Biometrika (2024), **1**, 1, *pp*. 1–26
Printed in Great Britain

Axiomatization of interventional probability distributions

BY KAYVAN SADEGHI AND TERRY SOO
Department of Statistical Science, UCL, Gower Street, London, WC1E 6BT, U.K.
k.sadeghi@ucl.ac.uk t.soo@ucl.ac.uk

SUMMARY

Causal intervention is an essential tool in causal inference. It is axiomatized under the rules of do-calculus in the case of structure causal models. We provide simple axiomatizations for families of probability distributions to be different types of interventional distributions. Our axiomatizations neatly lead to a simple and clear theory of causality that has several advantages: it does not need to make use of any modeling assumptions such as those imposed by structural causal models; it only relies on interventions on single variables; it includes most cases with latent variables and causal cycles; and more importantly, it does not assume the existence of an underlying true causal graph as we do not take it as the primitive object—moreover, a causal graph is derived as a by-product of our theory. We show that, under our axiomatizations, the intervened distributions are Markovian to the defined intervened causal graphs, and an observed joint probability distribution is Markovian to the obtained causal graph; these results are consistent with the case of structural causal models, and as a result, the existing theory of causal inference applies. We also show that a large class of natural structural causal models satisfy the theory presented here. The aim of this paper is axiomatization of interventional families, which is subtly different from causal modeling.

Some key words: Ancestral graph; Causal graph; Directed graph; Do-calculus; Interventional distribution; Markov property; Structural causal model.

1. INTRODUCTION

1·1. *Background*

A popular approach to infer causal relationships is to use the concept of intervention as opposed to observation. For example, as described in Peters et al. (2017), it can be observed that there is a correlation between smoking and the colour of the teeth, but no matter how much one whitens somebody's teeth, it would not affect their smoking habits. On the other hand, forcing someone to smoke would affect the colour of their teeth. Hence, smoking has a causal effect on the colour of the teeth, but not vice versa.

Interventions have generally been embedded in the setting of structural causal models, also known as structural equation models (Pearl, 1988; Spirtes et al., 2000). These are a system of assignments for a set of random variables ordered by an associated true causal graph, which is generally assumed to be unknown. Structural causal models utilise the theory of graphical (Markov) models, which are statistical models over graphs with nodes as random variables and edges that indicate some types of conditional dependencies; see Lauritzen (1996).

An axiomatic approach to interventions for structural causal models, known as Pearl's do-calculus (Pearl, 2009), has been developed for identifiability of interventional distributions from the observational ones; see also Huang & Valtorta (2006); Shpitser & Pearl (2006) for some further theoretical developments. There has also been a substantial amount of work on generalizing the concept of intervention from the case of directed acyclic graphs, sometimes called Bayesian networks, to more general graphs containing bidirected edges, which indicate the existence of latent variables (Zhang, 2008), and directed cycles (Bongers et al., 2021). However, most of these attempts stay within the setting of structural causal models or, at least, under the assumption that there exists an underlying true causal graph that somehow captures the causal relationships (Woodward, 2004).

Interventions on structural causal models have been defined to take various different forms; see Korb et al. (2004); Eberhardt & Scheines (2007). The type of intervention with which we are dealing here is hard in the sense that it destroys all the causes of the intervened variable, and is stochastic in the sense that it replaces the marginal distribution of the intervened variable with a new distribution; although, we will show in Remark 4 that an atomic, also called surgical, intervention, which forces the variable to have a specific value, can be easily adapted in this setting.

In this paper, without assuming any modeling assumptions such as those given in the setting of structural causal models, we give simple conditions for a family of joint distributions $\mathcal{P}_{\mathrm{do}} = \{P_{\mathrm{do}(1)}, \ldots, P_{\mathrm{do}(N)}\}$ to act as a well-behaved interventional family, so that one can think of $P_{\mathrm{do}(i)}$ as an interventional distribution on a single variable X_i, for each $i \in V = \{1, \ldots, N\}$ in a random vector X_V. We have confined ourselves, in this paper, to the case of a finite number of random variables, as most of the related theory was developed in the finite setting. As apparent from the context, our approach here is aligned with the interventional approach to causality rather than the counterfactual approach.

Here, we are not providing an alternative to the current mainstream setting or Pearlian setting, where it assumes the existence of a structural causal model and by applying an interventional approach, has led to extensive work on causal learning and estimation. We simply provide theoretical backing for this approach and generalize it beyond structural causal models, by providing certain axioms in order to derive as results some of the assumptions that have been used in the literature, specifically, the existence of the causal graph.

This paper carries certain important messages: There is no need to take the true causal graph as the primitive object—causal graph(s) can then be formally defined and derived from interventional families, rather than posited. The causal structure and graph can be solely derived from the family of interventional distributions; in other words, there is no need for an initial state, i.e., an underlying joint observational distribution P of X_V, to be be known for this purpose. However, we provide axioms such that the required consistency between the interventional family and the underlying observational distribution is satisfied when indeed the observational distribution is available, and such that one can measure the causal relationships. To derive the causal structure or graph, in most situations, one needs to rely only on single interventions once at a time. This is an advantage as much less information is used by only relying on single interventions. Indeed, there are real world situations in which one would like to consider intervening on several variables simultaneously; we believe a similar theory can be proposed in such cases.

We must emphasize that the work presented here is about axioms that interventional distributions should satisfy for the purpose of causal reasoning. These axioms should not be confused with a causal model whose goal is to provide correct interpretation of causal relationships and measuring their effects. This difference is quite subtle and could lead to confusion. Similarly, one should distinguish the causal graphs defined and derived here from a graph learned by structure learning from observational and, potentially, interventional data. The goal here is not structure learning (Spirtes et al., 2000; Colombo & Maathuis, 2014).

1·2. *Key results*

One of our central assumptions, Axiom 1, is that cause is transitive; see Hall (2000) for a philosophical discussion on the transitivity of the cause. Under the condition of singleton-transitivity and simple assumptions on conditional independence structure of $\mathcal{P}_{\mathrm{do}}$, we show that the causal relations are transitive; see Theorem 2. We provide a definition of causation similar to that in Peters et al. (2017). The concept of direct cause is defined in terms of the conditional independence properties of the interventional family, which is a departure from the widely-known definition (Woodward, 2004, page 55), and from this we define the intervened causal graph, and using these, we define the causal graph; see Section 4·2; this is a major relaxation of assumptions from the current paradigm where it is assumed that such a causal graph exists. The obtained causal graph allows bidirected edges and directed cycles without double edges consisting of a bidirected edge and an arrow. We call this family of graphs bowless directed mixed graphs. Later, in Section S6 of the supplementary material, we relax the bowlessness, property, in some cases. The generated graph is the true causal graph under the axiomatization, and we show that the definitions related to causal relationships and the graphical notions on the graph are interchangeable in Theorem 3.

One of our main theorems is that, under some additional assumptions, namely intersection and composition, intervened distributions $P_{\mathrm{do}(i)}$ in the interventional family are Markovian to the defined intervened graphs; see Theorem 4. We provide additional axioms, Axioms 2 and 3, to relate $\mathcal{P}_{\mathrm{do}}$ to an observed distribution P, and call the interventional family (strongly) observable. We show that the underlying distribution P for an observable interventional family is Markovian to the defined causal graph; see Theorem 5. Therefore, the established theory of causality using structural causal models, which mainly relies on the Markov property of the joint distribution of the structural causal model, could be followed from our theory.

We later provide additional axioms, in the supplementary material, for the case of ancestral causal graphs, to define what we call quantifiable interventional families that allow for measuring causal effects. We show that the quantifiable interventional families are strongly observational; see Theorem S1 in the supplementary material.

We also compare and contrast our theory with the structural causal model setting. We show that for structural causal models with certain simple properties, implied by faithfulness—such as the transitivity of the cause, the family of interventions on each node constitutes a strongly observable interventional family, and even without the transitivity assumption in the case of ancestral graphs, the causal graph generated by the theory presented in this paper is the same as the causal graph associated to the structural causal model; see Theorem 7.

Our theory is based only on intervening on single variables once at a time. We clearly identify cases, which can only be non-maximal and non-ancestral, where this theory may misidentify some direct causes; see Section 7.

1·3. *Related works*

To the best of our knowledge, most of the attempts to abstracting intervention or causality based on intervention in general are substantially different from our approach; see for example Rischel & Weichwald (2021) for a category-theory approach, and the more recent Park et al. (2023) for a measure-theoretic approach.

One such attempt is the seminal work of Dawid on the decision theoretic framework for causal inference; for example, see Dawid (2002, 2021). Our approach share the same concerns and spirit as that of Dawid's in focusing on the interventions rather than counterfactuals, as well as trying to justify the existence of the causal graph rather than assuming it, as we have heeded Dawid's caution (Dawid, 2010). However, the mechanics of the two works are different. First of all, we have not provided any statistical model like Dawid does. In addition, the goal of Dawid's work is mainly to enable "transfer of probabilistic information from an observational to an interventional setting," whereas, here, our starting point is interventions. Finally, our approach also covers a much more general class of causal graphs than directed acyclic graphs, considered by Dawid. We have not included influence diagrams, as proposed by Dawid, but believe that it should be possible to derive them together with their conditional-independence constraints using our causal graphs and Markov properties.

A more similar approach to ours is that of Bareinboim, Brito, and Pearl (Bareinboim et al., 2011). Like us the authors start with a family of interventional distributions with interventions defined on single variables. A difference is that they only work on atomic interventions, which requires alternative definitions to conditional independence, phrased as invariances—we believe this can be adapted to use stochastic intervention and regular conditional independence. Using this, they define the concept of direct cause, which is different, but of similar nature to how we define this concept. A major difference is that, in their paper, they provide different notions of compatibility of the interventional family and causal graphs by assuming certain conditions that include the global Markov property—in our work, we do not assume the global Markov property, and generate a graph directly from the interventional family by using direct cause, and prove the Markov property under certain axioms and conditions. Another important difference is that their work relies on the notion of observed initial state distribution to define interventions—as mentioned before, we do not need to rely on observational distribution to derive the causal graph.

The mentioned paper is purely on directed acyclic graphs, but in a more recent work (Bareinboim et al., 2022) the method was generalized to include arcs representing latent variables. The notion of Markov property has been replaced by semi-Markov property to deal with this generalization, but the difference between the two methods remains the same as described for the original paper.

Finally, we want to stress that in this paper, we are not using a counterfactual definition of causation. In Section S7, of the supplementary material, we will point out the differences between our work and what is known about axiomatization in the counterfactual setting, which is closely related to Lewis' work on modal logic.

2. PRELIMINARIES

2·1. *Conditional measures and independence*

We will work in the following setting. Let V be a finite set of size N. Let P be a probability measure on the product measurable space $\mathcal{X} = \prod_{i\in V} \mathcal{X}_i$. For $A \subseteq V$, we let $\mathcal{X}_A = \prod_{i\in A} \mathcal{X}_i$ and P^A be the marginal measure of P on $\mathcal{X}_A$ given by

$$P^A(W) = P(W \times \mathcal{X}_{V\setminus A})$$

for all measurable $W \subseteq \mathcal{X}_A$. We will use the notation $i \perp\!\!\!\perp_P j$ to mean that the marginal $P^{\{i,j\}} = P^{i,j}$ is the product measure $P^i \otimes P^j$ on $\mathcal{X}_i \times \mathcal{X}_j$, so that if $X = (X_1, \ldots, X_N)$ is a random vector defined on some probability space $(\Omega, \mathcal{F}, \mathrm{pr})$ taking values on $\mathcal{X}$ with law P, then X_i is independent of X_j (Dawid, 1979; Lauritzen, 1996).

