
A Potential Outcomes Calculus for Identifying Conditional Path-Specific Effects

Daniel Malinsky

Johns Hopkins University
Department of Computer Science
Baltimore, MD USA
malinsky@jhu.edu

Ilya Shpitser

Johns Hopkins University
Department of Computer Science
Baltimore, MD USA
ilyas@cs.jhu.edu

Thomas Richardson

University of Washington
Department of Statistics
Seattle, WA USA
thomasr@uw.edu

Abstract

The do-calculus is a well-known deductive system for deriving connections between interventional and observed distributions, and has been proven complete for a number of important identifiability problems in causal inference [1, 8, 18]. Nevertheless, as it is currently defined, the do-calculus is inapplicable to causal problems that involve complex nested counterfactuals which cannot be expressed in terms of the “do” operator. Such problems include analyses of path-specific effects and dynamic treatment regimes. In this paper we present the *potential outcome calculus* (po-calculus), a natural generalization of do-calculus for arbitrary potential outcomes. We thereby provide a bridge between identification approaches which have their origins in artificial intelligence and statistics, respectively. We use po-calculus to give a complete identification algorithm for conditional path-specific effects with applications to problems in mediation analysis and algorithmic fairness.

1 Introduction

Pearl’s do-calculus [6, 7, 8] is an abstract set of rules for reasoning about interventions that has proven to be influential in settings, such as computer science and artificial intelligence, where graphical models are used to represent causal relationships. In statistics and some social/biomedical sciences, the potential outcome framework [4, 15] is more commonly used to express causal assumptions and reason about interventions. Richardson and

Robins [11] have made an important contribution by unifying causal formalisms grounded in graphical causal models with the potential outcomes framework. In this paper we build on those connections, presenting a calculus for reasoning about interventions in the potential outcomes notation that is equivalent to Pearl’s do-calculus for standard interventions, but allows generalizations to nested causal quantities pertinent to evaluating (e.g.) dynamic treatment regimes or path-specific interventions (for which the “do” notation is insufficiently expressive). We show how the new calculus can be applied to problems in mediation analysis, specifically the identification of conditional path-specific causal effects. We introduce a procedure which is complete for expressing such quantities as functions of the observed data distribution, i.e., an algorithm which will produce an identifying expression for a conditional path-specific effect if and only if the effect is identifiable.

Conditional path-specific effects are quantified via conditional distributions over potential outcomes, where treatment variables are assigned to possibly distinct values for different causal pathways. In mediation analysis, functions of such distributions are used to isolate the effect of a drug, therapy, or other treatment assignment along a specific pathway in a specific subpopulation, defined by pre-treatment variables (such as age or gender) or post-treatment variables (such as adverse reactions to the treatment). Importantly, there are settings where the marginal path-specific effect is identified but the conditional path-specific effect is not identified; we later discuss one simple example shown in Fig. 1.

Another context in which conditional path-specific effects may be of interest is in the study of algorithmic fairness. Recent papers [2, 3, 21] have proposed to combat disparities perpetuated by some automated decision-making systems by identifying, estimating, and constraining unfair causal influences that propagate along certain pathways, e.g., the direct effect of gender on hiring outcomes or the indirect effect of race on criminal justice outcomes via geographical factors. It may also be desirable to constrain such

path-specific effects for certain subpopulations, which requires identifying conditional path-specific effects.

We begin by introducing potential outcomes, causal models, graphs, and some relevant results. Then we review the do-calculus, propose our potential outcome calculus, demonstrate they are equivalent, and give some simple derivations to establish the soundness of the rules in the language of potential outcomes. Finally, we introduce a formalism for expressing path-specific effects (PSEs) and a complete identification procedure for conditional PSEs.

2 Potential Outcomes, the Do Operator and Causal Models

Fix a set of indices $K \equiv \{1, \dots, k\}$ under a total ordering \prec . For each random variable $V_i, i \in K$, define a state space \mathfrak{X}_i , and the sets $\text{Pre}_i \equiv \{1, \dots, i-1\}$. Given $A \subseteq K$, we will denote subsets of random variables indexed by A with V_A and elements v_A of \mathfrak{X}_A by a (lowercase letters).

We assume the existence of all one-step-ahead *potential outcome* random variables (a.k.a. counterfactuals) of the form $V_i(\text{pa}_i) \equiv V_i(v_{\text{Pa}_i})$, where Pa_i is a fixed subset of Pre_i , and $\text{pa}_i \equiv v_{\text{Pa}_i}$ is any element in $\mathfrak{X}_{\text{Pa}_i}$. The variable $V_i(\text{pa}_i)$ denotes the value of V_i had the set of *direct causes* of V_i , or V_{Pa_i} , been set, possibly contrary to fact, to values pa_i . The existence of a total ordering \prec on indices, and the fact that $\text{Pa}_i \subseteq \text{Pre}_i$ precludes the existence of cyclic causation. (That is, we consider causal models that are *recursive*.) $V_i(\text{pa}_i)$ may be conceptualized as the output of a *structural equation* $f_i : \mathfrak{X}_{\text{Pa}_i \cup \{\epsilon_i\}} \mapsto \mathfrak{X}_i$, a function representing a causal mechanism that maps values of Pa_i , as well as the value of an exogenous disturbance variable ϵ_i , to values of V_i . We define causal models as sets of densities over the set of random variables

$$\mathbb{V} \equiv \{V_i(\text{pa}_i) \mid i \in \{1, \dots, k\}, \text{pa}_i \in \mathfrak{X}_{\text{Pa}_i}\}.$$

For simplicity of presentation, we assume \mathfrak{X}_i is always finite, and thus ignore the measure theoretic complications that arise with defining densities over sets of random variables above in the case where some state spaces on Pa_i are infinite.¹

Given a set of one-step-ahead potential outcomes \mathbb{V} , for any $A \subseteq K$ and $i \in K$ we define the potential outcome $V_i(a)$, the response of V_i had variables in V_A been set to a , by the definition known as *recursive substitution*:

$$V_i(a) \equiv V_i(a \cap \text{pa}_i, \{V_j(a) \mid j \in \text{Pa}_i \setminus A\}). \quad (1)$$

In words, this states that $V_i(a)$ is the potential outcome where variables Pa_i in A are set to their corresponding val-

ues in a , and all elements of Pa_i not in A are set to whatever values their recursively defined counterfactual versions would have had had A been set to a . Equivalently, $V_i(a)$ is the random variable induced by a modified set of structural equations: specifically the set of functions f_j for all $V_j \in A$ are replaced by constant functions f_j^* that set V_j to the corresponding value in a .

