-measurable mapping. If $\boldsymbol{\mathcal{Y}}$ is countably generated, then there exists a measurable mapping $\boldsymbol{g}:\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{Y}}$ such that $\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{g}(\boldsymbol{x})$ holds $\boldsymbol{\mu}$-a.e..*

PROOF. Let the σ-algebra of $\boldsymbol{\mathcal{Y}}$ be generated by the countable generating set $\{\boldsymbol{\mathcal{C}}_n\}_{n\in\mathbb{N}}$. The $\boldsymbol{\mu}$-measurable set $\boldsymbol{f}^{-1}(\boldsymbol{\mathcal{C}}_n)=\boldsymbol{\mathcal{A}}_n\cup\boldsymbol{\mathcal{E}}_n$ for some $\boldsymbol{\mathcal{A}}_n\in\boldsymbol{\Sigma}$ and some $\boldsymbol{\mathcal{E}}_n\in\boldsymbol{\mathcal{N}}$ and hence there is some $\boldsymbol{\mathcal{E}}_n\subseteq\boldsymbol{\mathcal{B}}_n\in\boldsymbol{\Sigma}$ such that $\boldsymbol{\mu}(\boldsymbol{\mathcal{B}}_n)=0$. Let $\hat{\boldsymbol{\mathcal{B}}}=\cup_{n\in\mathbb{N}}\boldsymbol{\mathcal{B}}_n$, $\hat{\boldsymbol{\mathcal{A}}}_n=\boldsymbol{\mathcal{A}}_n\setminus\hat{\boldsymbol{\mathcal{B}}}$ and $\hat{\boldsymbol{\mathcal{A}}}=\cup_{n\in\mathbb{N}}\hat{\boldsymbol{\mathcal{A}}}_n$, then $\boldsymbol{\mu}(\hat{\boldsymbol{\mathcal{B}}})=0$, $\hat{\boldsymbol{\mathcal{A}}}$ and $\hat{\boldsymbol{\mathcal{B}}}$ are disjoint and $\boldsymbol{\mathcal{X}}=\hat{\boldsymbol{\mathcal{A}}}\cup\hat{\boldsymbol{\mathcal{B}}}$. Now define the mapping $\boldsymbol{g}:\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{Y}}$ by

$$\boldsymbol{g}(\boldsymbol{x}):=\begin{cases}\boldsymbol{f}(\boldsymbol{x}) & \text{if } \boldsymbol{x}\in\hat{\boldsymbol{\mathcal{A}}},\\ \boldsymbol{y}_0 & \text{otherwise,}\end{cases}$$

[9] A measure $\boldsymbol{\mu}$ on a measurable space $(\boldsymbol{\mathcal{X}},\boldsymbol{\Sigma})$ is called *σ-finite* if $\boldsymbol{\mathcal{X}}=\cup_{n\in\mathbb{N}}\boldsymbol{\mathcal{A}}_n$, with $\boldsymbol{\mathcal{A}}_n\in\boldsymbol{\Sigma}$, $\boldsymbol{\mu}(\boldsymbol{\mathcal{A}}_n)<\infty$.

where for $\boldsymbol{y}_0$ we can take an arbitrary point in $\boldsymbol{\mathcal{Y}}$. This mapping $\boldsymbol{g}$ is measurable since for each generator $\boldsymbol{\mathcal{C}}_n$ we have

$$\boldsymbol{g}^{-1}(\boldsymbol{\mathcal{C}}_n) = \begin{cases} \hat{\boldsymbol{\mathcal{A}}}_n & \text{if } \boldsymbol{y}_0 \notin \boldsymbol{\mathcal{C}}_n, \\ \hat{\boldsymbol{\mathcal{A}}}_n \cup \hat{\boldsymbol{\mathcal{B}}} & \text{otherwise.} \end{cases}$$

is in $\boldsymbol{\Sigma}$. Moreover, $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{x})$ $\boldsymbol{\mu}$-almost everywhere. □

With this result at hand we can now prove the first measurable selection theorem.

