Supplement For: 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

Note that the Propositions with numbers ≤ 11 are ones that appear in the body of the main paper, and Propositions with higher numbers are only stated here in the Supplement.

Utility Results

First we define d-separation and m-separation for reference. Given two elements $V_i, V_j \in V$, and $X \subseteq V \setminus \{V_i, V_j\}$, we say that a path from V_i to V_j is blocked by X if $*\!\!\rightarrow X_k \rightarrow$, $\leftarrow X_k \leftarrow\!\!*$, or $\leftarrow X_k \rightarrow$ exists on the path, where $X_k \in X$, or if $*\!\!\rightarrow V_h \leftarrow\!\!*$ exists on the path, where $V_h \notin X$ and $\text{De}_h \cap X = \emptyset$. $*$ stands for either an arrowhead or tail edge-mark, allowing for bidirected edges. We say Y is m-separated from Z given X in $\mathcal{G}(V)$ if every path from an element of Y to an element of Z is blocked by X in $\mathcal{G}(V)$. d-separation is the special case where all edges are directed.

Proposition 4 *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 $A'' \supseteq A$, $(Y(a''), a' \perp\!\!\!\perp Z(a'') \mid X(a''))_{\mathcal{G}(a'')}$.*

Proof: Assume a m-connected path from an element in Y or a' to $Z(a)$ in $\mathcal{G}(a'')$. If this path does not intersect an element in $A'' \setminus A$, then it is also present in $\mathcal{G}(a)$. If this path does intersect $A'' \setminus A$, any element $A_i \in A''$ on this path cannot contain an outgoing edge on the path (since such edges do not exist in $\mathcal{G}(a'')$. As a result, all edges on the path also exist in $\mathcal{G}(a)$. Since the conditioning set is the same in both cases, the path is m-connected in $\mathcal{G}(a)$, which is a contradiction. □

Proposition 12 *Given a DAG $\mathcal{G}(V \cup H)$, $\mathcal{G}(V)(a) = \mathcal{G}(a)(V)$.*

Proof: By definition, both graphs agree on the set of random and fixed vertices. Note that $\mathcal{G}$ and $\mathcal{G}(a)$ have the same set of edges, and that $A \cap H = \emptyset$. Consequently, any edge from V_i to V_j in $\mathcal{G}(V)(a)$ corresponds to a marginally d-connected path from V_i to V_j with all intermediate vertices in H in $\mathcal{G}(V \cup H)$. And similarly, such a path exists for any edge from V_i to V_j in $\mathcal{G}(a)(V)$. This establishes the bijection between edges. □

Independence statements implied by d-separation on observed variable subsets of $\mathcal{G}(V \cup H)(a)$, for $A \subseteq V$ translate into m-separation statements of $\mathcal{G}(V)(a)$.

Proposition 13 *For any disjoint subsets $Y(a), Z(a), X(a)$ of $V(a)$ and a subset a' of a,*

$$(Y(a), a' \perp\!\!\!\perp Z(a) \mid X(a))_{\mathcal{G}(V \cup H)(a)} \Rightarrow (Y(a), a' \perp\!\!\!\perp Z(a) \mid X(a))_{\mathcal{G}(V)(a)}.$$

Proof: This follows immediately from the fact that m-separation statements in a latent projection ADMG $\mathcal{G}(V)$ are in a one-to-one correspondence with d-separation statements in a DAG $\mathcal{G}(V \cup H)$ on V, and the SWIG global Markov property. □

A Complete Identification Algorithm For Path-Specific Counterfactual Distributions In Hidden Variable Causal Models

Here we introduce a concise formulation of the complete identification algorithm for edge-consistent path-specific counterfactual distributions given in [6] via kernels, conditional graphs, and the fixing operation.