We denote by \mathbb{V}^* the set of all variables derived by (1) from \mathbb{V} , together with \mathbb{V} . In addition, for notational conciseness, we will use index sets to denote sets of potential outcomes themselves. That is, for $Y \subseteq K, A \subseteq K$, we will denote the set $\{V_i(a) \mid i \in Y\}$ by $Y(a)$. Note that we allow Y and A to intersect. Thus, we allow sets of potential outcomes of the form $A(a)$, which denote the sets $\{V_i(a) \mid i \in A\}$, where each $V_i(a)$ is defined using (1) above. In particular, if $A = \{V_i\}$ (a singleton), $V_i(v_i)$ is defined in our notation to be the random variable V_i , not the constant v_i .

In cases where Y and A do not intersect, the distribution $p(Y(a))$ has been denoted by Pearl as $p(Y \mid \text{do}(a))$ [8]. This formulation places emphasis on the intervention operator $\text{do}(a)$, which replaces structural equations by constants.

Recursive substitution provides a link between observed variables and potential outcomes. In particular, it implies the *consistency property*:² for any disjoint $A, B \subseteq K, i \in K \setminus (A \cup B), a \in \mathfrak{X}_A, b \in \mathfrak{X}_B$,

$$B(a) = b \text{ implies } V_i(a, b) = V_i(a). \quad (2)$$

Proposition 1 (consistency) *Given \mathbb{V}^* derived from \mathbb{V} via (1), then (2) holds.*

Proof: By (1), $V_i(a)$ and $V_i(a, b)$ are defined as

$$V_i(a_{\text{Pa}_i}, \{V_j(a) \mid j \in \text{Pa}_i \setminus (A \cup B)\}, \{V_j(a) = b_j \mid j \in B \cap \text{Pa}_i\})$$

$$\text{and } V_i(a_{\text{Pa}_i}, \{V_j(a, b) \mid j \in \text{Pa}_i \setminus (A \cup B)\}, b_{\text{Pa}_i}),$$

respectively. The conclusion follows immediately. \square

(1) implies that every $V_i(a)$ can be written as a function of a unique minimally causally relevant subset of a .

Proposition 2 (causal irrelevance) *Given \mathbb{V}^* derived from \mathbb{V} via (1), let $V_i(a) \in \mathbb{V}^*$, and let A^* be the maximal subset of A such that for every $A_j \in A^*$, there exists a sequence V_{w_1}, \dots, V_{w_m} that does not intersect A , where $A_j \in \text{Pa}_{w_1}, V_{w_i} \in \text{Pa}_{w_{i+1}}$, for $i = 1, \dots, m-1$, and $V_{w_m} \in \text{Pa}_i$. Then $V_i(a) = V_i(a^*)$.*

Proof: Follows by definition of A^* and (1). \square

A *functional causal model* (a.k.a. a non-parametric structural equation model with independent errors, NPSEM-IE)

¹The set of $p(\mathbb{V})$ for a particular set of Pa_i and an ordering \prec was called the *finest causally interpretable structured tree graph (FCISTG)* in [12].

²Some readers may be more familiar with the simpler formulation where $a = \emptyset$, so “ $B = b$ implies $V_i(b) = V_i$.” Our reasons for allowing multiple intervention sets will become clear in what follows.

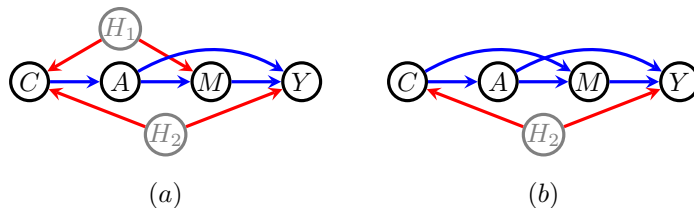


Figure 1: (a) A hidden variable causal DAG where $p(Y(a, M(a')))$ is identified, but $p(Y(a, M(a')) | C)$ is not identified. (b) A seemingly similar hidden variable causal DAG where both $p(Y(a, M(a')))$ and $p(Y(a, M(a')) | C)$ are identified.

asserts that the sets of variables

$$\{\{V_i(\text{pa}_i) \mid \text{pa}_i \in \mathfrak{X}_{\text{pa}_i}\} \mid i \in \{1, \dots, k\}\} \quad (3)$$

are mutually independent. Phrased in terms of structural equations, the functional causal model states that the joint distribution of the disturbance terms factorizes into a product of marginals: $p(\epsilon_1, \dots, \epsilon_k) = \prod_{i=1}^k p(\epsilon_i)$.

Alternative causal models, which make fewer assumptions than the functional model but are sufficient for all inferences we aim to make in this paper, are discussed in [11, 20]. We focus on the functional causal model here, since it is simpler to describe and the original setting of Pearl’s do-calculus. We discuss how our results apply to a weaker causal model [11] in the Supplement.

3 Graphical Models

Much conceptual clarity may be gained by viewing causal models as graphs. We will consider graphs with either directed edges only (\rightarrow), or mixed graphs with both directed and bidirected (\leftrightarrow) edges. Vertices correspond to random variables, and we simplify notation by using V_i to refer to both the graph vertex and corresponding random variable. In all cases we will require the absence of directed cycles, meaning that whenever the graph contains a path of the form $V_i \rightarrow \dots \rightarrow V_j$, the edge $V_j \rightarrow V_i$ cannot exist. Directed graphs with this property are called directed acyclic graphs (DAGs), and mixed graphs with this property are called acyclic directed mixed graphs (ADMGs). We will refer to graphs by $\mathcal{G}(V)$, where V is the set of random variables indexed by $\{1, \dots, k\}$. We will use the following standard definitions for sets of vertices in a graph:

$$\begin{aligned} \text{Pa}_i^{\mathcal{G}} &\equiv \{V_j \mid V_j \rightarrow V_i \text{ in } \mathcal{G}\} && \text{(parents of } V_i) \\ \text{An}_i^{\mathcal{G}} &\equiv \{V_j \mid V_j \rightarrow \dots \rightarrow V_i \text{ in } \mathcal{G}\} && \text{(ancestors of } V_i) \\ \text{De}_i^{\mathcal{G}} &\equiv \{V_j \mid V_j \leftarrow \dots \leftarrow V_i \text{ in } \mathcal{G}\} && \text{(descendants of } V_i) \end{aligned}$$