For $x_A \in \mathcal{X}_A$, we let $P(\cdot \mid x_A) = \mathrm{pr}(X \in \cdot \mid X_A = x_A)$ denote a regular conditional probability (Chang & Pollard, 1997) so that in particular, we have the disintegration

$$\mathrm{pr}(X \in F) = P(F) = \int_{\mathcal{X}_A} P(F \mid x_A) dP^A(x_A),$$

for $F \subseteq \mathcal{X}$ measurable. More generally, consider disjoint subsets A, B, and C of V. We will often consider the marginal of a conditional measure, and have a slight abuse of notation that:

$$P^A(\cdot \mid x_C) = \{P(\cdot \mid x_C)\}^A.$$

We write $A \perp\!\!\!\perp_P B \mid C$ to denote that the measure $P^{A,B}(\cdot \mid x_C) = P^{A \cup B}(\cdot \mid x_C)$ is a product measure on $\mathcal{X}_A \times \mathcal{X}_B$ for P^C-almost all $x_C \in \mathcal{X}_C$, so that if X has law P, then X_A is conditionally independent of X_B given X_C. Sometimes, we will simply say that A is conditionally independent of B given C in P. In addition, when independence fails, we write $A \not\!\perp\!\!\!\perp_P B \mid C$ and say that A and B are conditionally dependent given C in P.

2·2. *Structural independence properties of a distribution*

A probability distribution P is always a semi-graphoid (Pearl, 1988), i.e., it satisfies the four following properties for disjoint subsets A, B, C, and D of V:

1. $A \perp\!\!\!\perp_P B \mid C$ if and only if $B \perp\!\!\!\perp_P A \mid C$ (symmetry);
2. if $A \perp\!\!\!\perp_P B \cup D \mid C$, then $A \perp\!\!\!\perp_P B \mid C$ and $A \perp\!\!\!\perp_P D \mid C$ (decomposition);
3. if $A \perp\!\!\!\perp_P B \cup D \mid C$, then $A \perp\!\!\!\perp_P B \mid C \cup D$ and $A \perp\!\!\!\perp_P D \mid C \cup B$ (weak union);
4. if $A \perp\!\!\!\perp_P B \mid C \cup D$ and $A \perp\!\!\!\perp_P D \mid C$, then $A \perp\!\!\!\perp_P B \cup D \mid C$ (contraction).

The reverse implication of contraction clearly holds by decomposition and weak union. We also use three different properties of conditional independence that are not always satisfied by probability distributions:

5. if $A \perp\!\!\!\perp_P B \mid C \cup D$ and $A \perp\!\!\!\perp_P D \mid C \cup B$, then $A \perp\!\!\!\perp_P B \cup D \mid C$ (intersection);
6. if $A \perp\!\!\!\perp_P B \mid C$ and $A \perp\!\!\!\perp_P D \mid C$, then $A \perp\!\!\!\perp_P B \cup D \mid C$ (composition);
7. if $i \perp\!\!\!\perp_P j \mid C$ and $i \perp\!\!\!\perp_P j \mid C \cup \{k\}$, then $i \perp\!\!\!\perp_P k \mid C$ or $j \perp\!\!\!\perp_P k \mid C$ (singleton-transitivity),

where i, j, and k are single elements. A semi-graphoid distribution that satisfies intersection is called graphoid. If the distribution P is a regular multivariate Gaussian distribution, then P is a singleton-transitive compositional graphoid; for example, see Studený (2005) and Pearl (1988). If P has strictly positive density, it is always a graphoid; see, for example, Proposition 3.1 in Lauritzen (1996).

Remark 1. If P has full support over its state space, then it satisfies the intersection property; for a comprehensive discussion and necessary and sufficient conditions, see Peters (2015). $\Diamond$

Finally, we define the concept of ordered stabilities (Sadeghi, 2017). We say that P satisfies ordered upward- and downward-stability with respect to an order $\leq$ of V if the following hold:

- if $i \perp\!\!\!\perp_P j \mid C$, then $i \perp\!\!\!\perp_P j \mid C \cup \{k\}$ for every $k \in V \setminus \{i, j\}$ such that $i < k$ or $j < k$ (ordered upward-stability);
- if $i \perp\!\!\!\perp_P j \mid C$, then $i \perp\!\!\!\perp_P j \mid C \setminus \{k\}$ for every $k \in V \setminus \{i, j\}$ such that $i \nless k$, $j \nless k$, and $\ell \nless k$ for every $\ell \in C \setminus \{k\}$ (ordered downward-stability).

2·3. *Graphs and their properties*

We usually refer to a graph as an ordered pair $G = (V, E)$, where V is the node set and E is the edge set. When nodes i and j are the endpoints of an edge, we call them adjacent, and write $i \sim j$, and otherwise $i \nsim j$.

We consider two types of edges: arrows ($i \longrightarrow j$) and bidirected edges or arcs ($i \longleftrightarrow j$). We do not consider graphs that have simultaneous third type of edge: undirected edges or lines (i —— j). We only allow for the possibility of multiple edges between nodes when they are arrows in two different directions between i and j, i.e., $i \rightarrow j$ and $i \leftarrow j$, which we call *parallel arrows*. This means that, except in Section S6, we do not allow bows, i.e., a multiple edge of arrow and arc, to appear in the graph.

A subgraph of a graph G_1 is graph G_2 such that $V(G_2) \subseteq V(G_1)$ and $E(G_2) \subseteq E(G_1)$ and the assignment of endpoints to edges in G_2 is the same as in G_1. An induced subgraph by nodes $A \subseteq V$ is the subgraph that contains all and only nodes in A and all edges between two nodes in A.

A walk is a list $\langle v_0, e_1, v_1, \ldots, e_k, v_k \rangle$ of nodes and edges such that for $1 \leq i \leq k$, the edge e_i has endpoints v_{i-1} and v_i. A path is a walk with no repeated node or edge. When we define a path, we only write the nodes and not the edges. A maximal set of nodes in a graph whose members are connected by some paths constitutes a connected component of the graph. A cycle is a walk with no repeated nodes or edges except for $v_0 = v_k$.

We call the first and the last nodes endpoints of the path and all other nodes inner nodes. A path can also be seen as a certain type of connected subgraph of G; a subpath of a path π is an induced connected subgraph of π. For an arrow $j \rightarrow i$, we say that the arrow is from j to i. We also call j a parent of i, i a child of j and we use the notation $\mathrm{pa}(i)$ for the set of all parents of i in the graph. In the cases of $i \rightarrow j$ or $i \leftrightarrow j$ we say that there is an arrowhead at j or pointing to j. A path $\langle i = i_0, i_1, \ldots, i_n = j \rangle$, or a cycle where $i = j$, is directed from i to j if all $i_k i_{k+1}$ edges are arrows pointing from i_k to i_{k+1}. If there is a directed path from i to j, then node i is an ancestor of j and j is a descendant of i. We denote the set of ancestors of j by $\mathrm{an}(j)$; unlike some authors, we do not allow $j \in \mathrm{an}(j)$. Similarly, we define an ancestor of a set of nodes $A \subset V$ given by $\mathrm{an}(A) := \left(\bigcup_{j \in A} \mathrm{an}(j)\right) \setminus A$. If necessary, we might write an_G to specify that this is the set of ancestors in G.

A *strongly connected component* of a graph is the set of nodes that are mutually ancestors of each other, or it is a single node if that node does not belong to any directed cycle. It can be observed that nodes of the graph are partitioned into strongly connected components. We denote the members of the strongly connected component containing node i by $\mathrm{sc}(i)$.

A *tripath* is a path with three nodes. The inner node t in each of the three tripaths

$$i \rightarrow t \leftarrow j,\ i \leftrightarrow t \leftarrow j,\ i \leftrightarrow t \leftrightarrow j$$

is a collider, or a collider node, and the inner node of any other tripath is a non-collider, or a non-collider node, on the tripath or, more generally, on any path of which the tripath is a subpath; i.e. a node is a collider if two arrowheads meet. A path is called a *collider* path if all its inner nodes are colliders.

The most general class of graphs that naturally arises from the theory presented in this paper is what we call the *bowless directed mixed graph*, which consists of arrows and arcs, and the only multiple edges are parallel arrows; recall that a bow between two nodes consists of both an arrow and an arc. We will consider briefly how to modify and improve our graphs with bows in Section S6.

Ancestral graphs (Richardson & Spirtes, 2002) are graphs with arcs and arrows with no directed cycles and no arcs ij such that $i \in \mathrm{an}(j)$. Acyclic directed mixed graphs (Richardson, 2003) are graphs with arcs and arrows with no directed cycles. In other words, bowless directed mixed graphs unify acyclic directed mixed graphs without bows and directed cycles. Bowless directed mixed graphs also trivially contain the class of *directed ancestral graphs*, i.e., ancestral graphs (Richardson & Spirtes, 2002) without lines. These all also contain directed acyclic graphs (Kiiveri et al., 1984), which are graphs with only arrows and no directed cycles.

The class of bowless directed mixed graphs is a subclass of directed mixed graphs, introduced in Bongers et al. (2021), which is a very general class of graphs with arrows and arcs that allow for directed cycles. Later on, we will use some definitions and results originally defined for directed mixed graphs.

2·4. *Markov properties*

In this paper, we will need global and pairwise Markov properties for bowless directed mixed graphs. In order to introduce the global Markov property, we need to define the concept of σ-separation for directed mixed graphs. This notion of separation was originally defined for the larger class of directed graphs with hyperedges in Forre & Mooij (2017).

A path $\pi = \langle i = i_0, i_1, \ldots, i_n = j \rangle$ is said to be *σ-connecting given* C, which is disjoint from i, j, if all its collider nodes are in $C \cup \mathrm{an}(C)$, and all its non-collider nodes i_r are either outside C, or if there is an arrowhead at i_{r-1}, then $i_{r-1} \in \mathrm{sc}(i_r)$ and if there is an arrowhead at i_{r+1} on π, then $i_{r+1} \in \mathrm{sc}(i_r)$. For disjoint subsets A, B, and C of V, we say that A and B are σ-separated given C, and write $A \perp_\sigma B \mid C$, if there are no σ-connecting paths between A and B given C.

In the case where there are no directed cycles in the graph, σ-separation reduces to the m-separation of Richardson & Spirtes (2002); recall that π is m-connecting given C if all its collider nodes are in $C \cup \mathrm{an}(C)$, and all its non-collider nodes are outside C. In addition, if there are no arcs in the graph, i.e., the graph is a directed acyclic graph, it reduces to the well-known d-separation (Pearl, 1988).

We call two graphs Markov equivalent if they induce the same set of conditional separations.

A probability distribution P defined over V satisfies the *global Markov property* with respect to a bowless directed mixed graph G, or is simply *Markovian* to G, if for disjoint subsets A, B, and C of V, we have

$$A \perp_\sigma B \mid C \implies A \perp\!\!\!\perp B \mid C. \tag{1}$$

If G is an ancestral graph or a directed acyclic graph, then $\perp_\sigma$ will be replaced by $\perp_m$ or $\perp_d$, respectively, in the definition of the global Markov property.