THEOREM F.8 (Measurable selection theorem). *Let $\boldsymbol{\mathcal{E}}$ be a standard probability space with probability measure $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$, $\boldsymbol{\mathcal{X}}$ a standard measurable space and $\boldsymbol{\mathcal{S}} \subseteq \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{X}}$ a measurable set such that $\boldsymbol{\mathcal{E}} \setminus \boldsymbol{pr}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{S}})$ is a $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-null set, where $\boldsymbol{pr}_{\boldsymbol{\mathcal{E}}} : \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{E}}$ is the projection mapping on $\boldsymbol{\mathcal{E}}$. Then there exists a measurable mapping $\boldsymbol{g} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ such that $(\boldsymbol{e}, \boldsymbol{g}(\boldsymbol{e})) \in \boldsymbol{\mathcal{S}}$ for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$.*

PROOF. Take the subset $\hat{\boldsymbol{\mathcal{E}}} := \boldsymbol{\mathcal{E}} \setminus \boldsymbol{\mathcal{B}}$, for some measurable set $\boldsymbol{\mathcal{B}} \supseteq \boldsymbol{\mathcal{E}} \setminus \boldsymbol{pr}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{S}})$ and $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{B}}) = 0$, and note that $\hat{\boldsymbol{\mathcal{E}}}$ is a standard measurable space (see Corollary 13.4 in [12]) and $\hat{\boldsymbol{\mathcal{E}}} \subseteq \boldsymbol{pr}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{S}})$. Let $\hat{\boldsymbol{\mathcal{S}}} = \boldsymbol{\mathcal{S}} \cap (\hat{\boldsymbol{\mathcal{E}}} \times \boldsymbol{\mathcal{X}})$. Because the set $\hat{\boldsymbol{\mathcal{S}}}$ is measurable, it is in particular analytic (see Lemma F.3). It follows by the Jankov-von Neumann Theorem (see Theorem 18.8 or 29.9 in [12]) that $\hat{\boldsymbol{\mathcal{S}}}$ has a universally measurable uniformizing function, that is, there exists a universally measurable mapping $\hat{\boldsymbol{g}} : \hat{\boldsymbol{\mathcal{E}}} \to \boldsymbol{\mathcal{X}}$ such that for all $\boldsymbol{e} \in \hat{\boldsymbol{\mathcal{E}}}$, $(\boldsymbol{e}, \hat{\boldsymbol{g}}(\boldsymbol{e})) \in \hat{\boldsymbol{\mathcal{S}}}$. Hence, in particular, it is $\mathbb{P}_{\boldsymbol{\mathcal{E}}}\big|_{\hat{\boldsymbol{\mathcal{E}}}}$-measurable, where $\mathbb{P}_{\boldsymbol{\mathcal{E}}}\big|_{\hat{\boldsymbol{\mathcal{E}}}}$ is the restriction of $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$ to $\hat{\boldsymbol{\mathcal{E}}}$.

Now define the mapping $\boldsymbol{g}^* : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ by

$$\boldsymbol{g}^*(\boldsymbol{e}) := \begin{cases} \hat{\boldsymbol{g}}(\boldsymbol{e}) & \text{if } \boldsymbol{e} \in \hat{\boldsymbol{\mathcal{E}}} \\ \boldsymbol{x}_0 & \text{otherwise,} \end{cases}$$

where for $\boldsymbol{x}_0$ we can take an arbitrary point in $\boldsymbol{\mathcal{X}}$. Then this mapping $\boldsymbol{g}^*$ is $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-measurable. To see this, take any measurable set $\boldsymbol{\mathcal{C}} \subseteq \boldsymbol{\mathcal{X}}$, then

$$\boldsymbol{g}^{*-1}(\boldsymbol{\mathcal{C}}) = \begin{cases} \hat{\boldsymbol{g}}^{-1}(\boldsymbol{\mathcal{C}}) & \text{if } \boldsymbol{x}_0 \notin \boldsymbol{\mathcal{C}} \\ \hat{\boldsymbol{g}}^{-1}(\boldsymbol{\mathcal{C}}) \cup \boldsymbol{\mathcal{B}} & \text{otherwise.} \end{cases}$$

Because $\hat{\boldsymbol{g}}^{-1}(\boldsymbol{\mathcal{C}})$ is $\mathbb{P}_{\boldsymbol{\mathcal{E}}}\big|_{\hat{\boldsymbol{\mathcal{E}}}}$-measurable it is also $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-measurable and thus $\boldsymbol{g}^{*-1}(\boldsymbol{\mathcal{C}})$ is $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-measurable.