By convention, we assume $V_i \in \text{An}_i^{\mathcal{G}}$ and $V_i \in \text{De}_i^{\mathcal{G}}$. We will generally drop the superscript \mathcal{G} if the relevant graph is obvious and sometimes write \mathcal{G} in place of $\mathcal{G}(V)$ when the vertex set is clear. Given a DAG $\mathcal{G}(V)$, a statistical DAG model (a.k.a. a Bayesian network) associated with $\mathcal{G}(V)$ is

a set of distributions that are Markov relative to $\mathcal{G}(V)$, i.e., the set of distributions that can be written as the following product of conditional densities:

$$p(V) = \prod_{i=1}^k p(V_i \mid \text{Pa}_i). \quad (4)$$

Given $p(V)$ that is Markov relative to a DAG $\mathcal{G}(V)$, conditional independence relations (written: $(Y \perp\!\!\!\perp Z \mid X)$, where X, Y, Z are disjoint subsets of the index set K) satisfied by $p(V)$ can be derived using the well-known d-separation criterion [5], which we reproduce in the Supplement. We write $(Y \perp\!\!\!\perp_d Z \mid X)_{\mathcal{G}(V)}$ when Y is d-separated from Z given X in $\mathcal{G}(V)$. If $p(V)$ is Markov relative to $\mathcal{G}(V)$, then the following *global Markov property* holds: for any disjoint X, Y, Z

$$(Y \perp\!\!\!\perp_d Z \mid X)_{\mathcal{G}(V)} \Rightarrow (Y \perp\!\!\!\perp Z \mid X) \text{ in } p(V).$$

Functional causal models may also be associated with a DAG \mathcal{G} by identifying Pa_i with the graphical parents of V_i in $\mathcal{G}(V)$. Given a functional causal model for DAG \mathcal{G} , the joint distribution for any $V(a)$ derived from \mathbb{V} using (1) is identified via the following formula:

$$p(V(a)) = \prod_{i=1}^K p(V_i \mid \text{Pa}_i \setminus A, a \cap \text{pa}_i), \quad (5)$$

provided $\left(\prod_{a_j \in a} p(a_j \mid \text{Pa}_j \setminus A, a \cap \text{pa}_j)\right) > 0$. See [11] for a simple proof. The modified factorization (5) is known as the *extended g-formula* [11, 13]. Note that (5) has a term for every $V_i \in V$, just like (4).

The formula (5) resembles (4) and in fact may be viewed as a factorization of $p(V(a))$ with respect to a certain graph derived from \mathcal{G} . Such graphs, called Single World Intervention Graphs (SWIGs), were introduced in [11]. SWIGs are graphical representations of potential outcome densities that help unify the graphical and potential outcome formalisms. Given a set A of variables which are assigned to values a , a SWIG $\mathcal{G}(a)$ is constructed from $\mathcal{G}(V)$ by splitting all vertices in A into a random half and a fixed half, with the random half inheriting all edges with an incoming arrowhead and the fixed half inheriting all outgoing directed edges. Then, all random vertices V_i are re-labelled as

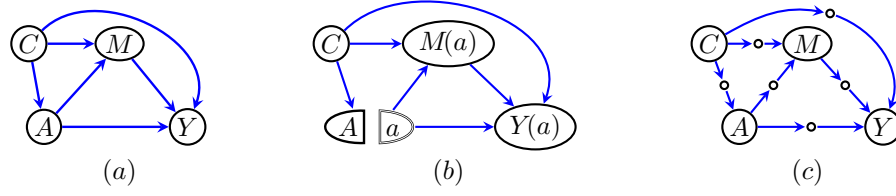


Figure 2: (a) A simple causal DAG \mathcal{G} , with a treatment A , an outcome Y , a vector C of baseline variables, and a mediator M . (b) A SWIG $\mathcal{G}(a)$ derived from (a) corresponding to the world where A is intervened on to value a . (c) An extended graph \mathcal{G}^e derived from (a).

$V_i(a)$ or equivalently (due to Proposition 2) as $V_i(a \cap \text{an}_i^*)$, where an_i^* consists of values of the ancestors of V_i in the split graph. In [11], unsplit vertices were drawn as circles, and split nodes as half circles, with fixed nodes denoted by a lowercase. Fixed nodes are enclosed by a double line. For an example of a SWIG representing the joint density $p(Y(a), M(a), C(a), A(a)) = p(Y(a), M(a), C, A)$, see Fig. 2 (b). Because of the resemblance of (5) to a DAG factorization, we say that $p(V(a))$ is Markov relative to a SWIG $\mathcal{G}(a)$ if $p(V(a))$ may be written as (5).

A SWIG $\mathcal{G}(a)$ is a DAG with a vertex set $\{V(a), a\}$, and may be viewed as a *conditional* graph, with vertices in $V(a)$ corresponding to random variables, and vertices in a corresponding to variables fixed to a value. We extend the notion of d-separation to allow fixed vertices. Specifically, we allow d-separation statements of the form $(Y(a), a' \perp\!\!\!\perp_d Z(a) \mid X(a))_{\mathcal{G}(a)}$, for disjoint random subsets $Y(a), Z(a), X(a)$ of $V(a)$ and a' a subset of a . Note that a possibly d-connecting path may only contain random nodes as non-endpoint vertices (as in [11] where fixed nodes are always blocked). Our extension here consists only in allowing fixed vertices to also appear as one endpoint in a d-separation statement. Just as (4) implied the global Markov property for a DAG, the modified factorization (5) implies a global Markov property for a SWIG.

Proposition 3 (SWIG global Markov property) *If $p(V(a))$ is Markov relative to $\mathcal{G}(a)$, then for any disjoint subsets $Y(a), Z(a), X(a)$ of $V(a)$ and a subset a' of a , if $(Y(a), a' \perp\!\!\!\perp_d Z(a) \mid X(a))_{\mathcal{G}(a)}$ then, for some $f(\cdot)$,*

$$p(Z(a) \mid Y(a), X(a)) = p(Z(a) \mid X(a)) = f(Z, X, a \setminus a').$$

Proof: The first equality is due to Theorem 12 in [11], the second follows from Theorem 19 in [10]. \square

Note that $f(Z, X, a \setminus a')$ is not necessarily equal to $p(Z(a \setminus a') \mid X(a \setminus a'))$.

The SWIG global Markov property implies the following intuitive result (proved in the Supplement) relating independence statements in $p(V(a))$ for various sets A . Specifically, the result is that interventions “always help” when it comes to conditional independence.