Kernels, Conditional Graphs, and Fixing

A kernel $q_V(V \mid W)$ is a mapping from $\mathfrak{X}_W$ to densities over V. Given $A \subseteq V$, we define conditioning and marginalization in the usual way:

$$q_V(A|W) \equiv \sum_{V \setminus A} q_V(V|W);\ q_V(V \setminus A|A, W) \equiv \frac{q_V(V|W)}{q_V(A|W)}.$$

A conditional graph $\mathcal{G}(V, W)$ is a graph with two types of vertices, random V and fixed W, with the property that for any fixed vertex in W, its set of parents is empty.[1] We will consider conditional ADMGS (CADMGs), or conditional DAGs (CDAGs) as a special case. A SWIG $\mathcal{G}(V(a))$ may be viewed as a conditional graph of the form $\mathcal{G}(V(a), a)$, where we denote the set of fixed vertices by a.

For a CADMG $\mathcal{G}(V, W)$, and $V_i \in V$, define

$$\mathrm{Dis}_i^{\mathcal{G}} \equiv \{V_j \mid V_j \leftrightarrow \ldots \leftrightarrow V_i \text{ in } \mathcal{G}\} \quad \text{(district of } V_i\text{)}.$$

Note that districts are only defined for, and may only contain, random vertices in V not fixed vertices in W. The set of districts in $\mathcal{G}$ is denoted by $\mathcal{D}(\mathcal{G})$.

A vertex $V_i \in V$ in a CADMG $\mathcal{G}(V, W)$ is said to be fixable if $\mathrm{De}_i \cap \mathrm{Dis}_i = \emptyset$. For such a vertex, define the operator $\phi_i(\mathcal{G})$ that yields a new CADMG $\mathcal{G}(V \setminus \{V_i\}, W \cup \{V_i\})$, obtained by removing all edges with arrowheads into V_i, and keeping all other edges in $\mathcal{G}(V, W)$.

Given a CADMG $\mathcal{G}(V, W)$, and a kernel $q_V(V \mid W)$, if V_i is fixable, define the operator $\phi_i(q_V; \mathcal{G})$ as yielding a new kernel

$$q_{V\setminus\{V_i\}}(V \setminus \{V_i\}|W \cup \{V_i\}) \equiv \frac{q_V(V \mid W)}{q_V(V_i \mid \mathrm{Mb}_i^{\mathcal{G}}, W)},$$

where $\mathrm{Mb}_i^{\mathcal{G}}$, the *Markov blanket* of V_i in $\mathcal{G}$, is defined to be $\mathrm{Dis}_i^{\mathcal{G}} \cup \{\mathrm{Pa}_j^{\mathcal{G}} \mid V_j \in \mathrm{Dis}_i^{\mathcal{G}}\}$.

A set of vertices $Z \subseteq V$ is said to be *fixable* in $\mathcal{G}(V, W)$, if there exists a fixable sequence $Z_1, \ldots, Z_k$ on vertices in Z such that Z_1 is fixable in $\mathcal{G}$, Z_2 is fixable in $\phi_1(\mathcal{G})$, Z_3 is fixable in $\phi_2(\phi_1(\mathcal{G}))$, and so on. Given a sequence α_Z for elements in Z, we define $\phi_{\alpha_Z}(\mathcal{G})$ and $\phi_{\alpha_Z}(q_V; \mathcal{G})$ in the natural way by operator composition. For any two valid fixing sequences α_Z, β_Z for a fixable set Z, $\phi_{\alpha_Z}(\mathcal{G}) = \phi_{\beta_Z}(\mathcal{G})$. Hence, for a fixable Z, we define $\phi_Z(\mathcal{G})$ to mean "fix elements in Z in $\mathcal{G}$ by any fixable sequence."