If, in addition to the global Markov property, the other direction of the implication holds, i.e., $A \perp_\sigma B \mid C \iff A \perp\!\!\!\perp B \mid C$, then we say that P and G are faithful. A weaker condition of adjacency-faithfulness (Ramsey et al., 2006; Zhang & Spirtes, 2008) states that for every edge between k and j in G, there are no independence statements $k \perp\!\!\!\perp_P j \mid C$ for any $C \subseteq V \setminus \{k, j\}$.

A distribution P satisfies the *pairwise Markov property* with respect to a bowless directed mixed graph G, if for every pair of non-adjacent nodes i, j in G, we have

$$i \perp\!\!\!\perp_P j \mid \mathrm{an}(\{i, j\}). \tag{2}$$

This is the same wording as that of the pairwise Markov property for the subclass of ancestral graphs; see Lauritzen & Sadeghi (2018).

We prove the equivalence of the pairwise and global Markov properties under compositional graphoids for maximal bowless directed mixed graphs, which shall be used later for causal graphs and Theorem 5.

THEOREM 1. *Let G be a bowless directed mixed graph, and P satisfy the intersection and composition properties. If P satisfies the pairwise Markov property with respect to G, then P is Markovian to G.*

The proof of Theorem 1, as with all our proofs will appear in the supplementary material.

We also define the converse of the pairwise Markov property. We say that P satisfies the *converse pairwise Markov property* with respect to G if an edge between i and j in G implies that

$$i \not\perp\!\!\!\perp_P j \mid \mathrm{an}(\{i, j\}). \tag{3}$$

Faithfulness and adjacency-faithfulness of $P_{\mathcal{C}}$ and $G_{\mathcal{C}}$ imply the converse pairwise Markov property; see Sadeghi & Soo (2022).

A graph is called maximal if the absence of an edge between i and j corresponds to a conditional separation statement for i and j, i.e. there exists for some C a statement of form $i \perp_{\sigma} j \mid C$. From the definition of σ-separation that graphs with chordless directed cycles, that is, having two non-adjacent nodes in a cycle, are not maximal.

We have the following corresponding converse to Theorem 1.

PROPOSITION 1. *Let G be a maximal bowless directed mixed graph. If P is Markovian to G, then P satisfies the pairwise Markov property with respect to G.*

We call a non-adjacent pair of nodes which cannot be σ-separated in a non-maximal graph, regardless of what to condition on, an *inseparable* pair, and a non-adjacent pair of nodes which can be σ-separated in a maximal or non-maximal graph a *separable* pair.

We define a *primitive inducing path* to be a path $\langle i, q_1, \cdots, q_r, j \rangle$, with at least three nodes, between i and j, where

(i) all edges $q_m q_{m+1}$ are either arcs or an arrow where $q_m \in \mathrm{sc}(q_{m+1})$ except for the first and last edges, which may be $i \rightarrow q_1$ or $q_r \leftarrow j$, without being in the same connected component;
(ii) for all inner nodes, we have $q_m \in \mathrm{an}(\{i, j\})$, i.e., they are in ancestors of i or j.

Primitive inducing paths were originally defined for the case of ancestral graphs, where they were allowed to be an edge (Richardson & Spirtes, 2002). We show in the supplementary material, the result below:

PROPOSITION 2. *In a bowless directed mixed graph, inseparable pairs are connected by primitive inducing paths.*

In addition, from Proposition 1, it follows that if P is Markovian to G, then a pair i, j being a separable pair is equivalent to the separation $i \perp_{\sigma} j \mid \mathrm{an}\{i, j\}$.

We say that a graph $G = (V, E)$ admits a valid order $\leq$ if for nodes i and j of G, $i \rightarrow j$ implies that $i > j$, and $i \leftrightarrow j$ implies that i and j are incomparable. This specifies the partial order via its cover relations and this order can is used as the order with respect to which ordered upward- and downward-stability hold for graph separations; see Sadeghi (2017).

Finally, for ancestral graphs, the m-separation holds for every set between parents and ancestors.

LEMMA 1. *For an ancestral graph and for separable nodes i and j, we have*

$$i \perp_{m} j \mid A,$$

for every A such that $\mathrm{pa}(\{i, j\}) \subseteq A \subseteq \mathrm{an}(\{i, j\})$.

We will use Lemma 1 in our analysis of structural causal models.

2·5. *Structural causal models*

The theory in this paper does not use or assume structural causal models, which are also known as the structural equation models (Pearl, 2009; Spirtes et al., 2000). We define structural causal models here as they are an interesting special case of our theory, for which intervention could be easily conceptualized.

Here, we define structural causal models for the class of bowless directed mixed graphs as a simplified version of structural causal models defined for directed mixed graphs in Bongers et al. (2021). Consider a graph G with N nodes, which in the context of causal inference is often referred to as the true causal graph. A structural causal model $\mathcal{C}$ associated with G is defined as a collection of N equations

$$X_i = \phi_i(X_{\mathrm{pa}_G(i)}, \epsilon_i), \quad i \in \{1, \ldots, N\},$$

where $\mathrm{pa}_G(i)$ is defined on G and ϵ_i might be called noises; for any subsets A and B, we require that $\epsilon_A \perp\!\!\!\perp \epsilon_B$ if and only if, in G, there is no arc between any node in A and any node in B. In this paper, we usually refer to a structural causal model as $\mathcal{C}$ and its joint distribution as $P_{\mathcal{C}}$.

In the more widely-used case where G is a directed acyclic graph, all the ϵ_i are jointly independent. For both mathematical and causal discussions on structural causal models with directed acyclic graphs, see Peters et al. (2017). When directed cycles are existent, some solvability conditions are required in order for the theory of structural causal models to work properly; for this and for more general discussion, see Bongers et al. (2021).

Standard interventions are defined quite naturally when functional equations are specified, as in the case of structural causal models: By intervening on X_i we replace the equation associated to X_i by $X_i = \tilde{X}_i$, where $\tilde{X}_i$ is independent of all other noises; it is not necessary that $\tilde{X}_i$ has the same distribution as X_i. We are concerned with a similar type of intervention in this paper—this is a special case of the so-called stochastic intervention (Korb et al., 2004), where some parental set of X_i might still exist after intervention on X_i; see also Peters et al. (2017). A more special type of intervention is called perfect intervention or surgical intervention, where it puts a point mass on a real value a—this is the original idea of do-calculus (Pearl, 2009), and is often denoted by $\mathrm{do}(X_i = a)$.

An important result for structural causal models, which facilitates causal inference using them, is that the joint distribution of a structural causal model is Markovian to its associated graph; see Verma & Pearl (1988); Pearl (2009) for the case of directed acyclic graphs, Sadeghi & Soo (2022) for directed ancestral graphs, and Bongers et al. (2021) for directed mixed graphs.

3. Interventional family of distributions

3·1. *Interventional families and the cause*

Again, let V be a finite set of size N. Consider a family of distributions $\mathcal{P}_{\mathrm{do}} = \{P_{\mathrm{do}(i)}\}_{i \in V}$, where each $P_{\mathrm{do}(i)}$ is defined over the same state space $\mathcal{X} = \prod_{i \in V} \mathcal{X}_i$. We refer to $\mathcal{P}_{\mathrm{do}}$ as an *interventional family* of distributions. For $(\tilde{X}_j)_{j \in V}$ a random vector with distribution $P_{\mathrm{do}(i)}$, we think of $P_{\mathrm{do}(i)}$ as the interventional distribution after intervening on some variable X_i.

For $k \in V$, we define the set of the *causes* of k as

$$\mathrm{cause}^{\mathcal{P}_{\mathrm{do}}}(k) = \mathrm{cause}(k) := \{i : \ i \neq k, \ i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k\};$$

we rarely, simultaneously, have to consider two different interventional families at once. Thus, if $i \in \mathrm{cause}(k)$, then, for $(\tilde{X}_j)_{j \in V}$, we have that $\tilde{X}_k$ is dependent on $\tilde{X}_i$. By convention, $k \notin \mathrm{cause}(k)$. For a subset $A \subseteq V$, we define $\mathrm{cause}(A) = \left(\bigcup_{k \in A} \mathrm{cause}(k)\right) \setminus A$.

The definition of the cause is identical to what is known in the literature as the existence of the total causal effect (Peters et al., 2017, Definition 6.12). Its combination with $\mathcal{P}_{\mathrm{do}}$ meets the intuition behind cause and intervention: after intervention on a variable X_i, it is dependent on a variable X_j if and only if it is a cause of that variable.

The above setting is compatible with the well-known intervention for structural causal models; for a comprehensive discussion on this, see Section 6. We will often illustrate our theory with simple examples of structural causal models and standard intervention on a single node.

We can also define the set of *effects* of i denoted by $\mathrm{eff}(i)$ by $i \in \mathrm{cause}(k) \iff k \in \mathrm{eff}(i)$. We take note of the following useful fact.

Remark 2. From the definition of cause, we have $\mathrm{eff}(i) = \{k : i \neq k,\ i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k\}$. ♢

Remark 3 *(Interventional families with the same cause).* Interventional families are defined for one joint distribution per intervention. This can be considered an advantage as not all interventions have to follow a single causal graph. Here, we provide an immediate condition for interventional families to define the same set of causes. In Section 5·3, we provide conditions for families of distributions that lead to the same causal graph.

Consider the interventional families $\mathcal{P}_{\mathrm{do}} = \{P_{\mathrm{do}(j)}\}_{j\in V}$ and $\mathcal{Q}_{\mathrm{do}} = \{Q_{\mathrm{do}(j)}\}_{j\in V}$ over the same state space $\mathcal{X}$. Causes and effects depend on the interventional family, and, by Remark 2, it follows that, for all $i, k \in V$, we have

$$\left(i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \iff i \not\perp\!\!\!\perp_{Q_{\mathrm{do}(i)}} k\right) \text{ if and only if } \mathrm{cause}^{\mathcal{P}_{\mathrm{do}}}(\ell) = \mathrm{cause}^{\mathcal{Q}_{\mathrm{do}}}(\ell),$$

for all $\ell \in V$, in which case we say they have the same causes. In the case of structural causal models, under standard interventions, the causes are usually invariant with respect to the choice of distribution, except for technical counterexamples; see Remark 11 and Example S1 in the supplementary material.