Proposition 4 (intervention monotonicity) *For any disjoint subsets $Y(a), Z(a), X(a)$ of $V(a)$ and a subset a' of a , if $(Y(a), a' \perp\!\!\!\perp Z(a) \mid X(a))_{\mathcal{G}(a)}$ then for any $A'' \supseteq A$, $(Y(a''), a' \perp\!\!\!\perp Z(a'') \mid X(a''))_{\mathcal{G}(a'')}$.*

Graphical Models With Hidden Variables

We also consider causal models where some variables are unmeasured (a.k.a. “latent” or “hidden” variables). Given a DAG $\mathcal{G}(V \cup H)$, define a *latent projection* mixed graph $\mathcal{G}(V)$ as follows. V is the vertex set of $\mathcal{G}(V)$, and for any $V_i, V_j \in V$ there is an edge $V_i \rightarrow V_j$ if there exists a directed path from V_i to V_j in $\mathcal{G}(V \cup H)$, with all intermediate nodes on the path in H ; there is an edge $V_i \leftrightarrow V_j$ if there exists a path from V_i to V_j of the form $V_i \leftarrow \dots \rightarrow V_j$, where every intermediate node on the path is in H and no consecutive edges on the path are of the form $\rightarrow H_k \leftarrow$ for $H_k \in H$. The latent projection $\mathcal{G}(V)$ obtained from a DAG $\mathcal{G}(V \cup H)$ is always an ADMG. Our results in this paper apply to ADMGs, and indeed this is the intended setting for Pearl’s do-calculus (he used the terminology “semi-Markovian models”).

The definition of d-separation naturally generalizes to ADMGs with minor modification for bidirected edges; the resulting criterion is called m-separation [9]. We write $(Y \perp\!\!\!\perp_m Z \mid X)_{\mathcal{G}(V)}$ if Y is m-separated from Z given X in ADMG $\mathcal{G}(V)$. In the following we sometimes drop the d or m subscripts and just write $\perp\!\!\!\perp$, where the relevant criterion is implicit.

Given an ADMG $\mathcal{G}(V)$, we define a SWIG $\mathcal{G}(V)(a)$ by the analogous node splitting construction as for DAGs. Specifically, each node is split into a random half and a fixed half, with random halves inheriting all incoming directed and bidirected edges, and fixed halves inheriting all outgoing directed edges. Alternatively given a SWIG $\mathcal{G}(a)$ derived from a DAG $\mathcal{G}(V \cup H)$, we define the latent projection operation in the natural way, yielding the SWIG $\mathcal{G}(a)(V)$ with random vertices V , fixed vertices a , and directed edges from $a_i \in a$ or $V_i \in V$ to $V_j \in V$ if there is a directed path from the corresponding vertices in $\mathcal{G}(a)$ with all intermediate vertices in H , and bidirected edges from

$V_i \in V$ to $V_j \in V$ if there exists a path from V_i to V_j of the form $V_i \leftarrow \dots \rightarrow V_j$, where every intermediate node on the path is in H and no consecutive edges on the path are of the form $\rightarrow H_k \leftarrow$ for $H_k \in H$. These operations commute, and we can derive independence statements via m-separation on $G(V)(a)$, as we prove in the Supplement.

4 Do-Calculus and Potential Outcomes Calculus

Pearl formulated the do-calculus originally as follows:

- 1 : $p(y \mid z, w, \text{do}(x)) = p(y \mid w, \text{do}(x))$
if $(Y \perp\!\!\!\perp Z \mid W, X)_{\mathcal{G}_{\overline{X}}}$
- 2 : $p(y \mid z, w, \text{do}(x)) = p(y \mid w, \text{do}(z), \text{do}(x))$
if $(Y \perp\!\!\!\perp Z \mid W, X)_{\mathcal{G}_{\overline{X}, \underline{Z}}}$
- 3 : $p(y \mid w, \text{do}(z), \text{do}(x)) = p(y \mid w, \text{do}(x))$
if $(Y \perp\!\!\!\perp Z \mid W, X)_{\mathcal{G}_{\overline{X}, \overline{Z(W)}}}$

where $\mathcal{G}_{\overline{X}}$ denotes the graph obtained from \mathcal{G} by removing all edges with arrowheads into X , $\mathcal{G}_{\underline{Z}}$ denotes the graph obtained from \mathcal{G} by removing all directed edges out of Z , and $Z(W) \equiv Z \setminus \text{An}_{\mathcal{G}_{\overline{X}}}(W)$.

Here we present the do-calculus entirely in terms of potential outcomes (the ‘‘potential outcomes calculus’’ or ‘‘po-calculus’’ for short). The conditions are phrased in terms of conditional independencies implied by SWIGs, e.g., $\mathcal{G}(x)$ for the SWIG where X is assigned value x . We restate the rules as follows:

- 1 : $p(Y(x) \mid Z(x), W(x)) = p(Y(x) \mid W(x))$
if $(Y(x) \perp\!\!\!\perp Z(x) \mid W(x))_{\mathcal{G}(x)}$
- 2 : $p(Y(x, z) \mid W(x, z)) = p(Y(x) \mid W(x), Z(x) = z)$
if $(Y(x, z) \perp\!\!\!\perp Z(x, z) \mid W(x, z))_{\mathcal{G}(x, z)}$
- 3 : $p(Y(x, z) \mid W(x, z)) = p(Y(x) \mid W(x))$
if $(Y(x, z_1), W(x, z_1) \perp\!\!\!\perp z_1)_{\mathcal{G}(x, z_1)}$ and
 $(Y(x, z_1) \perp\!\!\!\perp Z_2(x, z_1) \mid W(x, z_1))_{\mathcal{G}(x, z_1)}$
where $Z_1 = Z \setminus \text{An}_{\mathcal{G}(x)}(W)$,
 $Z_2 = Z \cap \text{An}_{\mathcal{G}(x)}(W)$

Recall that random variables in a SWIG $\mathcal{G}(x)$ are labelled $V_i(x)$ or equivalently as $V_i(x \cap \text{an}_i^*)$, where an_i^* consists of values of the ancestors of V_i in the split graph. We can view Rule 1 as the fragment of the SWIG global Markov property that pertains to random variables in $V(a)$. Rule 2 may be called ‘‘generalized conditional ignorability’’ because it is a general version of the standard ignorability assumption used in causal inference settings, where $(Y(a) \perp\!\!\!\perp A \mid C)$, or equivalently $(Y(a) \perp\!\!\!\perp A(a) \mid C(a))$, enables identification of (e.g.) the average treatment effect by adjusting for C . Note that Rule 3 does not have a simple interpretation, as it involves an equality of interventional distributions in

two distinct ‘‘worlds,’’ given an independence condition in a third. However, below we suggest an alternative, simpler rule which may be used without loss of generality, and is more intuitive. First, we state some basic results.

Proposition 5 *Rule 1 of po-calculus holds if and only if Rule 1 of do-calculus holds.*

Proof: Follows from the definition of $\mathcal{G}(x)$ and $\mathcal{G}_{\overline{X}}$, and the definition of m-separation. \square

Proposition 6 *Rule 2 of po-calculus holds if and only if Rule 2 of do-calculus holds.*

Proof: Follows from the definition of $\mathcal{G}(x, z)$ and $\mathcal{G}_{\overline{X}, \underline{Z}}$, and the definition of m-separation in $\mathcal{G}(x, z)$. \square

Proposition 7 *Rule 3 of po-calculus holds if and only if Rule 3 of do-calculus holds.*

Proof: Since path separation criteria on graphs quantify over elements in vertex sets, and since Z is a disjoint union of Z_1 ($Z(W)$ in Pearl’s terminology) and Z_2 , the precondition in Rule 3 of do-calculus may be written as two preconditions: $(Y \perp\!\!\!\perp Z_1 \mid W, X)_{\mathcal{G}_{\overline{X}, \overline{Z_1}}}$ and $(Y \perp\!\!\!\perp Z_2 \mid W, X)_{\mathcal{G}_{\overline{X}, \overline{Z_1}}}$.