Given a CADMG $\mathcal{G}(V, W)$, if $Z \subseteq V$ is fixable, then $R \equiv V \setminus Z$ is called a reachable set. A reachable set R such that $\mathcal{D}(\phi_Z(\mathcal{G}(V, W)))$ contains a single element is called *intrinsic*. If there exists a set of kernels

$$\{q_D(D|\,\mathrm{Pa}_D, W)|D \text{ is intrinsic in } \mathcal{G}(V, W)\},$$

where $\mathrm{Pa}_D \equiv \bigcup_{V_i \in D}\{\mathrm{Pa}_i \setminus D \mid V_i \in D\}$, such that for all fixable sets Z in $\mathcal{G}(V, W)$, and all fixable sequences α_Z, we have

$$\phi_{\alpha_Z}(q_V(V|W); \mathcal{G}(V, W)) = \prod_{D \in \mathcal{D}(\phi_Z(\mathcal{G}(V,W)))} q_D(D|\,\mathrm{Pa}_D, W),$$

we say $q_V(V \mid W)$ is in the *nested Markov model* of $\mathcal{G}(V, W)$.

[1] Note that some elements of V may have an empty parent set as well.

For any such $q_V(V \mid W)$, it can be shown that for any fixable Z in $\mathcal{G}(V, W)$, and any fixable sequences α, β for Z, $\phi_{\alpha_Z}(q_V(V|W); \mathcal{G}(V, W)) = \phi_{\beta_Z}(q_V(V|W); \mathcal{G}(V, W))$. As a result, we write $\phi_Z(q_V(V|W); \mathcal{G}(V, W))$ to mean "fix elements in Z in $q_V(V|W)$ using any fixable sequence."

Moreover, we have

$$\{q_D(D|\,\mathrm{Pa}_D, W)|D \text{ is intrinsic in } \mathcal{G}(V, W)\}$$
$$= \{\phi_{V\setminus D}(q_V(V \mid W); \mathcal{G}(V, W))| \text{ is intrinsic in } \mathcal{G}(V, W)\}.$$

We have the following important results.

Proposition 14 *If $q_{V\cup H}(V \cup H|W)$ is in the Markov model for the CDAG $\mathcal{G}(V \cup H, W)$, then $q_V(V|W) \equiv \sum_H q_{V\cup H}(V \cup H|W)$ is in the nested Markov model for the latent projection CADMG $\mathcal{G}(V, W)$.*

Proof: This is shown in [1]. □

The complete algorithm for an edge-consistent $p(Y(\pi, a, a'))$ for $Y \subseteq V$ is stated as follows.

Proposition 15 *Let $Y^* \equiv \mathrm{an}_{\mathcal{G}_{V\setminus A}}(Y)$. Then $p(Y(\pi, a, a'))$ is identified in $\mathcal{G}(V)$ if and only if for every $D \in \mathcal{D}(\mathcal{G}_{Y^*})$, $\mathrm{pa}_{\mathcal{G}}(D) \cap A$ are assigned to either a subset of a or a subset of a', and D is intrinsic in $\mathcal{G}(V)$. Moreover, if $p(Y(\pi, a, a'))$ is identified, we have*

$$p(Y(\pi, a, a')) = \sum_{Y^*\setminus Y} \prod_{D\in\mathcal{D}(\mathcal{G}_{Y^*})} \phi_{V\setminus D}(p(V); \mathcal{G}(V))\big|_{\tilde{a}_D}, \tag{8}$$

where $\tilde{a}_D$ is defined to be the appropriate subset of a associated with $\mathrm{pa}_{\mathcal{G}}(D) \cap A$ if those elements are assigned by the definition of $Y(\pi, a, a')$, and the appropriate subset of a' associated with $\mathrm{pa}_{\mathcal{G}}(D) \cap A$ otherwise.

Proof: This is shown in [6]. □

Note that the kernels $\phi_{V\setminus D}(.)$ are well defined by Proposition 14, since causal inference always starts with a causal model that implies a distribution that factorizes with respect to a (possibly hidden variable) DAG.

Remaining Proofs

Now we turn to proving results related to Sections 5 and 6 in the main paper.