By Lemma F.7 and the fact that standard measurable spaces are countably generated (see Proposition 12.1 in [12]), we prove the existence of a measurable mapping $\boldsymbol{g} : \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{X}}$ such that $\boldsymbol{g}^* = \boldsymbol{g}$ $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-a.e. and thus it satisfies $(\boldsymbol{e}, \boldsymbol{g}(\boldsymbol{e})) \in \boldsymbol{\mathcal{S}}$ for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$. □

This theorem rests on the assumption that the standard measurable space $\boldsymbol{\mathcal{E}}$ has a probability measure $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$. If this space becomes the product space $\boldsymbol{\mathcal{Y}} \times \boldsymbol{\mathcal{E}}$, for some standard measurable space $\boldsymbol{\mathcal{Y}}$ where only the space $\boldsymbol{\mathcal{E}}$ has a probability measure, then in general this theorem does not hold anymore. However, if we assume in addition that the fibers of $\boldsymbol{\mathcal{S}}$ in $\boldsymbol{\mathcal{Y}}$ are σ-compact for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$, then we can prove a second measurable selection theorem. A topological space is σ-compact if it is the union of countably many compact subspaces. For example, all countable discrete spaces, every interval of the real line, and moreover all the Euclidean spaces are σ-compact spaces.

THEOREM F.9 (Second measurable selection theorem). *Let $\boldsymbol{\mathcal{E}}$ be a standard probability space with probability measure $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ standard measurable spaces and $\boldsymbol{\mathcal{S}} \subseteq \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{Y}}$ a measurable set such that $\boldsymbol{\mathcal{E}} \setminus \mathcal{K}_\sigma$ is a $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-null set, where*

$$\mathcal{K}_\sigma := \{\boldsymbol{e} \in \boldsymbol{\mathcal{E}} : \forall \boldsymbol{x} \in \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{S}}_{(\boldsymbol{x},\boldsymbol{e})} \text{ is nonempty and } \sigma\text{-compact})\},$$

with $\boldsymbol{\mathcal{S}}_{(\boldsymbol{x},\boldsymbol{e})}$ denoting the fiber over $(\boldsymbol{x},\boldsymbol{e})$, that is

$$\boldsymbol{\mathcal{S}}_{(\boldsymbol{x},\boldsymbol{e})} := \{\boldsymbol{y} \in \boldsymbol{\mathcal{Y}} : (\boldsymbol{x},\boldsymbol{e},\boldsymbol{y}) \in \boldsymbol{\mathcal{S}}\}.$$

Then there exists a measurable mapping $\boldsymbol{g} : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{Y}}$ such that for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ we have $(\boldsymbol{x},\boldsymbol{e},\boldsymbol{g}(\boldsymbol{x},\boldsymbol{e})) \in \boldsymbol{\mathcal{S}}$.