By definition of Z_1 , it contains only non-ancestors of W in $\mathcal{G}_{\overline{X}}$ (and therefore also in $\mathcal{G}_{\overline{X}, \overline{Z_1}}$, which is an edge subgraph of $\mathcal{G}_{\overline{X}}$). Since Z_1 only has adjacent outgoing directed arrows in $\mathcal{G}_{\overline{X}, \overline{Z_1}}$, all elements of W are marginally m-separated from Z_1 in $\mathcal{G}_{\overline{X}, \overline{Z_1}}$. Thus, $(W(x, z_1) \perp\!\!\!\perp z_1)_{\mathcal{G}(x, z_1)}$ by the definition of $\mathcal{G}(x, z_1)$. Furthermore, no element of Z_1 can be an ancestor of Y in $\mathcal{G}_{\overline{X}, \overline{Z_1}}$. To see this, suppose an element Z_i of Z_1 were an ancestor of Y . Then since $(Y \perp\!\!\!\perp Z_1 \mid W, X)_{\mathcal{G}_{\overline{X}, \overline{Z_1}}}$, the directed path from Z_i must be blocked by W and X . W cannot be on this directed path because it is non-descendant of Z_1 , and X cannot be on the path because $\mathcal{G}_{\overline{X}, \overline{Z_1}}$ has no directed edges into X . So we conclude that Z_i is not an ancestor of Y in $\mathcal{G}_{\overline{X}, \overline{Z_1}}$ and therefore $(Y(x, z_1) \perp\!\!\!\perp z_1)_{\mathcal{G}(x, z_1)}$ by the definition of $\mathcal{G}(x, z_1)$. Thus, if do-calculus Rule 3 precondition holds, po-calculus Rule 3 precondition holds.

We now prove the converse. If $(Y(x, z_1) \perp\!\!\!\perp z_1)_{\mathcal{G}(x, z_1)}$ then Z_1 is not an ancestor of Y in $\mathcal{G}_{\overline{X}, \overline{Z_1}}$. Similarly if $(W(x, z_1) \perp\!\!\!\perp z_1)_{\mathcal{G}(x, z_1)}$ then Z_1 is not an ancestor of W in $\mathcal{G}_{\overline{X}, \overline{Z_1}}$. Since Z_1 only has adjacent edges that are outgoing directed edges, this implies $(Y, W \perp\!\!\!\perp Z_1 \mid X)_{\mathcal{G}_{\overline{X}, \overline{Z_1}}}$ holds. Since semi-graphoid axioms hold for m-separation, this implies $(Y \perp\!\!\!\perp Z_1 \mid W, X)_{\mathcal{G}_{\overline{X}, \overline{Z_1}}}$ holds. Finally, $(Y(x, z_1) \perp\!\!\!\perp Z_2(x, z_1) \mid W(x, z_1))_{\mathcal{G}(x, z_1)}$ holds if and only if $(Y \perp\!\!\!\perp Z_2 \mid W, X)_{\mathcal{G}_{\overline{X}, \overline{Z_1}}}$ holds, by the definitions of $\mathcal{G}(x, z_1)$, $\mathcal{G}_{\overline{X}, \overline{Z_1}}$, and m-separation. \square

We now briefly demonstrate the soundness of the three rules of the po-calculus using only potential outcomes machinery and our background assumptions.

Proposition 8 *Rules 1, 2, and 3 are sound.*

Proof: Proposition 3 licenses deriving conditional independence statements corresponding to the graphical conditions in each rule. Then we have the following derivations:

Rule 1: $p(Y(x)|Z(x), W(x)) = p(Y(x)|W(x))$
by $Y(x) \perp\!\!\!\perp Z(x) \mid W(x)$.

Rule 2: $p(Y(x, z)|W(x, z)) = p(Y(x, z)|Z(x, z) = z, W(x, z)) = p(Y(x)|Z(x), W(x))$
by $Y(x, z) \perp\!\!\!\perp Z(x, z) \mid W(x, z)$ and consistency.

Rule 3: $p(Y(x)|W(x)) = p(Y(x, z_1)|W(x, z_1))$
since $Y(x, z_1), W(x, z_1) \perp\!\!\!\perp z_1$.
 $= p(Y(x, z_1)|Z_2(x, z_1) = z_2, W(x, z_1))$
since $Y(x, z_1) \perp\!\!\!\perp Z_2(x, z_1)|W(x, z_1)$.
 $= p(Y(x, z_1, z_2)|Z_2(x, z_1, z_2) = z_2, W(x, z_1, z_2))$
by consistency.
 $= p(Y(x, z)|Z_2(x, z) = z_2, W(x, z))$
since $Y(x, z_1) \perp\!\!\!\perp Z_2(x, z_1) \mid W(x, z_1)$,
 $Z_2 \subseteq Z$, and so by Proposition 4,
 $= p(Y(x, z)|W(x, z))$

□

The proof of Proposition 8 has a number of interesting consequences. First, the soundness of Rule 2 follows by Rule 1 and consistency. Second, the soundness of Rule 3 follows by applications of Rule 1, Rule 2, consistency, causal irrelevance, and intervention monotonicity.

Causal irrelevance, as used in the proof, is implied by m-separation statements in the SWIG $\mathcal{G}(x, z_1)$; however this property, like consistency, follows by (1) alone and does not require any assumption regarding the distributions $p(V(a))$ for any $A \subseteq V$; specifically, (5) is not required. As a result the three rules of po-calculus, taken as a whole, are consequences of consistency and causal irrelevance, which hold in any recursive causal model, together with the SWIG Markov property for random variables in $V(a)$. (Intervention monotonicity follows from these.)

The proof of Proposition 8 also implies that a simpler reformulation of po-calculus suffices without loss of generality. Specifically, this reformulation replaces Rule 3 by the following simpler rule (encoding causal irrelevance in graphical form):

$$3^* : p(Y(x, z)) = p(Y(x)) \quad \text{if } (Y(x, z) \perp\!\!\!\perp z)_{\mathcal{G}(x, z)}.$$

A benefit of translating the do-calculus exactly into our potential outcomes formulation is that the do-calculus rules as stated have been shown to be sufficient for a wide class of possible derivations on distributions expressible in terms of the do operator [1, 18]. However, since we phrased the

rules for arbitrary potential outcomes, they may be applied to causal contrasts not expressible in standard do notation. We illustrate this by applying these rules to mediation analysis.