Consider a fixed $i \in V$, and suppose that measures $P_{\mathrm{do}(i)}$ and $Q_{\mathrm{do}(i)}$ have the same null sets. Notice that the equality

$$Q_{\mathrm{do}(i)}(\cdot \mid x_i) = P_{\mathrm{do}(i)}(\cdot \mid x_i)$$

for almost every $x_i \in \mathcal{X}_i$, is sufficient for the effects of i to be same in both families. Moreover, with the disintegration

$$dP_{\mathrm{do}(i)}(x) = dP_{\mathrm{do}(i)}(x_{V\setminus\{i\}} \mid x_i)dP^i_{\mathrm{do}(i)}(x_i)$$

we see that effects of i depend only on the corresponding conditional distribution, and are invariant under the marginal distributions on i with the same null sets. ♢

Remark 4 *(Atomic interventions).* Observe that the dependence

$$i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k$$

is equivalent to the existence of disjoint measurable subsets $W^*, W^{**} \subset \mathcal{X}_i$ of positive measures under $P^i_{\mathrm{do}(i)}$ satisfying the inequality

$$P^k_{\mathrm{do}(i)}(\cdot \mid x_i^*) \neq P^k_{\mathrm{do}(i)}(\cdot \mid x_i^{**}), \tag{4}$$

for all $x_i^* \in W^*$ and all $x_i^{**} \in W^{**}$. Since $P_{\mathrm{do}(i)}(\cdot \mid x_i^*)$ and $P_{\mathrm{do}(i)}(\cdot \mid x_i^{**})$ are probability measures on $\mathcal{X}_{V\setminus\{i\}}$, they can be thought of as atomic interventions on i, where the values at i are fixed, at x_i^* and x_i^{**}, respectively. Thus inequality (4) has the interpretation that i is a cause of k if and only if there exists atomic interventions that witness an effect on k; that is, as a function of x_i, the conditional probability, $P^k_{\mathrm{do}(i)}(\cdot \mid x_i)$, is non-constant.

From Remark 3, without loss of generality, given atomic interventions, $A_{\mathrm{do}(x_i)}$, which are measures on $\mathcal{X}_{V\setminus\{i\}}$ indexed by $x_i \in \mathcal{X}_i$, we can extend these to an intervention, defined on the complete space, $\mathcal{X}$, via the disintegration

$$dP_{\mathrm{do}(i)}(x) := dA_{\mathrm{do}(x_i)}(x_{V\setminus\{i\}})dR(x_i),$$

where R is a suitably chosen probability measure on $\mathcal{X}_i$ Specifically, in the case where $\mathcal{X}_i$ is finite, R can be taken to be a uniform measure on $\mathcal{X}_i$. ◊

We call a subset $S \subseteq V$ a *causal cycle* if for every $i, k \in S$, we have $k \in \mathrm{cause}(i) \cap \mathrm{eff}(i)$. We write $\mathrm{cc}(i)$ to denote the causal cycle containing i.

Under the composition property, we have the following independence. Let $\mathrm{neff}(i) := V \setminus \mathrm{eff}(i)$ denote the subset of V that contains members that are not an effect of i. 400

PROPOSITION 3 (NON-EFFECTS UNDER COMPOSITION). *Let $\mathcal{P}_{\mathrm{do}}$ be a family of distributions. If $P_{\mathrm{do}(i)}$ satisfies the composition property, then*

$$i \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} \mathrm{neff}(i).$$

3·2. *Transitive interventional families*

We now say that $\mathcal{P}_{\mathrm{do}}$ is a *transitive interventional* family if the following axiom holds.

AXIOM 1 (TRANSITIVITY OF CAUSE). For distinct $i, j, k \in V$, if $i \in \mathrm{cause}(j)$ and $j \in \mathrm{cause}(k)$, then $i \in \mathrm{cause}(k)$. 405

The distributions in transitive interventional families are restricted by Axiom 1, which places constraints between different $P_{\mathrm{do}(i)}$, since $\mathrm{cause}(k)$ depends on all $P_{\mathrm{do}(i)}$.

Under singleton-transitivity, we have sufficient conditions for Axiom 1 to hold. These conditions are not satisfied in general, even in the case of a structural causal model with standard interventions. 410

THEOREM 2 (TRANSITIVITY OF CAUSE UNDER SINGLETON-TRANSITIVITY). *Let $\mathcal{P}_{\mathrm{do}}$ be an interventional family whose members $P_{\mathrm{do}(i)}$ satisfy singleton-transitivity. Assume, for distinct $i, j, k \in V$ such that $i \notin \mathrm{cause}(k)$ and $j \in \mathrm{cause}(k)$, we have:*

(a) $i \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid j$; *and*
(b) $j \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k$. 415

Then $\mathcal{P}_{\mathrm{do}}$ is a transitive interventional family.

PROPOSITION 4. *Transitivity of the cause induces a strict preordering $\lesssim$ on V by*

$$\begin{cases} i < k \iff k \in \mathrm{cause}(i) \text{ and } k \notin \mathrm{cause}(i); \\ i \sim k \iff k \in \mathrm{cause}(i) \cap \mathrm{eff}(i). \end{cases} \tag{5}$$

COROLLARY 1. *Let $\mathcal{P}_{\mathrm{do}}$ be a transitive interventional family that satisfies the composition property. If $i \notin \mathrm{cause}(k)$, then*

$$i \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \mathrm{cause}(k).$$

In the next example, we show that we cannot drop the singleton transitivity assumption in Theorem 2.

Example 1 *(Failure of transitivity without singleton transitivity).* Suppose 1 is the cause of 2, and 2 is a cause of 3. It may not be the case that 1 is a cause of 3. Consider the structural causal model, with $X_1 \longrightarrow X_2 \longrightarrow X_3$, where X_1 is Bernoulli $p \in (0, 1)$, conditional on $X_1 = x_1$, we sample a Poisson random variable $N = n$ with mean $x_1 + 1$, and then we sample $X_2 = (X_2^0, \dots, X_2^n)$ as an i.i.d. sequence of $n + 1$ Bernoulli random variable(s) with parameter $1/2$, and finally set $X_3 := X_2^0$. The first random variable X_1 is independent of the final result X_3. The standard interventions where we simply substitute a distributional copy of X_i for each i gives that 420 425

1 is clearly the cause of 2, and 2 a cause of 3. However, since $P_{\mathrm{do}(1)} = P$, singleton-transitivity fails: $1 \perp\!\!\!\perp_P 3$ and $1 \perp\!\!\!\perp_P 3 \mid 2$, but we have neither $1 \perp\!\!\!\perp_P 2$ nor $3 \perp\!\!\!\perp_P 2$. ◊

4. CAUSAL GRAPHS AND INTERVENED MARKOV PROPERTIES

4·1. *Intervened and direct cause*

Notice that only knowing the causal ordering cannot yield a graph, since, for example, for $i \longrightarrow j \longrightarrow k$, there is no way to distinguish the two graphs corresponding to whether an additional $i \longrightarrow k$ exists in the graph or not. For this reason, we need to define the concept of direct cause.

In order to define direct and intervened cause in the general case, we need an iterative procedure:

1. For each $i, k \in V$, start with $\mathrm{dcause}(k) := \mathrm{cause}(k)$, $\iota\mathrm{cause}_i(k) := \mathrm{cause}(k)$, and $S(\mathcal{P}_{\mathrm{do}})$ an empty graph with node set V;
2. Redefine
$$\mathrm{dcause}(k) := \{i : \; i \in \mathrm{dcause}(k), \; i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(k) \setminus \{i\}\};$$
3. Generate $S(\mathcal{P}_{\mathrm{do}})$ by setting arrows from i to k, i.e., $i \longrightarrow k$ if $i \in \mathrm{dcause}(k)$;
4. Generate the graph $S_i(\mathcal{P}_{\mathrm{do}})$ by removing all arrows pointing to i from $S(\mathcal{P}_{\mathrm{do}})$;
5. Redefine $\iota\mathrm{cause}_i(k) := \mathrm{an}_{S_i(\mathcal{P}_{\mathrm{do}})}(k)$;
6. If $S(\mathcal{P}_{\mathrm{do}})$ is modified by Step 3, then go to Step 2; otherwise, output $\mathrm{dcause}(k)$, $\iota\mathrm{cause}_i(k)$, $S(\mathcal{P}_{\mathrm{do}})$, and $S_i(\mathcal{P}_{\mathrm{do}})$.

We call $\mathrm{dcause}(k)$ the set of the *direct causes* of k, and $\iota\mathrm{cause}_i(k)$ the set of *intervened causes* of k after intervention on i. We also call $S(\mathcal{P}_{\mathrm{do}})$ the *causal structure*, and $S_i(\mathcal{P}_{\mathrm{do}})$ the *i-intervened causal structure*.

In the iteration $\mathrm{dcause}(k)$ is getting smaller, which ensures that the procedure will stop.

By convention, $k \notin \mathrm{dcause}(k) \cup \iota\mathrm{cause}_i(k)$, for any $i \in V$. We also let $\mathrm{dcause}(A) = (\bigcup_{k \in A} \mathrm{dcause}(k)) \setminus A$, and $\iota\mathrm{cause}_i(A) = (\bigcup_{k \in A} \iota\mathrm{cause}_i(k)) \setminus A$.

By definition, cause is a universal concept: no matter how large the system of random variables is, as long as it contains the two investigated random variables, the marginal dependence of those variables stays intact. On the other hand, direct cause depends on the system of variables in which the two investigated variables lie.

The below example shows why we need $\iota\mathrm{cause}_i(k)$ as opposed to $\mathrm{cause}(k)$ in the definition of $\mathrm{dcause}(k)$; see also Section S3 in the supplementary material.

Example 2. Let the graph of Figure 1, below, be the graph associated to a structural causal model with standard interventions. Under faithfulness, notice that, $i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \mathrm{cause}(k) \setminus \{i\}$, where $\mathrm{cause}(k) \setminus \{i\} = \{j, \ell\}$. However, i is clearly not a direct cause of k. On the other hand, $\iota\mathrm{cause}_i(k) \setminus \{i\} = \{j\}$, and $i \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(k) \setminus \{i\}$.

In the iterative procedure above, in the first round, there will be an arrow from i to k in $S(\mathcal{P}_{\mathrm{do}})$. In the second round, i will be removed. ◊

Remark 5. For causal graphs, as defined in the next subsection, that are ancestral, $\mathrm{dcause}(k)$ can simply be defined by

$$i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \mathrm{cause}(k) \setminus \{i\}$$

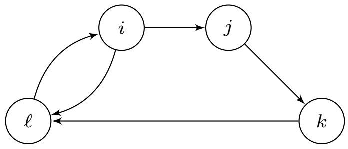

Fig. 1. A graph for which detecting the direct cause requires an iterative procedure.

rather than using the iterative procedure and $\iota\mathrm{cause}_i(k)$ in the conditioning set; see Section S3 in the supplementary material for the equivalence of the two methods under certain conditions. $\diamondsuit$

Remark 6. As it is seen in Section 7, there may still be arrows generated here that arguably should not be considered direct causes in the case of non-maximal non-ancestral graphs. We study and identify these cases and offer adjustments in that section. $\diamondsuit$

The next proposition can be thought of as a proxy for the pairwise Markov property, and will be used in our proofs of the Markov property for our casual graphs.

PROPOSITION 5. *Let $\mathcal{P}_{\mathrm{do}}$ be a transitive interventional family, and $P_{\mathrm{do}(i)}$ satisfy the composition property. For distinct $i, k \in V$, if $i \notin \mathrm{dcause}(k)$, then*

$$i \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(k) \setminus \{i\}.$$

4·2. *Intervened causal and causal graphs*

We are now ready to define a graph that demonstrates the causal relationships by capturing the direct causes as well as non-causal dependencies due to latent variables.

Given an interventional family $\mathcal{P}_{\mathrm{do}}$, and $i \in V$, we define the *i-intervened graph*, denoted by $G_i(\mathcal{P}_{\mathrm{do}})$ to be the i-intervened causal structure, where, in addition, for each pair of non-adjacent nodes $j, k \in V$ that are distinct from i, we place an arc between j and k, i.e., $j \leftarrow\!\!\!\!\succ k$, if

$$j \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(\{j,k\}), \tag{6}$$

Thus with (6), we put an arc if one of the interventions $P_{\mathrm{do}(i)}$ suggests the presence of a latent variable.