PROOF. Take the subset $\hat{\boldsymbol{\mathcal{E}}} := \boldsymbol{\mathcal{E}} \setminus \boldsymbol{\mathcal{B}}$, for some measurable set $\boldsymbol{\mathcal{B}} \supseteq \boldsymbol{\mathcal{E}} \setminus \mathcal{K}_\sigma$ and $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{B}}) = 0$. Note that $\hat{\boldsymbol{\mathcal{E}}}$ is a standard measurable space, $\hat{\boldsymbol{\mathcal{E}}} \subseteq \mathcal{K}_\sigma$ and $\hat{\boldsymbol{\mathcal{S}}} = \boldsymbol{\mathcal{S}} \cap (\boldsymbol{\mathcal{X}} \times \hat{\boldsymbol{\mathcal{E}}} \times \boldsymbol{\mathcal{Y}})$ is measurable. By assumption, for each $(\boldsymbol{x},\boldsymbol{e}) \in \boldsymbol{\mathcal{X}} \times \hat{\boldsymbol{\mathcal{E}}}$ the fiber $\hat{\boldsymbol{\mathcal{S}}}_{(\boldsymbol{x},\boldsymbol{e})}$ is nonempty and σ-compact and hence by applying the Theorem of Arsenin-Kunugui (see Theorem 35.46 in [12]) it follows that the set $\hat{\boldsymbol{\mathcal{S}}}$ has a measurable uniformizing function, that is, there exists a measurable mapping $\hat{\boldsymbol{g}} : \boldsymbol{\mathcal{X}} \times \hat{\boldsymbol{\mathcal{E}}} \to \boldsymbol{\mathcal{Y}}$ such that for all $(\boldsymbol{x},\boldsymbol{e}) \in \boldsymbol{\mathcal{X}} \times \hat{\boldsymbol{\mathcal{E}}}$, $(\boldsymbol{x},\boldsymbol{e},\hat{\boldsymbol{g}}(\boldsymbol{x},\boldsymbol{e})) \in \hat{\boldsymbol{\mathcal{S}}}$. Now define the mapping $\boldsymbol{g} : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{E}} \to \boldsymbol{\mathcal{Y}}$ by

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{e}) := \begin{cases} \hat{\boldsymbol{g}}(\boldsymbol{x},\boldsymbol{e}) & \text{if } \boldsymbol{e} \in \hat{\boldsymbol{\mathcal{E}}} \\ \boldsymbol{y}_0 & \text{otherwise,} \end{cases}$$

where for $\boldsymbol{y}_0$ we can take an arbitrary point in $\boldsymbol{\mathcal{Y}}$. This mapping $\boldsymbol{g}$ inherits the measurability from $\hat{\boldsymbol{g}}$ and it satisfies for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$ and for all $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ that $(\boldsymbol{x},\boldsymbol{e},\boldsymbol{g}(\boldsymbol{x},\boldsymbol{e})) \in \boldsymbol{\mathcal{S}}$. □

The next two lemmas provide some useful properties for the "for $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-almost every $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$" quantifier.

LEMMA F.10. *Let $\boldsymbol{\phi} : \boldsymbol{\mathcal{E}} \to \tilde{\boldsymbol{\mathcal{E}}}$ be a measurable map between two standard measurable spaces. Let $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$ be a probability measure on $\boldsymbol{\mathcal{E}}$ and let $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}} = \mathbb{P}_{\boldsymbol{\mathcal{E}}} \circ \boldsymbol{\phi}^{-1}$ be its push-forward under $\boldsymbol{\phi}$. Let $\tilde{P} : \tilde{\boldsymbol{\mathcal{E}}} \to \{0,1\}$ be a property, that is, a (measurable) boolean-valued function on $\tilde{\boldsymbol{\mathcal{E}}}$. Then the property $P = \tilde{P} \circ \boldsymbol{\phi}$ on $\boldsymbol{\mathcal{E}}$ holds $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-a.e. if and only if the property $\tilde{P}$ holds $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}$-a.e..*