5 Path-Specific Effects and Extended Graphs

The identification theory for path-specific effects generally proceeds by considering nested, path-specific potential outcomes. Fix a set of treatment variables A , and a subset of *proper causal paths* π from any element in A . A proper causal path only intersects A at the source node. Next, pick a pair of value sets a and a' for elements in A . For any $V_i \in V$, define the potential outcome $V_i(\pi, a, a')$ by setting A to a for the purposes of paths in π , and to a' for the purposes of proper causal paths from A to Y not in π . Formally, the definition is as follows, for any $V_i \in V$:

$$\begin{aligned} V_i(\pi, a, a') &\equiv a \text{ if } V_i \in A \\ V_i(\pi, a, a') &\equiv \\ &V_i(\{V_j(\pi, a, a') \mid V_j \in \text{Pa}_i^\pi\}, \{V_j(a') \mid V_j \in \text{Pa}_i^{\bar{\pi}}\}) \end{aligned} \quad (6)$$

where $V_j(a') \equiv a'$ if $V_j \in A$, and given by (1) otherwise, Pa_i^π is the set of parents of V_i along an edge which is a part of a path in π , and $\text{Pa}_i^{\bar{\pi}}$ is the set of all other parents of V_i .

A counterfactual $V_i(\pi, a, a')$ is said to be *edge inconsistent* if counterfactuals of the form $V_j(a_k, \dots)$ and $V_j(a'_k, \dots)$ occur in $V_i(\pi, a, a')$, otherwise it is said to be *edge consistent*. It is well known that a joint distribution $p(V(\pi, a, a'))$ containing an edge-inconsistent counterfactual $V_i(\pi, a, a')$ is not identified in the functional causal model (nor weaker causal models) with a corresponding graphical criterion on π and $\mathcal{G}(V)$ called the ‘recanting witness’ [16, 20]. For example, in Fig. 2 (a), given $\pi = \{C \rightarrow A \rightarrow Y\}$, $Y(\pi, c, c') \equiv Y(c', M(c', A(c')), A(c))$, while given $\pi = \{A \rightarrow Y\}$, $Y(\pi, a, a') \equiv Y(C, a, M(a', C))$. Note that $Y(\pi, c, c')$ is edge inconsistent due to the presence of $A(c)$ and $A(c')$, while $Y(\pi, a, a')$ is edge consistent.

Counterfactuals defined by (6) form the basis for direct, indirect, and path-specific effects estimated in the mediation analysis literature. There are generalizations where elements in A are set to arbitrary values for different paths, under the name of *path interventions* [20]. Similarly, edge consistent counterfactuals $V(\pi, a, a')$ generalize to responses to *edge interventions* [20]. We do not discuss this further here in the interests of space, although the results presented below generalize without issue. Note that edge consistent counterfactuals cannot, in general, be phrased in terms of the do operator.

We have the following the result, proven in [20].

Theorem 1 *If $V(\pi, a, a')$ is edge consistent, then under*

the functional causal model for DAG \mathcal{G} ,

$$p(V(\pi, a, a')) = \prod_{i=1}^K p(V_i | a \cap \text{pa}_i^\pi, a' \cap \text{pa}_i^{\bar{\pi}}, \text{Pa}_i^{\mathcal{G}} \setminus A). \quad (7)$$

As an example, the distribution $p(Y(\pi, a, a')) = p(Y(C, a, M(a', C)))$ of the edge consistent counterfactual in Fig. 2 (a) is identified as a marginal distribution derived from (7), specifically $\sum_{C, M} p(Y | a, M, C) p(M | a', C) p(C)$. The po-calculus as presented above may be applied to any sort of potential outcome, including nested potential outcomes representing path-specific effects. In the following, we exploit an equivalence between path-specific potential outcomes and standard potential outcomes defined from an *extended graph* \mathcal{G}^e , which is constructed from \mathcal{G} following [14]. This both simplifies complex nested potential outcome expressions and enables us to leverage a series of prior results to identify conditional PSEs.

Given an ADMG $\mathcal{G}(V)$, define for each $A_i \in A \subseteq V$ the set of variables $A_i^{\text{Ch}} \equiv \{A_i^j \mid V_j \in \text{Ch}_i\}$, and let $A^{\text{Ch}} \equiv \bigcup_{A_i \in A} A_i^{\text{Ch}}$. We define the extended graph of $\mathcal{G}(V)$, written $\mathcal{G}^e(V \cup A^{\text{Ch}})$, as the graph with the vertex set $V \cup A^{\text{Ch}}$, with edges of the form $A_i \rightarrow A_i^j \rightarrow V_j$ if and only if $A_i \rightarrow V_j$ is present in $\mathcal{G}(V)$, for $A_i \in A$, $V_j \in V$; furthermore, $V_i \leftrightarrow V_j$ in $\mathcal{G}^e(V \cup A^{\text{Ch}})$ if and only if $V_i \leftrightarrow V_j$ is present for $V_i, V_j \in V$ in $\mathcal{G}(V)$. As an example, the extended graph for the DAG in Fig. 2 (a), with $A = V$, is shown in Fig. 2 (c). For conciseness, we will generally drop explicit references to vertices $V \cup A^{\text{Ch}}$, and denote extended graph of $\mathcal{G}(V)$ by \mathcal{G}^e . Extended graphs as we define them here are straightforward generalizations of those presented in [14], where they only consider “node copies” of a single “treatment” variable, whereas here extended graphs have “copies” corresponding to every parent-child relationship of a set of treatments A .

The edges $A_i \rightarrow A_i^j$ in \mathcal{G}^e are understood to represent *deterministic* relationships. More precisely, we associate a causal model with \mathcal{G}^e as follows. For \mathcal{G} we had associated a set of potential outcomes \mathbb{V} , and for \mathcal{G}^e we have \mathbb{V}^e . For every $V_i(\text{pa}_i) \in \mathbb{V}$, we let $V_i(\text{pa}_i)$ be in \mathbb{V}^e . Note that this is well-defined, since V_i in \mathcal{G} and \mathcal{G}^e share the number of parents, and the parent sets for every V_i share state spaces. In addition, for every $A_i^j \in A^{\text{Ch}}$, we let $A_i^j(a_i)$ for $a_i \in \mathfrak{X}_{A_i}$ be in \mathbb{V}^e . By assumption, every $A_i^j \in A^{\text{Ch}}$ has a single parent A_i , and we further require that $p(A_i^j(a_i))$ is a deterministic density, with $p(A_i^j(a_i) = a_i) = 1$. To fix intuitions, consider the example of Pearl’s discussed in [14]. They consider an analysis where A_i corresponds to smoking status, and affects hypertensive status V_j as well as myocardial infarction status V_k through nicotine A_i^j and non-nicotine A_i^k components respectively. The relationships $A_i \rightarrow A_i^j$ and $A_i \rightarrow A_i^k$ are deterministic relationships between smoking and exposure to nicotine/non-

nicotine components. [14] go on to consider potential outcomes of the form $V_k(a_i^j, a_i^k)$ (where the “node copies” A_i^j and A_i^k are assigned to perhaps different values) inspired by a hypothetical intervention on the nicotine components of cigarette exposure that fixes non-nicotine components at some reference value (e.g., a new nicotine-free cigarette). In this case, the path-specific effect of smoking on outcome via nicotine components is easy to write down and identify, at the price of introducing new variables and deterministic relationships into the model.