We also define the *causal graph*, denoted by $G(\mathcal{P}_{\mathrm{do}})$, to be the causal structure, where, in addition, for each pair of nodes j, k that are not adjacent by an arrow, we place an arc between them if the jk-arc exists in $G_i(\mathcal{P}_{\mathrm{do}})$ for every $i \in V$ that is distinct from j and k.

Remark 7. The graph $G_i(\mathcal{P}_{\mathrm{do}})$ does not contain arrows pointing to i in $G(\mathcal{P}_{\mathrm{do}})$ and all arcs with i as an endpoint in $G(\mathcal{P}_{\mathrm{do}})$.

We also note that the existence of the jk-arc in two different intervened graphs $G_i(\mathcal{P}_{\mathrm{do}})$ and $G_\ell(\mathcal{P}_{\mathrm{do}})$ may not coincide when there is a primitive inducing path as the below example shows. $\diamondsuit$

Example 3. Consider a structural causal model with the graph presented in Figure 2, below, with standard intervention. Assume that the joint distribution of the structural causal model is faithful to this graph. It is easy to observe that $j \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(\{j,k\})$, but $j \perp\!\!\!\perp_{P_{\mathrm{do}(\ell)}} k \mid \iota\mathrm{cause}_i(\{j,k\})$. This implies that there is a jk-arc in $G_i(\mathcal{P}_{\mathrm{do}})$, but not in $G_\ell(\mathcal{P}_{\mathrm{do}})$. $\diamondsuit$

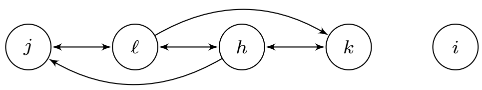

Fig. 2. A non-maximal graph associated to a structural causal model.

For an interventional family $\mathcal{P}_{\mathrm{do}}$, our definitions now allow us to use the notions $\mathrm{pa}_{G(\mathcal{P}_{\mathrm{do}})}(k)$ and $\mathrm{dcause}(k)$ interchangeably. For a transitive $\mathcal{P}_{\mathrm{do}}$, we have that, moreover, other causal terminologies on $\mathcal{P}_{\mathrm{do}}$ and the graph terminologies on $G(\mathcal{P}_{\mathrm{do}})$ can be used interchangeably:

THEOREM 3 (INTERCHANGEABLE TERMINOLOGY). *For a transitive interventional $\mathcal{P}_{\mathrm{do}}$ where $P_{\mathrm{do}(i)}$ satisfy the composition property, we have the following:*

(i) $i \in \mathrm{an}_{G(\mathcal{P}_{\mathrm{do}})}(k) \iff i \in \mathrm{cause}(k)$ *in* $\mathcal{P}_{\mathrm{do}}$.
(ii) $i \in \mathrm{sc}_{G(\mathcal{P}_{\mathrm{do}})}(k) \iff k \in \mathrm{cc}(i)$ *in* $\mathcal{P}_{\mathrm{do}}$.

As an immediate consequence of the above result, we see that if the set of causes of a variable is non-empty, then at least one of the causes must act as the direct cause.

COROLLARY 2. *Let $\mathcal{P}_{\mathrm{do}}$ be a transitive interventional family where $P_{\mathrm{do}(i)}$ satisfy the composition property. For $k \in V$, if* $\mathrm{cause}(k) \neq \emptyset$*, then* $\mathrm{dcause}(k) \neq \emptyset$.

In the next example, we illustrate how all the different edges can easily occur. We remark that we can generate cycles quite easily, but in the standard structural causal model setting, their existence is non-trivial and requires solvability conditions (Bongers et al., 2021).

Example 4 *(A simple example of an arrow, cycle, and arc).* Consider a fixed joint distribution P for random variables (X_1, X_2, X_3), where X_1 is not independent of X_2, and X_3 is independent of (X_1, X_2) We may think of the joint distribution (X_1, X_2) as generated via the following functional equations: $(X_1, X_2 = \phi(X_1, U))$ or $(X_1 = \psi(X_2, V), X_2)$, where ϕ and ψ are deterministic functions, U and V are uniformly distributed on $[0, 1]$.

Thus corresponding to ϕ, we have a structural causal model, where $X_1 \rightarrowtail X_2$, and X_3 is isolated, and similarly, corresponding to ψ, we have a structural causal model, where $X_2 \rightarrowtail X_1$, and X_3 is isolated. Next, we see how these structural causal model interact with various interventions, and see the corresponding causal graphs that can be defined.

Via the function ϕ, we consider standard interventions, where X_1, X_2, and X_3 are replaced with a distributional copies of themselves, giving the family $(P = P^{\phi}_{\mathrm{do}(1)}, P^{\phi}_{\mathrm{do}(2)}, P = P^{\phi}_{\mathrm{do}(3)})$ and via the function ψ we consider standard interventions giving the family $(P^{\psi}_{\mathrm{do}(1)}, P = P^{\psi}_{\mathrm{do}(2)}, P = P^{\psi}_{\mathrm{do}(3)})$; in both cases, we place an arrow between 1 and 2 in the expected direction. Note that

$$P^{\mathrm{ind}} := P^1 \otimes P^2 \otimes P^3 = P^{\phi}_{\mathrm{do}(2)} = P^{\psi}_{\mathrm{do}(1)}.$$

Consider also the non-standard interventional family $(P^{\phi}_{\mathrm{do}(1)}, P^{\psi}_{\mathrm{do}(2)}, P)$; 1 is a direct cause of 2, and 2 is also a direct cause of 1, so we obtain a parallel edge.

Finally, to have an arc, we consider the interventional family

$$(P_{\mathrm{do}(1)}, P_{\mathrm{do}(2)}, P_{\mathrm{do}(3)}) = (P^{\mathrm{ind}}, P^{\mathrm{ind}}, P);$$

there are no causes, and no arrows are placed, but $P_{\mathrm{do}(3)}$ detects the dependence in X_1 and X_2, and places an arc between them. ♢

In the following example, we stress that the generated causal graph heavily depends on the interventional family that is considered, even under standard interventions arising from simple structural causal model.

Example 5 *(Two different graphs, one underlying distribution).* Let $\epsilon_1, \epsilon_2, \epsilon_3$ be independent Bernoulli random variables with parameter $1/2$. Consider the random variables $X_1 = \epsilon_1$, $X_2 = X_1 + \epsilon_2$, and $X_3 = X_1 + \epsilon_3$; they have a joint distribution P, and can be thought of as a structural causal model $\mathcal{C}$ with $X_1 \longrightarrow X_2$ and $X_1 \longrightarrow X_3$. Thus we can define a intervention family $\mathcal{P}_{\mathrm{do}}$ corresponding to standard interventions on $\mathcal{C}$; it is not difficult to verify in this case that the resulting casual graph will be the same as the graph for $\mathcal{C}$.

However, it is not difficult to construct another structural causal model $\mathcal{C}'$ with $X_3' \longrightarrow X_2'$, $X_3' \longrightarrow X_1'$, and $X_2' \longrightarrow X_1'$, where (X_1', X_2', X_3') also have the joint distribution P. Thus we can define another intervention family $\mathcal{P}'_{\mathrm{do}}$ corresponding to standard interventions on $\mathcal{C}'$; again it is not difficult to verify in this case that the resulting casual graph will be the same as the graph for $\mathcal{C}'$.

We exploited the simple fact that that the joint distribution P does not uniquely determine a corresponding structural causal model—not even up to adjacency. Notice that $\mathcal{C}$ and $\mathcal{C}'$ do not have the same number of edges. ♢

The generated graph is indeed a bowless directed mixed graph.

PROPOSITION 6 (THE CAUSAL GRAPH IS BOWLESS DIRECTED MIXED GRAPH). *The causal graph and intervened graphs generated from a transitive interventional family are bowless directed mixed graphs.*

Remark 8. By just knowing the family $\mathcal{P}_{\mathrm{do}}$, in most situations, there is no way to distinguish an arrow and a bow. For example, in the case of only two nodes i and j, dependence in either of the interventional distributions indicates a direct cause without specifying the existence of a latent variable causing both nodes; see also Example S4. This reflects in our definition on causal graphs where the existence of an arc between two nodes is tested only when there is no arrow between them. The exception is for non-maximal graphs. We detail this in Section S6, where we also discuss identifying bows when, in addition, there exist observational distributions. ♢

Remark 9. If we assume that $\mathrm{cc}(k) = \{k\}$ for every $k \in V$, so that there is no causal cycle, then the causal graph is a bowless acyclic directed mixed graph. If we assume that for every $j, k \in V$ that are not direct causes of each other and every i, we have

$$j \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(\{j, k\}),$$

then the causal graph does not contain arcs—in this case it is seen that the interventional family does not detect any latent variables that cause both j and k, since they are not dependent. Assuming both of these conditions results in a directed acyclic graph. ♢

4·3. *The Markov property with respect to the intervened causal graphs*

We can now present a main result of this paper.

THEOREM 4. (INTERVENTIONAL DISTRIBUTIONS ARE MARKOVIAN TO INTERVENED GRAPHS). *Let $\mathcal{P}_{\mathrm{do}} = \left\{P_{\mathrm{do}(i)}\right\}_{i \in V}$ be a transitive interventional family. For each $i \in V$, if $P_{\mathrm{do}(i)}$ satisfies the intersection property and the composition property, then $P_{\mathrm{do}(i)}$ is Markovian to the i-intervened graph, $G_i(\mathcal{P}_{\mathrm{do}})$.*

5. OBSERVABLE INTERVENTIONAL FAMILIES

5·1. *The Markov property with respect to the causal graph*

In Section 4, the causal structures are completely defined by interventional families; here we provide additional assumptions for the causal graph to be Markovian with respect to an underlying distribution that can be observed.

We say that an interventional family of distributions $\mathcal{P}_{\rm do}$ on the state space $\mathcal{X} = \prod_{i \in V} \mathcal{X}_i$ is *observable* with respect to an *underlying* distribution P on $\mathcal{X}$ such that the following axiom holds.

AXIOM 2.

(a) For every separable pair $j, k \in V$, in $G(\mathcal{P}_{\rm do})$, and every distinct $i \in V$, we have

$$j \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(\{j,k\}) \Rightarrow j \perp\!\!\!\perp_P k \mid \mathrm{cause}(\{j,k\}).$$

(b) For every k and every $i \in \mathrm{cause}(k)$, we have

$$i \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(\{i,k\}) \Rightarrow i \perp\!\!\!\perp_P k \mid \mathrm{cause}(\{i,k\}).$$

We remark that if P is a product measure on $\mathcal{X}$, then every interventional family will be observable with respect to P; in practice we want to consider underlying distributions that have some relation to the interventional family, such as when P is the distribution of a structural causal model.