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 $\mathcal{G}^e(V \cup A^{\mathrm{Ch}})$. Then $p(V) = p^e(V, A^{\mathrm{Ch}})$ and $p(V(\pi, a, a')) = p^e(V(a^\pi))$.*

Proof: By definition of the causal model for $\mathcal{G}$, we have

$$p(V(\pi, a, a') = v) = \sum_{\epsilon_i : f_i(a_{\mathrm{Pa}_i^\pi}, a'_{\mathrm{Pa}_i^{\overline{\pi}}}, v_{\mathrm{Pa}_i \setminus A}) = v_i} p(\epsilon_1, \ldots, \epsilon_k),$$

where for each V_i, Pa_i^{π} is the subset of $\mathrm{Pa}_i \cap A$ with an edge from Pa_i to V_i in π, and $\mathrm{Pa}_i^{\overline{\pi}}$ is the subset of $\mathrm{Pa}_i \cap A$ with an edge from Pa_i to V_i not in π. Similarly, by definition of the restricted causal model for $\mathcal{G}^e(V \cup A^{\mathrm{Ch}})$, we have

$$p^e(V(a^{\pi}) = v) = \sum_{\epsilon_i : f_i(a^{\pi}_{\mathrm{Pa}_i \cap A^{\mathrm{Ch}}}, v_{\mathrm{Pa}_i \setminus A}) = v_i} p(\epsilon_1, \ldots, \epsilon_k).$$

The equivalence follows immediately. Note that the same argument establishes $p(V) = p^e(V, A^{\mathrm{Ch}})$, by letting π be the empty set of paths, and $A = \emptyset$. □

Proposition 16 *Assume there exists elements $p_1(\mathbb{V}), p_2(\mathbb{V})$ in the causal model for $\mathcal{G}$ such that $p_1(V) = p_2(V)$, but $p_1(V(\pi, a, a')) \neq p_2(V(\pi, a, a'))$. Then $p(V(a^{\pi}))$ is not identified in the restricted causal model for $\mathcal{G}^e(V \cup A^{\mathrm{Ch}})$.*

Proof: Follows immediately by Proposition 9. □

We state formally our claim in the main paper that the latent projection and extended graph operations commute.

Proposition 17 *Fix a DAG $\mathcal{G}(V \cup H)$, and let $A \subseteq V$. Then $\mathcal{G}^e(V \cup A^{\mathrm{Ch}})$, the latent projection onto $V \cup A^{\mathrm{Ch}}$ of $\mathcal{G}^e(V \cup H \cup A^{\mathrm{Ch}})$ is equal to the extended graph $\mathcal{G}(V \cup A^{\mathrm{Ch}})^e$ applied to the latent projection $\mathcal{G}(V)$.*

Proof: By definition, the two graphs have the same vertices. That the two graphs share the same edges follows from the definition of $\mathcal{G}^e$, which stipulates that the only edge into each variable in A^{Ch} is from the corresponding variable in A, i.e., there are no directed paths from any H into any element of A^{Ch} not through some element of A. So, all bidirected edges induced by the latent projection operation are between vertices in V, which are shared between the two graphs. □

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 $\mathcal{G}^e(V, A^{\mathrm{Ch}})$. Moreover if $p(Y(a^{\pi}))$ is identified, we have*

$$p^e(Y(a^{\pi})) = \sum_{Y^* \setminus Y} \prod_{D \in \mathcal{D}(\mathcal{G}^e_{Y^*})} \phi_{(V \cup A^{\mathrm{Ch}}) \setminus D}(p^e(V, A^{\mathrm{Ch}}); \mathcal{G}^e)\big|_{\tilde{a}_D} \tag{9}$$

where $Y^* \equiv \mathrm{an}_{\mathcal{G}^e_{V \setminus A^{\mathrm{Ch}}}}(Y)$, and $\tilde{a}_D$ is defined to be the appropriate subset of a^{π} associated with $\mathrm{pa}_{\mathcal{G}}(D) \cap A^{\mathrm{Ch}}$.