PROOF. Assume the property $P = \tilde{P} \circ \boldsymbol{\phi}$ holds $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-a.e., then $\boldsymbol{\mathcal{C}} = \{\boldsymbol{e} \in \boldsymbol{\mathcal{E}} : P(\boldsymbol{e}) = 1\}$ contains a measurable set $\boldsymbol{\mathcal{C}}^*$ with $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-measure 1, that is, $\boldsymbol{\mathcal{C}}^* \subseteq \boldsymbol{\mathcal{C}}$ and $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{C}}^*) = 1$. By Lemma F.3, $\boldsymbol{\phi}(\boldsymbol{\mathcal{C}}^*)$ is analytic. By Lemma F.6, there exist measurable sets $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}$ such that $\boldsymbol{\mathcal{A}} \subseteq \boldsymbol{\phi}(\boldsymbol{\mathcal{C}}^*) \subseteq \boldsymbol{\mathcal{B}}$ and $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}(\boldsymbol{\mathcal{A}}) = \mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}(\boldsymbol{\mathcal{B}})$. Because $\boldsymbol{\phi}$ is measurable, $\boldsymbol{\phi}^{-1}(\boldsymbol{\mathcal{A}})$ and $\boldsymbol{\phi}^{-1}(\boldsymbol{\mathcal{B}})$ are both measurable. Also, $\boldsymbol{\phi}^{-1}(\boldsymbol{\mathcal{A}}) \subseteq \boldsymbol{\phi}^{-1}(\boldsymbol{\phi}(\boldsymbol{\mathcal{C}}^*)) \subseteq \boldsymbol{\phi}^{-1}(\boldsymbol{\mathcal{B}})$. As $\boldsymbol{\mathcal{C}}^* \subseteq \boldsymbol{\phi}^{-1}(\boldsymbol{\phi}(\boldsymbol{\mathcal{C}}^*))$, we must have that $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\phi}^{-1}(\boldsymbol{\mathcal{B}})) \geq \mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\mathcal{C}}^*) = 1$. Hence $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}(\boldsymbol{\mathcal{A}}) = \mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}(\boldsymbol{\mathcal{B}}) = 1$. Note that as $\boldsymbol{\mathcal{C}}^* \subseteq \boldsymbol{\mathcal{C}}$, $\boldsymbol{\mathcal{A}} \subseteq \boldsymbol{\phi}(\boldsymbol{\mathcal{C}}^*) \subseteq \boldsymbol{\phi}(\boldsymbol{\mathcal{C}}) \subseteq \{\tilde{e} \in \tilde{\boldsymbol{\mathcal{E}}} : \tilde{P}(\tilde{e}) = 1\}$. Hence the set $\tilde{\boldsymbol{\mathcal{C}}} := \{\tilde{e} \in \tilde{\boldsymbol{\mathcal{E}}} : \tilde{P}(\tilde{e}) = 1\}$ contains a measurable set of $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}$-measure 1, in other words, $\tilde{P}$ holds $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}$-a.s..

The converse is easier to prove. Suppose $\tilde{\boldsymbol{\mathcal{C}}} = \{\tilde{e} \in \tilde{\boldsymbol{\mathcal{E}}} : \tilde{P}(\tilde{e}) = 1\}$ contains a measurable set $\tilde{\boldsymbol{\mathcal{C}}}^*$ with $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}$-measure 1, that is, $\tilde{\boldsymbol{\mathcal{C}}}^* \subseteq \tilde{\boldsymbol{\mathcal{C}}}$ and $\mathbb{P}_{\tilde{\boldsymbol{\mathcal{E}}}}(\tilde{\boldsymbol{\mathcal{C}}}^*) = 1$. Because $\boldsymbol{\phi}$ is measurable, the set $\boldsymbol{\phi}^{-1}(\tilde{\boldsymbol{\mathcal{C}}}^*)$ is measurable and $\mathbb{P}_{\boldsymbol{\mathcal{E}}}(\boldsymbol{\phi}^{-1}(\tilde{\boldsymbol{\mathcal{C}}}^*)) = 1$, and furthermore, $\boldsymbol{\phi}^{-1}(\tilde{\boldsymbol{\mathcal{C}}}^*) \subseteq \boldsymbol{\phi}^{-1}(\tilde{\boldsymbol{\mathcal{C}}}) = \boldsymbol{\mathcal{C}}$. □