We now show the following two results. First, we show that an edge-consistent $V(\pi, a, a')$ may be represented without loss of generality by a counterfactual response to an intervention on a subset of A^{Ch} in \mathcal{G}^e with the causal model defined above. Second, we show that this response is identified by the same functional (7).

Given an edge consistent $V(\pi, a, a')$, define \mathcal{G}^e via $A \subseteq V$. We define a^π that assigns a_i to $A_i^j \in A^{\text{Ch}}$ if $A_i \rightarrow V_j$ in $\mathcal{G}(V)$ is in π , and assigns a'_i to $A_i^j \in A^{\text{Ch}}$ if $A_i \rightarrow V_j$ in $\mathcal{G}(V)$ is not in π . The resulting set of counterfactuals $V(a^\pi)$ is well defined in the model for \mathbb{V}^e , and we have the following result, proved in the Supplement.

Proposition 9 *Fix an element $p(\mathbb{V})$ in the causal model for a DAG $\mathcal{G}(V)$, and consider the corresponding element $p^e(\mathbb{V}^e)$ in the restricted causal model associated with a DAG $\mathcal{G}^e(V \cup A^{\text{Ch}})$. Then $p(V) = p^e(V \cup A^{\text{Ch}})$ and $p(V(\pi, a, a')) = p^e(V(a^\pi))$.*

Corollary 1 *Given an extended DAG \mathcal{G}^e ,*

$$p(V(a^\pi)) = \prod_{i=1}^K p^e(V_i | a^\pi \cap \text{pa}_i, \text{Pa}_i^{\mathcal{G}^e} \setminus A).$$

Proof: This follows from Proposition 9, and the fact that the functional in (7) in $p(V)$ is equal to $\prod_{i=1}^K p^e(V_i | a^\pi \cap \text{pa}_i, \text{Pa}_i^{\mathcal{G}^e} \setminus A)$ in $p^e(V \cup A^{\text{Ch}})$. \square

In the causal models derived from DAGs with unobserved variables (e.g., $\mathcal{G}(V \cup H)$), identification of distributions on potential outcomes such as $p(V(a))$ or $p(V(\pi, a, a'))$ may be stated without loss of generality on the latent projection ADMG $\mathcal{G}(V)$. A complete algorithm for identification of path-specific effects in hidden variable models was given in [16] and presented in a more concise form in [19]. We describe this form in detail in the Supplement. We also note (and prove in the Supplement) that the latent projection and the extended graph operations commute.

We now show that identification theory for $p(V(\pi, a, a'))$ in latent projection ADMGs $\mathcal{G}(V)$ may be restated, without loss of generality, in terms of identification of $p(V(a^\pi))$ in $\mathcal{G}^e(V \cup A^{\text{Ch}})$.

Proposition 10 *For any $Y \subseteq V$, $p(Y(\pi, a, a'))$ is identified in the ADMG $\mathcal{G}(V)$ if and only if $p(Y(a^\pi))$ is identified*

in the ADMG $\mathcal{G}^e(V, A^{\text{Ch}})$. Moreover if $p(Y(a^\pi))$ is identified, it is by the same functional as $p(Y(\pi, a, a'))$.

Note that this Proposition is a generalization of Corollary 1 from DAGs to latent projection ADMGs. The proof of this claim, and all claims in the next section, are given in the Supplement.

6 Identification of Conditional PSEs

Having established that we can identify path-specific effects by working with potential outcomes derived from the \mathcal{G}^e model, we turn to the identification of conditional path-specific effects using the po-calculus. In [17], the authors present the conditional identification (IDC) algorithm for identifying quantities of the form $p(Y(x)|W(x))$ (in our notation), given an ADMG. Since conditional path-specific effects correspond to exactly such quantities defined on the extended model \mathcal{G}^e , we can leverage their scheme for our purposes. The idea is to reduce the conditional problem, identification of $p(Y(a^\pi)|W(a^\pi))$, to an unconditional (joint) identification problem for which a complete identification algorithm already exists.

The algorithm has three steps: first, exhaustively apply Rule 2 of the po-calculus to reduce the conditioning set as much as possible; second, identify the relevant joint distribution using Proposition 10 and the complete algorithm in [19]; third, divide that joint by the marginal distribution of the remaining conditioning set to yield the conditional path-specific potential outcome distribution. The procedure is presented formally as Algorithm 1, with the subroutine corresponding to Proposition 10 named PS-ID.

Note that we make use of SWIGs defined from extended graphs, e.g., $\mathcal{G}^e(a^\pi, z)$. Beginning with \mathcal{G}^e the SWIG $\mathcal{G}^e(a^\pi, z)$ is constructed by the usual node-splitting operation: split nodes Z and A_i^z into random and fixed halves, where A_i^z is has fixed copy a if $A_i \rightarrow V_j$ in $\mathcal{G}(V)$ is in π , and a'_i if $A_i \rightarrow V_j$ in $\mathcal{G}(V)$ is not in π . Relabeling of random nodes proceeds as previously described.

The following two results are adapted from [17]; they are simply translated into potential outcomes and applied to extended graphs \mathcal{G}^e .

Proposition 11 *If $(Y(x, z) \perp\!\!\!\perp Z(x, z) \mid W(x, z))_{\mathcal{G}^e(x, z)}$ and $T \subseteq W$ then $(Y(x, t) \perp\!\!\!\perp T(x, t) \mid Z(x, t), W_1(x, t))_{\mathcal{G}^e(x, t)}$ if and only if $(Y(x, z, t) \perp\!\!\!\perp T(x, z, t) \mid W_1(x, z, t))_{\mathcal{G}^e(x, z, t)}$, where $W_1 = W \setminus T$.*

Corollary 2 *For any $\mathcal{G}^e(x)$ and any conditional distribution $p(Y(x)|W(x))$, there exists a unique maximal set $Z(x) = \{Z_i(x) \in W(x) \mid p(Y(x)|W(x)) = p(Y(x, z_i)|W(x, z_i) \setminus \{Z_i(x, z_i)\})\}$ such that Rule 2 applies for $Z(x, z)$ in $\mathcal{G}^e(x, z)$ for $p(Y(x, z)|W(x, z))$.*

Algorithm 1 PS-IDC($Y, a^\pi, W, \mathcal{G}^e$)

Input: outcome Y , path-specific setting a^π , conditioning set W , and graph \mathcal{G}

Output: $p(Y(a^\pi)|W(a^\pi))$

1: **if** $\exists Z \in W$ s.t.