Proof: Assume $p(Y(\pi, a, a'))$ is identified in $\mathcal{G}(V)$ via (8). The conclusion follows from Proposition 9, and the fact that the functional in (8) in $p(V)$ is equal to (9) in $p^e(V, A^{\mathrm{Ch}})$.

Assume $p(Y(\pi, a, a'))$ is not identified, and fix a witness of this fact, which is either a hedge or a district with a recanting set of parents in A. If the witness is a hedge, the construction in [5] yields $p_1(\mathbb{V})$ and $p_2(\mathbb{V})$, such that $p_1(V) = p_2(V)$, but $p_1(Y(\pi, a, a')) \neq p_2(Y(\pi, a, a'))$ If the witness is a recanting district, the construction in [3], described also in [6], yields $p_1(\mathbb{V})$ and $p_2(\mathbb{V})$, such that $p_1(V) = p_2(V)$, but $p_1(Y(\pi, a, a')) \neq p_2(Y(\pi, a, a'))$. In both cases, this immediately implies the conclusion by Corollary 16. □

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 $W_1 = W \setminus T$.*

Proof: The set of possible d-connecting paths from $Y(x,z,t)$ to $T(x,z,t)$ in $\mathcal{G}^e(x,z,t)$ is a subset of the set of possible d-connecting paths from $Y(x,t)$ to $T(x,t)$ in $\mathcal{G}^e(x,t)$. For any such path that exists in both graphs, if it is blocked by $W_1(x,t)$ in $\mathcal{G}^e(x,t)$, it will be blocked by $W_1(x,z,t)$ in $\mathcal{G}^e(x,z,t)$. If it is blocked by $Z(x,t)$ in $\mathcal{G}^e(x,t)$, the path will be blocked in $\mathcal{G}^e(x,z,t)$ by construction of $\mathcal{G}^e(x,z,t)$. If it is blocked by collider without $Z(x,t), W_1(x,t)$ descendants in $\mathcal{G}^e(x,t)$, the same will remain true in $\mathcal{G}^e(x,z,t)$. Thus, if $(Y(x,t) \perp\!\!\!\perp T(x,t) \mid Z(x,t), W_1(x,t))_{\mathcal{G}^e(x,t)}$, then $(Y(x,z,t) \perp\!\!\!\perp T(x,z,t) \mid W_1(x,z,t))_{\mathcal{G}^e(x,z,t)}$.

Now, assume for contradiction, $(Y(x,z,t) \perp\!\!\!\perp T(x,z,t) \mid W_1(x,z,t))_{\mathcal{G}^e(x,z,t)}$, but $(Y(x,t) \not\perp\!\!\!\perp T(x,t) \mid Z(x,t), W_1(x,t))_{\mathcal{G}^e(x,t)}$, with a witnessing d-connecting path from some $Y_1(x,t)$ to some $T_1(x,t)$. If this path is not a possible d-connecting path in $\mathcal{G}^e(x,z,t)$, it must contain a non-collider through an element of Z, and thus is blocked by $Z(x,t)$ in $\mathcal{G}^e(x,t)$. If this path is a possible d-connecting path in $\mathcal{G}^e(x,z,t)$ it must be blocked by a collider which contains no descendants in $W_1(x,z,t)$ in $\mathcal{G}^e(x,z,t)$, but remains open due to this collider containing descendants in $Z(x,t)$ in $\mathcal{G}^e(x,t)$.