THEOREM 5 (THE UNDERLYING DISTRIBUTION IS MARKOVIAN TO THE CAUSAL GRAPH). *Let $\mathcal{P}_{\rm do}$ be a transitive observable interventional family, with underlying joint distribution P that satisfies the intersection property and the composition property. Then P is Markovian to the causal graph, $G(\mathcal{P}_{\rm do})$.*

5·2. *Causal graphs for observational distributions*

The arcs in the causal graph $G(\mathcal{P}_{\rm do})$ can be generated from the observational distribution P and the causes, if the inverse of Axiom 2 holds. Assume $\mathcal{P}_{\rm do}$ is an observable interventional family of distributions with respect to an underlying observational measure P. We say that $\mathcal{P}_{\rm do}$ is a *strongly-observable interventional* family if the following additional axiom holds.

AXIOM 3.

(a) For every distinct $i, j, k \in V$, we have

$$j \perp\!\!\!\perp_P k \mid \mathrm{cause}(\{j,k\}) \Rightarrow j \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(\{j,k\}).$$

(b) For every k and every $i \in \mathrm{cause}(k)$, we have

$$i \perp\!\!\!\perp_P k \mid \mathrm{cause}(\{i,k\}) \Rightarrow i \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(\{i,k\}).$$

Our next example shows that there are univariate observable interventional families that are not given by standard interventions and that univariate observable interventional families may not satisfy Axiom 3.

Example 6 *(Interventions on joint distributions).* Suppose X_1, X_2, and X_3 are jointly independent (Bernoulli) random variables, with law P, which will serve as an underlying distribution. Now consider the (non-standard) intervention, where $P_{\mathrm{do}(1)}$ changes the joint distribution of X_2 and X_3 to one of *dependence*, but leaves the marginal distributions of X_2 and X_3, and X_1 alone; furthermore, we leave X_1 independent of (X_2, X_3). Although we normally think of

$P_{\mathrm{do}(1)}$ as an interventional on X_1, our general definition allows for somewhat counter-intuitive constructions.

Let $P_{\mathrm{do}(2)}$ and $P_{\mathrm{do}(3)}$ be standard interventions on X_2 and X_3, respectively, that simply leave the original independent distribution unchanged. The family $\{P_{\mathrm{do}(1)}, P_{\mathrm{do}(2)}, P_{\mathrm{do}(3)}\}$ satisfies Axioms 1 and 2 trivially, but Axiom 3 is not satisfied. $\Diamond$

Define $G(P)$ by adding to the causal structure $S(\mathcal{P}_{\mathrm{do}})$, arcs between j and k, i.e., $j \leftrightarrow k$, for nodes j and k not adjacent by an arrow, if

$$j \not\perp\!\!\!\perp_P k \mid \mathrm{cause}(\{j,k\}). \tag{7}$$

Clearly, (7) suggests the presence of a latent variable; compare with (6).

PROPOSITION 7. *Suppose that $\mathcal{P}_{\mathrm{do}}$ is a strongly-observable interventional family with the underlying distribution P, and $N \geq 3$, so that there are at least three nodes. Then $G(P) = G(\mathcal{P}_{\mathrm{do}})$ when these graphs are maximal.*

Note that for strongly-observable interventional families, the existence of the jk-arc in $G(\mathcal{P}_{\mathrm{do}})$ and $G(\mathcal{P})$ may differ for non-maximal graphs only when j, k is an inseparable pair. This implies that these two graphs are Markov equivalent.

5·3. *Congruent interventional families*

Consider the interventional families $\mathcal{P}_{\mathrm{do}} = \{P_{\mathrm{do}(j)}\}_{j\in V}$ and $\mathcal{Q}_{\mathrm{do}} = \{Q_{\mathrm{do}(j)}\}_{j\in V}$ over the same state space $\mathcal{X}$. It is immediate that if both families are strongly observable with respect to a single underlying distribution P, and if the causal graphs $G(\mathcal{P}_{\mathrm{do}})$ and $G(\mathcal{Q}_{\mathrm{do}})$ are maximal, then they have the same adjacencies. Motivated by Axioms 2 and 3, we say that the families are *congruent* if they have the same causes; see also Remark 3.

1.) For every k and every $i \in \mathrm{cause}^{\mathcal{P}_{\mathrm{do}}}(k) = \mathrm{cause}^{\mathcal{Q}_{\mathrm{do}}}(k)$, we have

$$i \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i^{\mathcal{P}_{\mathrm{do}}}(\{i,k\}) \iff i \perp\!\!\!\perp_{Q_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i^{\mathcal{Q}_{\mathrm{do}}}\{i,k\}.$$

2.) For every distinct $i, j, k \in V$, we have

$$j \perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i^{\mathcal{P}_{\mathrm{do}}}(\{j,k\}) \iff j \perp\!\!\!\perp_{Q_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}^{\mathcal{Q}_{\mathrm{do}}}\{j,k\}.$$

We collect our observations in the following proposition.

PROPOSITION 8. *Consider two interventional families over the same state space.*

1.) *The families are congruent if and only if their causal graphs are the same.*
2.) *If the families are strongly observable with respect to the same underlying distribution, and their causal graphs are maximal, then the graphs have the same adjacencies; furthermore, if the families have the same causes, and if the causal graphs are ancestral, then the graphs are the same, and the families are congruent.*

6. SPECIALIZATION TO STRUCTURAL CAUSAL MODELS

6·1. *Introduction*

In this section, we relate the standard intervention on structural causal models to the setting presented in this paper.

Let $\mathcal{C}$ be a structural causal model with random vector X taking values on $\mathcal{X} = \prod_{i\in V} \mathcal{X}_i$ with joint distribution $P_{\mathcal{C}}$, and associated graph $G_{\mathcal{C}}$. Consider again the standard intervention in structural causal models, where intervention on $i \in V$ replaces the equation $X_i = \phi_i(X_{\mathrm{pa}_G(i)}, \epsilon_i)$ by

$X_i = \tilde{X}_i$, where $\tilde{X}_i$ is independent of all other noises. In the setting of this paper, the new system of equations after intervening on i yields the joint distribution $P_{\mathcal{C}[\mathrm{do}(i)=\tilde{X}_i]}$, and consequently one obtains the family of distributions $\mathcal{P}_{\mathcal{C}[\mathrm{do}=\tilde{X}]} := \{P_{\mathcal{C}[\mathrm{do}=\tilde{X}_1]}, \ldots, P_{\mathcal{C}[\mathrm{do}=\tilde{X}_N]}\}$.

Remark 10. The definition of the set $\mathrm{cause}(i)$, defined for $\mathcal{P}_{\mathcal{C}[\mathrm{do}=\tilde{X}]}$, in the setting of this paper is identical to the definition of the set of "causes" of i in the structural causal model setting (Pearl, 2009; Peters et al., 2017). ♢

Remark 11. Under some weak, but technical assumptions, in the sense of compatibility, we suspect it is possible to show that $\mathrm{cause}(A)$ is invariant under the choice of $\tilde{X}$; see also Proposition 6.13 in Peters et al. (2017), which has technical counterexamples.

Therefore, under invariance, if we are only interested in the causal structure, we can simply refer to some canonical family, where intervention $\tilde{X}_i$ has the same distribution as X_i, which we denote by $\mathcal{P}_{\mathrm{do}}(\mathcal{C}) = \{P_{\mathrm{do}(1)}, \ldots, P_{\mathrm{do}(N)}\}$. ♢

6·2. *Structural causal models and interventional families*

The first question that needs to be addressed is when $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ satisfies different axioms and key assumptions related to interventional families. Cancellations may occur in structural causal models so that cause is not transitive, as required in Axiom 1. We do not provide conditions for structural causal models such that cause is transitive—the main results in this section do not require transitivity of cause. The following example shows that standard interventions on structural causal models do not satisfy the conditions of Theorem 2 nor do they lead to quantifiable interventional families, as defined in the supplementary material.

Example 7. Consider the collider $X_1 \rightarrow X_3 \leftarrow X_2$, where $X_3 = X_1 \oplus X_2 \bmod 2$. Consider the underlying joint distribution $P_{\mathcal{C}}$ where X_1 is Bernoulli with parameter $p_1 = 1/100$ and X_2 is Bernoulli with parameter $p_2 = 1/2$. Consider the standard interventions where $p_1 \to 1/2$ and $p_2 \to 1/100$. Although these are standard interventions, the resulting family does not satisfy Axiom S1 in the supplementary material. Observe that X_2 is a direct cause of X_3, but X_1 is not a cause of X_3. However, $P_{\mathrm{do}(1)}(x_3 = 1 \mid x_2 = 1, x_1 = 1) = 0$ and $P_{\mathcal{C}}(x_3 = 1 \mid x_2 = 1) = P_{\mathcal{C}}(x_1 = 0) = 99/100$. We also see that the conditions of Theorem 2 are not satisfied. ♢

We will need to introduce a concept related to faithfulness on the edge level. We say that $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ satisfies the *edge-cause* condition with respect to $G_{\mathcal{C}}$ if an arrow from i to j in $G_{\mathcal{C}}$ implies that $i \in \mathrm{dcause}(j)$, i.e., $i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} j$ and $i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} j \mid \iota\mathrm{cause}_i(j) \setminus \{i\}$.

In Section S5 in the supplementary material, we will discuss simple conditions on the structural causal model that imply the edge-cause condition. In particular, it is easy to see that if $P_{\mathrm{do}(i)}$ are faithful to the intervened graphs $(G_{\mathcal{C}})_i$, where the upcoming arrows and arcs to i are removed, the edge-cause condition is satisfied. The same can be said with the weaker condition of adjacency-faithfulness of $P_{\mathrm{do}(i)}$ and $(G_{\mathcal{C}})_i$.

PROPOSITION 9 (ANCESTORS AND CAUSES). *For a structural causal model $\mathcal{C}$ and the family $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$, we have that*

$$\mathrm{dcause}(k) \subseteq \mathrm{pa}_{G_{\mathcal{C}}}(k) \textit{ and } \mathrm{cause}(k) \subseteq \mathrm{an}_{G_{\mathcal{C}}}(k), \tag{8}$$

for every $k \in V$. In addition, if $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ satisfies the edge-cause condition with respect to $G_{\mathcal{C}}$ then

$$\mathrm{dcause}(k) = \mathrm{pa}_{G_{\mathcal{C}}}(k) \textit{ and } \mathrm{pa}_{G_{\mathcal{C}}}(k) \subseteq \mathrm{cause}(k). \tag{9}$$

Moreover, $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ is transitive if and only if

$$\mathrm{cause}(k) = \mathrm{an}_{G_{\mathcal{C}}}(k). \tag{10}$$

The following example shows that the inequality may be strict in (8) and (9) does not hold without the edge-cause condition.

Example 8 *(Independence and cause in a structural causal model).* Consider the collider

$$X_1 \rightarrow X_3 \leftarrow X_2,$$

where $X_3 = X_1 \oplus X_2 \bmod 2$. Consider the underlying joint distribution $P_{\mathcal{C}}$ where X_1 is Bernoulli with parameter $p_1 = 1/2$ and X_2 is Bernoulli with parameter $p_2 = 1/2$. Consider the standard interventions where nothing happens: $p_1 \to 1/2$ and $p_2 \to 1/2$. Then X_3 has no causes, and thus its set of ancestors is not equal to its causes. However, it is easy to verify that Axiom S1 in the supplementary material is satisfied. $\diamondsuit$

We also recall that the joint distribution $P_{\mathcal{C}}$ of a structural causal model $\mathcal{C}$ is Markovian to $G_{\mathcal{C}}$. We need a corresponding result to the Markov property of the joint distribution of the structural causal models for the intervened distribution.