LEMMA F.11 (Some properties for the for-almost-every quantifier). *Let* $\boldsymbol{\mathcal{X}} = \mathcal{X} \times \tilde{\mathcal{X}}$ *and* $\boldsymbol{\mathcal{E}} = \mathcal{E} \times \tilde{\mathcal{E}}$ *be products of nonempty standard measurable spaces and* $\mathbb{P}_{\boldsymbol{\mathcal{E}}} = \mathbb{P}_{\mathcal{E}} \times \mathbb{P}_{\tilde{\mathcal{E}}}$ *be the product measure of probability measures* $\mathbb{P}_{\mathcal{E}}$ *and* $\mathbb{P}_{\tilde{\mathcal{E}}}$ *on* $\mathcal{E}$ *and* $\tilde{\mathcal{E}}$*, respectively. Denote by "*$\forall\!\!\!\forall \boldsymbol{e}$*" the quantifier "for* $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$*-almost every* $\boldsymbol{e} \in \boldsymbol{\mathcal{E}}$*" and by "*$\forall \boldsymbol{x}$*" the quantifier "for all* $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$*", and similarly for their components, for example, "*$\forall\!\!\!\forall e$*" for "for* $\mathbb{P}_{\mathcal{E}}$*-almost every* $e \in \mathcal{E}$*" and "*$\forall x$*" for "for all* $x \in \mathcal{X}$*". Then we have the following properties:*

1. $\forall\!\!\!\forall e : P(e) \implies \exists e : P(e)$ *(similarly to* $\forall x : P(x) \implies \exists x : P(x)$*);*
2. $\forall\!\!\!\forall e : P(e) \iff \forall\!\!\!\forall \boldsymbol{e} : P(e)$ *(similarly to* $\forall x : P(x) \iff \forall \boldsymbol{x} : P(x)$*);*
3. $\exists x \forall\!\!\!\forall e : P(x,e) \implies \forall\!\!\!\forall e \exists x : P(x,e)$ *(similarly to* $\exists x \forall e : P(x,e) \implies \forall e \exists x : P(x,e)$*);*
4. $\forall\!\!\!\forall e \forall x : P(x,e) \implies \forall x \forall\!\!\!\forall e : P(x,e)$ *(similarly to* $\forall e \forall x : P(x,e) \implies \forall x \forall e : P(x,e)$*);*
5. $\forall\!\!\!\forall \boldsymbol{e} : P(\boldsymbol{e}) \implies \exists \tilde{e} \forall\!\!\!\forall e : P(\boldsymbol{e})$ *(similarly to* $\forall \boldsymbol{x} : P(\boldsymbol{x}) \implies \exists \tilde{x} \forall x : P(\boldsymbol{x})$*);*
6. $\forall\!\!\!\forall e \forall x : P(x,e) \iff \forall\!\!\!\forall \boldsymbol{e} \forall \boldsymbol{x} : P(x,e)$*;*
7. $\forall\!\!\!\forall \boldsymbol{e} \forall \boldsymbol{x} : P(\boldsymbol{x},\boldsymbol{e}) \implies \exists \tilde{e} \exists \tilde{x} \forall\!\!\!\forall e \forall x : P(\boldsymbol{x},\boldsymbol{e})$*,*

where P *denotes a property, that is, a measurable boolean-valued function, on the corresponding measurable spaces and we write* $\boldsymbol{e}$ *and* $\boldsymbol{x}$ *for* $(e,\tilde{e})$ *and* $(x,\tilde{x})$*, respectively.*

PROOF. We only prove the statements that may not be immediately obvious.

Property 2. Let $pr_{\mathcal{E}} : \boldsymbol{\mathcal{E}} \to \mathcal{E}$ be the projection mapping on $\mathcal{E}$. Then by Lemma F.10 we have

$$\forall\!\!\!\forall e : P(e) \iff \forall\!\!\!\forall \boldsymbol{e} : P \circ \mathrm{pr}_{\mathcal{E}}(\boldsymbol{e}) \iff \forall\!\!\!\forall \boldsymbol{e} : P(e)\,.$$