But this implies the existence of a d-connecting path in $\mathcal{G}^e(x,t)$ from an element $Y_1(x,t)$ in $Y(x,t)$ to an element $Z_1(x,t)$ in $Z(x,t)$ given $W_1(x,t)$, and thus also given $W(x,t)$ (since no element in $T(x,t)$ will block this path by construction). Since we can choose $Z_1(x,t)$ to be the closest element in $Z(x,t)$ to $Y_1(x,t)$ involved in the witnessing path, we obtain that $Y_1(x,z) \not\perp\!\!\!\perp Z_1(x,z) \mid W(x,z)$, which is a contradiction. □

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 $p(Y(x,z)|W(x,z))$.*

Proof: Fix two maximal sets $Z_1(x)$ and $Z_2(x)$ such that Rule 2 applies for $Z(x,z)$ in $\mathcal{G}^e(x,z)$ for $p(Y(x,z)|W(x,z))$. If $Z_1(x) \neq Z_2(x)$, fix $T(x) \in Z_1(x) \setminus Z_2(x)$. By the previous proposition, Rule 2 applies for $Z_2(x) \cup T(x)$, contradicting our assumption. □

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 $\mathcal{G}^e$.*

Proof: The proof strategy follows that of the completeness argument in [4]. We expand the argument here to be more transparent. In addition, we must handle an additional case of non-identifiability that arises in mediation problems, that has to do with structures called *recanting districts* in [3].

If $p(Y(a^\pi, z), W(a^\pi, z) \setminus Z(a^\pi, z))$ is identified in $\mathcal{G}^e$, then $p(Y(a^\pi) \mid W(a^\pi))$ is identifiable in $\mathcal{G}^e$ since

$$\begin{aligned} p(Y(a^\pi)|W(a^\pi)) &= p(Y(a^\pi, z)|W(a^\pi, z) \setminus Z(a^\pi, z)) \\ &= \frac{p(Y(a^\pi, z), W(a^\pi, z) \setminus Z(a^\pi, z))}{p(W(a^\pi, z) \setminus Z(a^\pi, z))}. \end{aligned}$$

Now assume $p(Y(a^\pi, z), W(a^\pi, z) \setminus Z(a^\pi, z))$ is not identified in $\mathcal{G}^e$. Either $p(W(a^\pi, z))$ is identified or not. If $p(W(a^\pi, z))$ is identified, $p(Y(a^\pi, z)|W(a^\pi, z) \setminus Z(a^\pi, z))$ is identified if and only if $p(Y(a^\pi, z), W(a^\pi, z) \setminus Z(a^\pi, z))$ is. Since the latter is false by assumption, our conclusion follows.

Assume $p(W(a^\pi, z))$ is not identified. Let $\tilde{a} = a \cup z$, and $\tilde{\pi}$ be the set comprised of π and all outgoing directed edges from elements in Z. Then the distribution $p(W(a^\pi, z))$ is equal to $p(W(\tilde{a}^{\tilde{\pi}}))$, which in turn is equivalent to $p(W(\tilde{\pi}, \tilde{a}, a'))$.

$p(W(\tilde{\pi}, \tilde{a}, a'))$ could fail to be identified in the causal model for $\mathcal{G}$ for two reasons. Either there could exist a hedge structure [5] for $p(W(a))$, or there could exist a recanting district structure [3] in $\mathcal{D}(\mathcal{G}_{W^*})$, where $W^* \equiv \text{An}_W^{\mathcal{G}_{V \setminus A}}$. We consider these cases in turn.

If there exists a hedge structure, fix a district D in $\mathcal{D}(\mathcal{G}_{W^*})$, where $W^* \equiv \text{An}_W^{\mathcal{G}_{V \setminus A}}$, such that there is a larger district D' containing D that forms the hedge structure with D. Further, find the minimal subset W' of W such that the set of all childless vertices in the hedge structure (contained in D') is in $\text{An}_{W'}^{\mathcal{G}_{V \setminus A}}$. Let H be the smallest set of vertices that contains W', D', and such that the set of childless vertices in the hedge structure is in $\text{An}_{W'}^{\mathcal{G}_H}$.