LEMMA 2. *Let $\mathcal{C}$ be a structural causal model. For each $i \in V$, its intervened distribution $P_{\mathrm{do}(i)}$ is Markovian to $(G_{\mathcal{C}})_i$ and $G_{\mathcal{C}}$.*

THEOREM 6 (STRONGLY OBSERVABLE STRUCTURAL CAUSAL MODELS). *Let $\mathcal{C}$ be a structural causal model associated to graph $G_{\mathcal{C}}$. Assume that $P_{\mathcal{C}}$ satisfies the converse pairwise Markov property with respect to $G_{\mathcal{C}}$. Also assume that $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ satisfies the edge-cause condition with respect to $G_{\mathcal{C}}$. Then $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ is a strongly observable interventional family if $G_{\mathcal{C}}$ is ancestral. In addition, if $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ is transitive, then the result holds for bowless directed mixed graphs.*

6·3. *Causal graphs and graphs associated to structural causal models*

In this subsection, we present the ultimate relationship between interventions in the structural causal model setting and the setting in this paper, which relates the "true causal graph" $G_{\mathcal{C}}$ with the causal graph $G(\mathcal{P}_{\mathrm{do}})$ defined in this paper. We first need some lemmas.

LEMMA 3. *Let $\mathcal{C}$ be a structural causal model. We have that $\mathrm{pa}_{G(\mathcal{P}_{\mathrm{do}}(\mathcal{C}))}(k) \subseteq \mathrm{pa}_{G_{\mathcal{C}}}(k)$, for every $k \in V$. In addition, if $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ satisfies the edge-cause condition with respect to its associated graph $G_{\mathcal{C}}$, then $\mathrm{pa}_{G(\mathcal{P}_{\mathrm{do}}(\mathcal{C}))}(k) = \mathrm{pa}_{G_{\mathcal{C}}}(k)$.*

To prove a part of the next main result that deals with directed ancestral graphs, we need the following lemma.

LEMMA 4. *Let $\mathcal{C}$ be a structural causal model, and assume that $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ satisfies the edge-cause condition with respect to its associated directed ancestral graph $G_{\mathcal{C}}$. Then for every $i, j \in V$, we have*

$$i \perp\!\!\!\perp_{P_{\mathcal{C}}} j \mid \mathrm{cause}(\{i, j\}) \Rightarrow i \perp\!\!\!\perp_{P_{\mathcal{C}}} j \mid \mathrm{an}_{\mathcal{C}}(\{i, j\}).$$

THEOREM 7 (EQUALITY OF CAUSAL AND STRUCTURAL CAUSAL MODEL GRAPHS). *Consider a structural causal model $\mathcal{C}$ with the joint distribution $P_{\mathcal{C}}$. If $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ satisfies the edge-cause condition, and $P_{\mathcal{C}}$ satisfies the converse pairwise Markov property with respect to the maximal directed ancestral graph $G_{\mathcal{C}}$, then $G(\mathcal{P}_{\mathrm{do}}(\mathcal{C})) = G_{\mathcal{C}}$. In addition, if $\mathcal{P}_{\mathrm{do}}(\mathcal{C})$ is transitive, then the result holds for maximal bowless directed mixed graphs. Moreover, without the maximality assumption, it holds that $G(\mathcal{P}_{\mathrm{do}}(\mathcal{C}))$ and $G_{\mathcal{C}}$ are Markov equivalent.*

The following example shows that we, indeed, require to assume that $G_{\mathcal{C}}$ is maximal.

Example 9. Consider the non-maximal graph of Figure 3 to be $G_{\mathcal{C}}$. Assume standard intervention and also faithfulness of $P_{\mathcal{C}}$ and $G_{\mathcal{C}}$. In $G(\mathcal{P}_{\mathrm{do}}(\mathcal{C}))$, there exists an arc between j and k since no matter what one intervenes on, j and k always stay dependent given any conditioning set. Notice that here we require two discriminating paths between j and k. If there were only one discriminating path between j and k, for example with no h' and ℓ', then by intervening on any node on the discriminating path, such as h, one obtains the required independence $j \perp\!\!\!\perp_{P_{\mathrm{do}(h)}} k \mid \iota\mathrm{cause}_h(\{j, k\})$. ◊

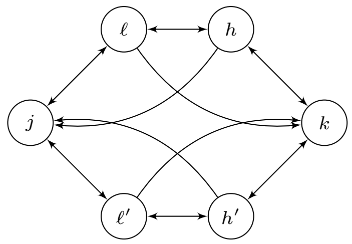

Fig. 3. A non-maximal graph associated to a structural causal model.

COROLLARY 3. *Let $\mathcal{C}$ be a structural causal model. If its joint distribution $P_{\mathcal{C}}$ and its associated bowless directed mixed graph $G_{\mathcal{C}}$ are faithful, and its intervened distribution $P_{\mathrm{do}(i)}$ and its associated intervened graphs $(G_{\mathcal{C}})_i$ are faithful, for every i, then $G(\mathcal{P}_{\mathrm{do}}(\mathcal{C})) = G_{\mathcal{C}}$.*

7. IDENTIFYING CASES THAT NEED EXTRA OR MULTIPLE CONCURRENT INTERVENTIONS

Our theory is based on the interventional family $\mathcal{P}_{\mathrm{do}} = \{P_{\mathrm{do}(i)}\}_{i \in V}$, which only allows single interventions.

We note again that, for ancestral causal graphs, under certain conditions, direct cause can be simply defined using $i \not\!\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \mathrm{cause}(k) \setminus \{i\}$; see Section S3. The following example shows that for the case of structural causal models for non-ancestral graphs, this definition might misidentify some direct causes.

Example 10. In Example 2 and the graph of Figure 1, we observed that an iterative procedure is needed to obtain $i \not\!\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(k) \setminus \{i\}$, which does not coincide with $i \not\!\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \mathrm{cause}(k) \setminus \{i\}$ in this case.

In a structural causal model associated to the below graph with standard intervention, we see that even $i \not\!\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \iota\mathrm{cause}_i(k) \setminus \{i\}$ misidentifies i as the direct cause of k since i has no parents in the graph. We observe that, in this case the independence $i \perp\!\!\!\perp_{P_{\mathrm{do}(j)}} k \mid \mathrm{cause}(k) \setminus \{i\}$ holds. ◊

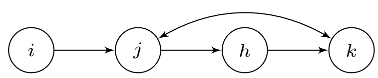

Fig. 4. A non-maximal graph associated to a structural causal model.

Below, we classify the cases for non-ancestral graphs, where direct cause cannot be defined in the way described above.

The result below shows that all cases where $\mathrm{cause}(k)$ is not sufficient to define $\mathrm{dcause}(k)$ result in primitive inducing paths:

PROPOSITION 10. *For a transitive interventional family $\mathcal{P}_{\mathrm{do}}$, assume $P_{\mathrm{do}(i)}$ is Markovian to the i-intervened graph, $G_i(\mathcal{P}_{\mathrm{do}})$, an assumption which can be discharged by satisfying conditions of Theorem 4. For non-adjacent pair of nodes i and k, let $i \in \mathrm{cause}(k) \setminus \mathrm{dcause}(k)$. If $i \not\perp\!\!\!\perp_{P_{\mathrm{do}(i)}} k \mid \mathrm{cause}(k) \setminus \{i\}$, then there is a primitive inducing path between i and k in $G(\mathcal{P}_{\mathrm{do}})$.*

For $i \in \mathrm{cause}(k)$, there could be two types of primitive inducing paths between i and k, described in the above proposition. Notice again that such cases can happen only for non-ancestral graphs and when the true causal graph is non-maximal.

(1) If this primitive inducing path is not a primitive inducing path in $G_i(\mathcal{P}_{\mathrm{do}})$, as in the case of Figure 1, i.e., an inner node of the primitive inducing path is only an ancestor of k via i, then the iterative procedure to define direct cause works as it is designed to ensure that a direct cause is not placed in the causal graph incorrectly.

(2) If the primitive inducing path is a primitive inducing path in $G_i(\mathcal{P}_{\mathrm{do}})$, as in the case in Figure 4, then the current theory is incomplete in the sense that some direct causes may be misidentified as our theory considers $i \in \mathrm{dcause}(k)$, and it always places an arrow from i to k in $G(\mathcal{P}_{\mathrm{do}})$. We have two sub-cases here:

   (a) If there exists maximum one such primitive inducing path between each pair of nodes, then we propose the following adjustment to the definition of direct cause, which fixes this issue:
   We define i to be a direct cause of k if for every j, that may be i but not k, it holds that $i \not\perp\!\!\!\perp_{P_{\mathrm{do}(j)}} k \mid \mathrm{cause}(k) \setminus \{i\}$.
   Hence, as a procedure, one can generate the intervened causal structure $S_i(\mathcal{P}_{\mathrm{do}})$, and if there is a primitive inducing path between i and k then performs the extra test

$$i \perp\!\!\!\perp_{P_{\mathrm{do}(j)}} k \mid \mathrm{cause}(k) \setminus \{i\},$$

   for a j on primitive inducing path. If it holds, then the arrow from i to k will be removed. We note that, for observable interventional families, and for separable pairs, by Axiom 2 and (S2) in the supplementary material, the original definition of direct cause is equivalent to this new one. However, this definition, is an extension of the original definition for inseparable pairs.
   Notice also that in such cases, similar to bows, by only knowing $\mathcal{P}_{\mathrm{do}}$, it is not possible to distinguish an arrow from an arc between i and k. We treat such cases as a direct cause from i to k.

   (b) If there are more than one such primitive inducing paths between i and k then, In the case of faithful structural causal models, no matter which j one intervenes on, $i \not\perp_{\sigma} k \mid \mathrm{cause}(k) \setminus \{i\}$. Hence, Markov property does not imply independence of i and k given $\mathrm{cause}(k) \setminus \{i\}$.
   In such cases, one should intervene concurrently on one node on each of these primitive inducing paths to determine whether i and k become separated given $\mathrm{cause}(k) \setminus \{i\}$. Hence our theory based on single interventions is not identifying such direct causes. To be more precise, we have the following remark.

Remark 12. Let $\mathcal{P}_{\mathrm{do}}$ be a transitive interventional family. Assume $i \in \mathrm{cause}(k)$. Suppose that there are r primitive inducing paths $\pi_1, \ldots, \pi_r$ between i and k. In order to identify whether $i \in \mathrm{dcause}(k)$, the existence of a concurrent intervention $P_{\mathrm{do}(j_1, \cdots, j_r)}$ of size r, is necessary, where each j_s, for $1 \leq s \leq r$, is an inner node of π_s. $\lozenge$

8. DISCUSSION

The notion of the cause defined here is equivalent to the definition of the cause provided in the literature for structural causal models. We have provided transitivity of the cause as the first axiom on the given family, as it leads to reasonable causal graphs. Although, in general, causation does not seem to be transitive (Hall, 2000), it seems to us that examples for which causation is not transitive do not satisfy singleton-transitivity. We refrain from philosophical discussions here, but in our opinion, representing causal structure using graphs implicitly implies that one is focusing on the cases where cause in indeed transitive as the directed paths in directed graphs are transitive.