Property 4: We have

$$\begin{aligned} &\forall\!\!\!\forall e \forall x : P(x,e) \\ &\implies \exists\, \mathbb{P}_{\mathcal{E}}\text{-null set } N \,\forall e \in \mathcal{E} \setminus N \,\forall x : P(x,e) \\ &\implies \exists\, \mathbb{P}_{\mathcal{E}}\text{-null set } N \,\forall x \,\forall e \in \mathcal{E} \setminus N : P(x,e) \\ &\implies \forall x \,\exists\, \mathbb{P}_{\mathcal{E}}\text{-null set } N \,\forall e \in \mathcal{E} \setminus N : P(x,e) \\ &\implies \forall x \forall\!\!\!\forall e : P(x,e)\,. \end{aligned}$$

Property 5: Let $\boldsymbol{N}$ be a measurable $\mathbb{P}_{\boldsymbol{\mathcal{E}}}$-null set such that $P(\boldsymbol{e})$ holds for all $\boldsymbol{e} \in \boldsymbol{\mathcal{E}} \setminus \boldsymbol{N}$. Define for $\tilde{e} \in \tilde{\mathcal{E}}$ the set $N_{\tilde{e}} := \{e \in \mathcal{E} : (e,\tilde{e}) \in \boldsymbol{N}\}$. Note that the sets $N_{\tilde{e}}$ are measurable. From Fubini's theorem it follows that for $\mathbb{P}_{\tilde{\mathcal{E}}}$-almost every $\tilde{e} \in \tilde{\mathcal{E}}$ we have $\mathbb{P}_{\mathcal{E}}(N_{\tilde{e}}) = 0$. That is, there exists a measurable $\mathbb{P}_{\tilde{\mathcal{E}}}$-null set $\tilde{N}$ such that $\mathbb{P}_{\mathcal{E}}(N_{\tilde{e}}) = 0$ for all $\tilde{e} \in \tilde{\mathcal{E}} \setminus \tilde{N}$. Hence, there exists $\tilde{e} \in \tilde{\mathcal{E}} \setminus \tilde{N}$ such that $\mathbb{P}_{\mathcal{E}}(N_{\tilde{e}}) = 0$; for all $e \in \mathcal{E} \setminus N_{\tilde{e}}$, $P(\boldsymbol{e})$ then holds. This means $\exists \tilde{e} \forall\!\!\!\forall e : P(\boldsymbol{e})$.

Property 7: We have

$$\begin{aligned} \forall\!\!\!\forall \boldsymbol{e} \forall \boldsymbol{x} : P(\boldsymbol{x},\boldsymbol{e}) &\implies \exists \tilde{e} \forall\!\!\!\forall e \forall \boldsymbol{x} : P(\boldsymbol{x},\boldsymbol{e}) \implies \exists \tilde{e} \forall\!\!\!\forall e \forall \tilde{x} \forall x : P(\boldsymbol{x},\boldsymbol{e}) \\ &\implies \exists \tilde{e} \forall \tilde{x} \forall\!\!\!\forall e \forall x : P(\boldsymbol{x},\boldsymbol{e}) \implies \exists \tilde{e} \exists \tilde{x} \forall\!\!\!\forall e \forall x : P(\boldsymbol{x},\boldsymbol{e})\,, \end{aligned}$$

where in the first equivalence we used Property 5, in the third equivalence we used Property 4 and in the last equivalence we used Property 1. □

REFERENCES

[1] BLOM, T., BONGERS, S. and MOOIJ, J. M. (2019). Beyond Structural Causal Models: Causal Constraints Models. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI-19)* (R. P. ADAMS and V. GOGATE, eds.). AUAI Press.

[2] BLOM, T., VAN DIEPEN, M. M. and MOOIJ, J. M. (2020). Conditional Independences and Causal Relations implied by Sets of Equations. *arXiv.org preprint* arXiv:2007.07183 [cs.AI].

[3] BOLLEN, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons, New York, USA.

[4] BONGERS, S., BLOM, T. and MOOIJ, J. M. (2021). Causal Modeling of Dynamical Systems. *arXiv.org preprint* arXiv:1803.08784v3 [cs.AI].