Assume without loss of generality that each vertex in $\mathcal{G}_H$ has at most one child. We construct elements $p_1(\mathbb{H})$ and $p_2(\mathbb{H})$ in the causal model in $\mathcal{G}_H$ as follows. In $p_1(\mathbb{H})$ each structural equation is a bit parity function of the parents, and each bidirected arc corresponds to a binary latent common parent where each such latent is involved in precisely two functions. Moreover, each such latent variable ϵ_{ij} that is a parent of V_i and V_j is drawn from a uniform distribution $p(\epsilon_{ij})$. In $p_2(\mathbb{H})$ the same is true, except no element in $D' \setminus D$ is involved in the structural equation for any element in D, and no ϵ_{ij} that is a parent of an element in $D' \setminus D$ and an element in D exists. It has been shown in [5] that if $p_1(\mathbb{H})$ and $p_2(\mathbb{H})$ are constructed in this way, they induce $p_1(H), p_1(W'(a_{H \cap A}))$ and $p_2(H), p_2(W'(a_{H \cap A}))$ respectively, such that $p_1(H) = p_2(H)$ (i.e., the induced observational distributions are the same), but $p_1(W'(a_{H \cap A})) \neq p_2(W'(a_{H \cap A}))$ (i.e., the induced potential outcome distributions are distinct).

Specifically, let R be the set of childless vertices in $\mathcal{G}_{D'}$. Then it has been shown that $p_1(D') = p_2(D')$ is a distribution uniform over any assignment to D' such that the number of 1 values in R is even. At the same time, $p_1(R(a_{H \cap A}))$ is a uniform distribution over assignments with even number of 1 values, while $p_2(R(a_{H \cap A}))$ is a uniform distribution. Since each element in $H \setminus R$ has a single parent in $\mathcal{G}_H$, the bit parity function for those elements simply reduces to the identity function. Note that more general structural equations suffice for the argument, as long as the linear transformation that maps $p(D'(a_{H \cap A}))$ to $p(W'(a_{H \cap A})) \equiv \sum_{D'} p(W'(a_{H \cap A}) \mid D'(a_{H \cap A}))p(D'(a_{H \cap A}))$ is one to one.

Consider a path π in $\mathcal{G}(a)$ from some element W_i in W' to an element Y_j in Y, such that W_i is m-connected to Y given W, and the edge on the path adjacent to W_i has an arrowhead into W_i (Pearl called such paths *backdoor paths*). Such a path must exist by construction of W. In addition, consider the smallest subset W'' of W such that W_i is m-connected to Y_j given W'' in $\mathcal{G}(a)$. Pick the smallest set H' containing H such that the above m-connection statement holds in $\mathcal{G}(a)_{H'}$. We now extend $p_1(\mathbb{H})$ and $p_2(\mathbb{H})$ to $p_1(\mathbb{H}')$ and $p_2(\mathbb{H}')$ to show $p(Y_j(a_{H' \cap A}) \mid W'(a_{H' \cap A}))$ is not identified.

We have three base cases. The first case assumes the first node Z_j on π not in H is a parent of an element Z_i in H. Let the structural equation corresponding to Z_i be the bit parity function of all its parents in $\mathcal{G}_{H'}$, including Z_j in both $p_1(\mathbb{H}')$ and $p_2(\mathbb{H}')$, and let $p(Z_j)$ be the uniform distribution on a binary variable.