$(Y(a^\pi, z) \perp\!\!\!\perp Z(a^\pi, z) \mid W(a^\pi, z))_{\mathcal{G}^e(a^\pi, z)}$

return PS-IDC($Y, a^\pi \cup z, W \setminus Z, \mathcal{G}^e$)

2: **else** let $p'(Y(a^\pi), W(a^\pi)) \leftarrow$ PS-ID($Y \cup W, a^\pi, \mathcal{G}^e$)

return $p'(Y(a^\pi), W(a^\pi)) / \sum_y p'(Y(a^\pi), W(a^\pi))$

The following is similar to Theorem 6 in [17], but extended to path-specific queries in extended graphs. The proof is in the Supplement.

Theorem 2 *Let $p(Y(\pi, a, a') \mid W(\pi, a, a'))$ be a conditional path-specific distribution in the causal model for \mathcal{G} , and let $p(Y(a^\pi) \mid W(a^\pi))$ be the corresponding distribution in the extended causal model for $\mathcal{G}^e(V \cup A^{\text{Ch}})$. Let Z be the maximal subset of W such that $p(Y(a^\pi) \mid W(a^\pi)) = p(Y(a^\pi, z) \mid W(a^\pi, z) \setminus Z(a^\pi, z))$. Then $p(Y(a^\pi) \mid W(a^\pi))$ is identifiable in \mathcal{G}^e if and only if $p(Y(a^\pi, z), W(a^\pi, z) \setminus Z(a^\pi, z))$ is identifiable in \mathcal{G}^e .*

We then have by Corollary 2, Theorem 2, and completeness of the identification algorithm for path-specific effects [19]:

Theorem 3 *Algorithm 1 is complete.*

As an example, $p(Y(a, M(a')))$ is identified from $p(C, A, M, Y)$ in the causal model in Fig. 1 (a), via

$$\sum_M \frac{\sum_C p(Y, M \mid a, C)p(C) \sum_C p(M \mid a', C)p(C)}{\sum_C p(M \mid a, C)p(C)}.$$

However $p(Y(a, M(a'))|C)$ is not identified, since PS-IDC concludes $p(Y(a, M(a')), C)$ must first be identified, and this joint distribution is not identified via results in [16]. On the other hand, $p(Y(a, M(a'))|C)$ is identified from $p(C, A, M, Y)$ in a seemingly similar graph in Fig. 1 (b), via $\sum_M p(Y \mid M, a, C)p(M \mid a', C)$.

7 Conclusion

In this paper we introduced the potential outcomes calculus, a generalization of do-calculus that applies to arbitrary potential outcomes. We have shown that potential outcome calculus is equivalent to Pearl's do-calculus for standard interventional quantities, and is a logical consequence of the properties of consistency and causal irrelevance, as well as the global Markov property associated with SWIGs. Finally, we used the potential outcomes calculus to give a sound and complete algorithm for conditional distributions defined on potential outcomes associated with path-specific effects. This algorithm may be viewed as a path-specific generalization of the identification algorithm for conditional interventional distributions in [17].

8 Acknowledgments

The authors would like to thank the American Institute of Mathematics for supporting this research via the SQuaRE program. This project is sponsored in part by the National Institutes of Health grant R01 AI127271-01 A1, and the Office of Naval Research grants N00014-18-1-2760 and N00014-15-1-2672. The authors would like to thank James M. Robins for helpful discussions.

References

- [1] Yimin Huang and Marco Valtorta. Pearl’s calculus of interventions is complete. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pages 217–224. AUAI Press, 2006.
- [2] Razieh Nabi, Daniel Malinsky, and Ilya Shpitser. Learning optimal fair policies. *arXiv preprint arXiv:1809.02244*, 2018.
- [3] Razieh Nabi and Ilya Shpitser. Fair inference on outcomes. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pages 1931–1940. AAAI Press, 2018.
- [4] Jerzy Neyman. On the application of probability theory to agricultural experiments: essay on principles (1923), section 9. Reprinted in English, with Discussion. *Statistical Science*, pages 463–480, 1990.
- [5] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan and Kaufmann, San Mateo, 1988.
- [6] Judea Pearl. A probabilistic calculus of actions. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence (UAI-94)*, pages 452–462. Morgan Kaufmann, 1994.
- [7] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–709, 1995.
- [8] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009.
- [9] Thomas S. Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1):145–157, 2003.
- [10] Thomas S. Richardson, Robin J. Evans, James M. Robins, and Ilya Shpitser. Nested Markov properties for acyclic directed mixed graphs. *preprint: arXiv:1701.06686*, 2017.
- [11] Thomas S. Richardson and Jamie M. Robins. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *preprint: http://www.csss.washington.edu/Papers/wp128.pdf*, 2013.
- [12] James M. Robins. A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512, 1986.
- [13] James M. Robins, Miguel A. Hernan, and Uwe Siebert. Effects of multiple interventions. In *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, volume 2, chapter 28, pages 2191–2230. World Health Organization, 2004.
- [14] James M. Robins and Thomas S. Richardson. Alternative graphical causal models and the identification of direct effects. In *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*. Oxford University Press, 2011.
- [15] D. B. Rubin. Causal inference and missing data (with discussion). *Biometrika*, 63:581–592, 1976.
- [16] Ilya Shpitser. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cognitive Science (Rumelhart Special Issue)*, 37:1011–1035, 2013.
- [17] Ilya Shpitser and Judea Pearl. Identification of conditional interventional distributions. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pages 437–444. AUAI Press, 2006.
- [18] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence (AAAI-06)*, pages 1219–1226. AAAI Press, 2006.
- [19] Ilya Shpitser and Eli Sherman. Identification of personalized effects associated with causal pathways. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI-18)*. AUAI Press, 2018.
- [20] Ilya Shpitser and Eric J. Tchetgen Tchetgen. Causal inference with a graphical hierarchy of interventions. *Annals of Statistics*, 44(6):2433–2466, 2016.
- [21] Lu Zhang, Yongkai Wu, and Xintao Wu. A causal framework for discovering and removing direct and indirect discrimination. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, pages 3929–3935, 2017.