[5] COHN, D. L. (2013). *Measure Theory*, 2nd ed. Birkhäuser, Boston, USA.

[6] DAWID, A. P. (2002). Influence Diagrams for Causal Modelling and Inference. *International Statistical Review* **70** 161–189.

[7] EVANS, R. J. (2016). Graphs for Margins of Bayesian Networks. *Scandinavian Journal of Statistics* **43** 625–648.

[8] FORRÉ, P. and MOOIJ, J. M. (2017). Markov Properties for Graphical Models with Cycles and Latent Variables. *arXiv.org preprint* arXiv:1710.08775 [math.ST].

[9] GEIGER, D. (1990). Graphoids: A Qualitative Framework for Probabilistic Inference Technical Report No. R-142, Computer Science Department, University of California, Los Angeles, USA.

[10] GOLUB, G. and KAHAN, W. (1965). Calculating the Singular Values and Pseudo-Inverse of a Matrix. *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis* **2** 205–224.

[11] HYTTINEN, A., EBERHARDT, F. and HOYER, P. O. (2012). Learning Linear Cyclic Causal Models with Latent Variables. *Journal of Machine Learning Research* **13** 3387–3439.

[12] KECHRIS, A. S. (1995). *Classical Descriptive Set Theory*. *Graduate Texts in Mathematics* **156**. Springer-Verlag, New York, USA.

[13] KOSTER, J. T. A. (1996). Markov Properties of Nonrecursive Causal Models. *The Annals of Statistics* **24** 2148–2177.

[14] KOSTER, J. T. A. (1999). On the Validity of the Markov Interpretation of Path Diagrams of Gaussian Structural Equations Systems with Correlated Errors. *Scandinavian Journal of Statistics* **26** 413–431.

[15] LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. Clarendon Press, Oxford.

[16] LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H. G. (1990). Independence Properties of Directed Markov Fields. *Networks* **20** 491–505.

[17] MEEK, C. (1995). Strong Completeness and Faithfulness in Bayesian Networks. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)* (P. BESNARD and S. HANKS, eds.) 411–418. Morgan Kaufmann, San Francisco, CA, USA.

[18] NEAL, R. M. (2000). On Deducing Conditional Independence from d-Separation in Causal Graphs with Feedback. *Journal of Artificial Intelligence Research* **12** 87–91.

[19] PEARL, J. (1985). A Constraint Propagation Approach to Probabilistic Reasoning. In *Proceedings of the First Conference on Uncertainty in Artificial Intelligence (UAI-85)* (L. KANAL and J. LEMMER, eds.) 31–42. AUAI Press, Corvallis, Oregon, USA.

[20] PEARL, J. (2009). *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge University Press, New York, USA.

[21] PEARL, J. and DECHTER, R. (1996). Identifying Independence in Causal Graphs with Feedback. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI-96)* (E. HORVITZ and F. JENSEN, eds.) 420–426. Morgan Kaufmann, San Francisco, CA, USA.

[22] PENROSE, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* **51** 406–413.

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

[24] RICHARDSON, T. S. (1996). Models of Feedback: Interpretation and Discovery, PhD thesis, Carnegie Mellon University.

[25] RICHARDSON, T. S. and ROBINS, J. M. (2014). ACE Bounds; SEMs with Equilibrium Conditions. *Statistical Science* **29** 363-366.

[26] SPIRTES, P. (1994). Conditional Independence in Directed Cyclic Graphical Models for Feedback Technical Report No. CMU-PHIL-54, Carnegie Mellon University.

[27] SPIRTES, P. (1995). Directed Cyclic Graphical Representations of Feedback Models. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI-95)* (P. BESNARD and S. HANKS, eds.) 499–506. Morgan Kaufmann, San Francisco, CA, USA.

[28] SPIRTES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, Massachusetts.

[29] SPIRTES, P., RICHARDSON, T., MEEK, C., SCHEINES, R. and GLYMOUR, C. (1998). Using Path Diagrams as a Structural Equation Modelling Tool. *Sociological Methods & Research* **27** 182–225.