Although direct cause has been defined here with an iterative procedure in the general case, under the assumption that the causal graph is directed ancestral, causal graphs can be defined without an iterative procedure. The major departure in this paper from the literature is that the direct cause of a variable is defined using single interventions and by conditioning on other causes of the variable. As opposed to the defined causal relationships, which stand true in a larger system of variables, the direct causal relationship clearly depends on the system—it seems that one can always add a new variable to the system that breaks the direct causal relationship by sitting between the two variables as an intermediary.

Our original motivation to write this paper was to relax the common assumption that there exists a true causal graph, and, thereby, the goal of causal inference is solely to learn or estimate this graph. We do not need any such assumption under our axiomatization as we define causal graphs using intervened graphs, which themselves are defined using the concept of the direct cause for arrows and the pairwise dependencies given the joint causes in the interventional distributions for arcs. Arcs represent latent variables, and our generated graphs also allow for causal cycles. Transitivity ensures that the causal relationships in the interventional family and the ancestral relationships in graph are interchangeable.

We believe this setting can be extended to causal graphs that unify anterial graphs (Lauritzen & Sadeghi, 2018) with cyclic graphs. Such a graph represents, in addition, symmetric causal relationships implied by feedback loops; see Lauritzen & Richardson (2002). In order to do so, some extension of Markov properties, presented here for bowless directed mixed graphs, for this larger class of graphs is needed.

For the case where causal cycles exist, one advantage of the setting presented here is that it is easy to provide examples for cyclic graphs under our axiomatization; see Example 4. This is in contrast with the case of structural causal models with cycles, where, for this purpose, strong solvability assumptions must be satisfied (Bongers et al., 2021).

We remark again that there is no need for an underlying or observational distribution P to define the causal graph. We provide a minimalist and a maximalist approach to place an arc in the casual graph based on whether the arc exists in intervened graphs. This will only lead to different causal graphs where arcs between inseparable pairs of nodes are or are not present.

Although we argue that causal graphs only need $P_{\mathrm{do}(i)}$ to be defined, and have defined them only using $P_{\mathrm{do}(i)}$, we show that under the axioms of (strongly) observable interventional families, the arcs in causal graph can be directly defined using the observed distribution P.

Our definition of the arc, partly ensures automatically that the main results of the paper, i.e., Markov properties of interventional and observational distributions with respect to intervened causal and causal graphs are satisfied. However, under an alternative definition of the arc we provided in the supplementary file—which places an arc when the endpoint variables are always dependent regardless of what we condition on—no assumptions related to Markov property is in place. In this case, we need the extra assumption of ordered upward- and downward-stability to obtain the Markov property. These Markov-property results are analogous to the case of structural causal models, and, consequently, this allows the developed theory for causality in the structural causal models to be embedded in the general setting of this paper.

We mostly work on definitions and axioms for interventional families that are only related to conditional independence structure of interventional and observed distributions; although they are sufficient for generating and making sense of causal graphs, for "measuring" causal effects, which we do not discuss, they are not sufficient. For that purpose, for the case of directed ancestral causal graphs, we provide, in the supplementary material, the axiom of (bivariate) quantifiable interventional families, which relates the univariate (and bivariate) conditional-marginals of the distributions in the family to those of an underlying distribution P. In principle, P could be learned via observation, and in the case of directed acyclic graphs, it is determined uniquely by the interventional family. The extension of this axiom to bowless directed mixed graphs seems quite technical, and requires further study.

The satisfaction of the axioms for a family of distributions does not mean that the family provides the correct interventions—refer again to Example 4 to observe that all three types of edges, as the causal graph for different interventional families of two variables with the same underlying distribution P, can occur. Finding the correct interventional families is a question for mathematical and statistical modeling. One can think of this as being analogous to Kolmogorov probability axioms (Kolmogorov, 1960): a measurable space satisfying Kolmogorov axioms does not mean that it provides the correct probability for the experiment at hand. This is not the case in the structural causal model setting, as in the presence of densities, interventional distributions with full support are equivalent, modulo technical counterexamples (Peters et al., 2017) —this is because the causal graph in this setting is assumed to exist and already set in place. Example 5 shows that even the skeleton of the causal graph could change by the change of interventions with the same underlying distribution.

When we relate structural causal models to the setting of this paper, we find that if the structural causal model satisfies some weaker version of faithfulness given by the edge-cause and converse pairwise Markov property, then, in the case where the natural graph associated with the structural causal model is ancestral, the causal graph, given by standard interventions, is the same as the structural causal model graph; if the graph is not ancestral, then if the interventions are transitive, we can recover this result for maximal bowless directed mixed graphs. These results demonstrate that our theory is compatible with the standard theory, for a large class of structural causal model.

We have not provided conditions on structural causal models under which the cause is transitive, although it is not used for the main results related to structural causal models being embedded in the setting of this paper. Our initial investigation revealed that this is quite a technical problem. This is nevertheless beyond the scope of this paper, and is a subject of future work.

Finally, although an advantage of our theory is that it only relies on single interventions, our theory might misidentify direct causes between primitive inducing paths in the intervened graphs for non-maximal non-ancestral causal graphs. We provide an adjustment to deal with this when there is only one primitive inducing path exists between a pair. If there are more primitive inducing paths between a pair, we showed that we need multiple concurrent intervention of the size of the number of primitive inducing paths between the pair.

Similarly, our theory does not include some cases where multiple concurrent interventions could act as the cause of a random variable whereas none of them individually act as the cause; for example; see Example 7. Understanding these cases, and developing a similar theory for such cases is a subject of future work.

ACKNOWLEDGMENTS

The authors are grateful to Patrick Forré for a helpful discussion on pairwise Markov properties for graphs with directed cycles, and to Philip Dawid, Thomas Richardson, and Jiji Zhang for helpful general discussions related to this manuscript. We thank the referees for their encouragement and helpful comments. Work of the first author is supported in part by the EPSRC grant EP/W015684/1.

SUPPLEMENTARY MATERIAL

We provide the theory needed to prove the results presented in Section 2·4, including the proof of equivalence of the pairwise and global Markov property under compositional graphoids for maximal bowless directed mixed graphs. We provide certain alternatives and adjustments to the definition of causal graphs. We specialize the definitions and results for directed ancestral graphs. We provide additional axioms of quantifiable interventional families for the purpose of measuring causal effects. We consider the addition of bows. We contrast, in more detail, our work with the fundamentally different work on the axiomatization in the counterfactual setting. We also provide all the proofs.

REFERENCES

BAREINBOIM, E., BRITO, C. & PEARL, J. (2011). Local characterizations of causal Bayesian networks. Berlin, Heidelberg: Springer-Verlag.

BAREINBOIM, E., CORREA, J. D., IBELING, D. & ICARD, T. (2022). *On Pearl's Hierarchy and the Foundations of Causal Inference*. New York, NY, USA: Association for Computing Machinery, 1st ed., p. 507–556.

BONGERS, S., FORRÉ, P., PETERS, J. & MOOIJ, J. M. (2021). Foundations of structural causal models with cycles and latent variables. *Ann. Statist.* **49**, 2885 – 2915.

CHANG, J. T. & POLLARD, D. (1997). Conditioning as disintegration. *Statist. Neerlandica* **51**, 287–317.

COLOMBO, D. & MAATHUIS, M. H. (2014). Order-independent constraint-based causal structure learning. *J. Mach. Learn. Res.* **15**, 3741–3782.

DAWID, A. (2010). Beware of the DAG! *Journal of Machine Learning Research - Proceedings Track* **6**, 59–86.

DAWID, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **41**, 1–31.

DAWID, A. P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review* **70**, 161–189.

DAWID, P. (2021). Decision-theoretic foundations for statistical causality. *Journal of Causal Inference* **9**, 39–77.

EBERHARDT, F. & SCHEINES, R. (2007). Interventions and causal inference. *Philosophy of Science* **74**, 981–995.

FORRE, P. & MOOIJ, J. M. (2017). Markov properties for graphical models with cycles and latent variables. *arXiv:1710.08775* .

HALL, N. (2000). Causation and the price of transitivity. *Journal of Philosophy* **97**, 198.

HUANG, Y. & VALTORTA, M. (2006). Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06. Arlington, Virginia, USA: AUAI Press.

KIIVERI, H., SPEED, T. P. & CARLIN, J. B. (1984). Recursive causal models. *J. Aust. Math. Soc., Ser. A* **36**, 30–52.

KOLMOGOROV, A. N. (1960). *Foundations of the Theory of Probability*. Chelsea Pub Co, 2nd ed.

KORB, K. B., HOPE, L. R., NICHOLSON, A. E. & AXNICK, K. (2004). Varieties of causal intervention. In *PRICAI 2004: Trends in Artificial Intelligence*, C. Zhang, H. W. Guesgen & W.-K. Yeap, eds. Berlin, Heidelberg: Springer Berlin Heidelberg.

LAURITZEN, S. L. (1996). *Graphical Models*. Oxford, United Kingdom: Clarendon Press.

LAURITZEN, S. L. & RICHARDSON, T. S. (2002). Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 321–348.

LAURITZEN, S. L. & SADEGHI, K. (2018). Unifying Markov properties for graphical models. *Ann. Statist.* **46**, 2251–2278.

PARK, J., BUCHHOLZ, S., SCHÖLKOPF, B. & MUANDET, K. (2023). A measure-theoretic axiomatisation of causality.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems : networks of plausible inference*. San Mateo, CA, USA: Morgan Kaufmann Publishers.

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

PETERS, J. (2015). On the intersection property of conditional independence and its application to causal discovery. *J. Causal Inference* **3**, 97–108.

PETERS, J., JANZING, D. & SCHÖLKOPF, B. (2017). *Elements of Causal Inference - Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning Series. Cambridge, MA, USA: The MIT Press.

RAMSEY, J., SPIRTES, P. & ZHANG, J. (2006). Adjacency-faithfulness and conservative causal inference. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06. Arlington, Virginia, USA: AUAI Press.

RICHARDSON, T. S. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.* **30**, 145–157.

RICHARDSON, T. S. & SPIRTES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30**, 962–1030.

RISCHEL, E. & WEICHWALD, S. (2021). Compositional abstraction error and a category of causal models. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence,*, Proceedings of Machine Learning Research. PMLR.

SADEGHI, K. (2017). Faithfulness of probability distributions and graphs. *J. Mach. Learn. Res.* **18**, 1–29.

SADEGHI, K. & SOO, T. (2022). Conditions and assumptions for constraint-based causal structure learning. *Journal of Machine Learning Research* **23**, 1–34.

SHPITSER, I. & PEARL, J. (2006). Identification of conditional interventional distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06. Arlington, Virginia, USA: AUAI Press.

SPIRTES, P., GLYMOUR, C. & SCHEINES, R. (2000). *Causation, Prediction, and Search*. MIT press, 2nd ed.

STUDENÝ, M. (2005). *Probabilistic Conditional Independence Structures*. London, United Kingdom: Springer-Verlag.

VERMA, T. & PEARL, J. (1988). Causal networks: semantics and expressiveness. *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence (UAI)* **4**, 352–359.

WOODWARD, J. (2004). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

ZHANG, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artif. Intell.* **172**, 1873–1896.

ZHANG, J. & SPIRTES, P. (2008). Detection of unfaithfulness and robust causal inference. *Minds and Machines* **18**, 239–271.

[*Received* 3 *December* 2023]