In this case, the observed data distributions are $p_1(H \mid Z_j)p_1(Z_j)$ and $p_2(H \mid Z_j)p_2(Z_j)$. $p_1(Z_j) = p_2(Z_j)$ by construction. Next, note that $p_1(H \mid Z_j = 0) = p_2(H \mid Z_j = 0)$ equal to the distributions $p_1(H) = p_2(H)$ given in the previous construction. Specifically these distributions are uniform on all assignments to R with an even number of 1 values. By symmetry, $p_1(H \mid Z_j = 1) = p_2(H \mid Z_j = 1)$, with the distributions being uniform on all assignments to R with an odd number of 1 values. By above construction and results in [5], $p_1(H(a_{H \cap A}) \mid Z_j(a_{H \cap A}) = 0) = p_1(H \mid Z_j = 0)$, while $p_2(H(a_{H \cap A}) \mid Z_j(a_{H \cap A}) = 0)$ is a uniform distribution. Since $p_1(Z_j(a_{H\cap A})) = p_1(Z_j) = p_2(Z_j) = p_2(Z_j(a_{H\cap A}))$, we have that $p_1(Z_j(a_{H\cap A}), R(a_{H\cap A}))$ only has positive probability if the number of 1 values in $\{Z_j\} \cup R$ is even, while $p_2(Z_j(a_{H\cap A}), R(a_{H\cap A}))$ is a uniform distribution. This implies $p_1(Z_j(a_{H\cap A}) = 0 \mid R(a_{H\cap A}) = 0) = 1$, while $p_2(Z_j(a_{H\cap A}) = 0 \mid R(a_{H\cap A}) = 0) < 1$, which establishes the base case.

The second case assumes the first node Z_j on π not in H is a child of an element Z_i in H. We also consider the third case where $Y_j \in H$, here by letting $Y_j = Z_i$. If $p(Z_j(a_{H'\cap A}) \mid W'(a_{H'\cap A}))$ (or $p(Y_j(a_{H'\cap A}) \mid W'(a_{H'\cap A})))$ is not identified, we are done. Otherwise, we assume $p(Z_j(a_{H'\cap A}) \mid W'(a_{H'\cap A}))$ is identified. Consider the edge subgraph $\mathcal{G}'_{H'}$ of $\mathcal{G}_{H'}$ that lacks the outgoing directed edges from Z_i within H.

Since the childless vertices in the hedge structure are in $\mathrm{An}^{\mathcal{G}_H}_{W'}$, if Z_i is not in the hedge structure in H, it must be on a directed path in $\mathcal{G}_H$ from some childless vertex in the hedge structure to an element of W'. Since we assumed each vertex in $\mathcal{G}_H$ has at most one child, removing the outgoing arrow from Z_i in $\mathcal{G}_{H'}$ results in $\mathcal{G}'_{H'}$ containing the hedge structure for $p(Z_j(a_{H'\cap A}), W''(a_{H'\cap A}))$, where $W'' = W' \setminus \{W_i\}$ and W_i is $W' \cap \mathrm{De}_{\mathcal{G}_{H'}}(Z_j)$.

If $p(W''(a_{H'\cap A}))$ is identified, we are done, since we established the base case where $p(Z_j(a_{H'\cap A}) \mid W''(a_{H'\cap A}))$ is not identified. If $p(W''(a_{H'\cap A}))$ is not identified, note that W'' is a strictly smaller set then W', and we restart the base case argument, finding a hedge or a recanting district for this smaller set, constructing a new set H, and a new backdoor path to an element in Y. Since the new subset of W is strictly smaller, we can only do this a finite number of times before encountering another base case.

If Z_i is in the hedge structure in H, then the resulting graph $\mathcal{G}'_{H'}$ contains a hedge structure for $p(Z_j(a_{H'\cap A}), W'(a_{H'\cap A}))$ with the set of childless vertices of the previous hedge and also Z_i (since it is now childless in H). Given the hedge construction above, we have $p_1(Z_j(a_{H'\cap A}) = 0 \mid W'(a_{H'\cap A}) = 0) < 1$, while $p_2(Z_j(a_{H'\cap A}) = 0 \mid W'(a_{H'\cap A}) = 0) = 1$, and we are done.

We now consider the inductive cases on the path π. Consider Z_k and Z_{k+1} on the path, where Z_{k+1} is closer to Y_j on the path. We have the following cases.

If Z_{k+1} is a parent of Z_k, or Z_{k+1} is a child of Z_k, then in $\mathcal{G}'_{H'}$:

$$p_1(Z_{k+1}(a_{H'\cap A})|W'(a_{H'\cap A})) = \sum_{Z_k} p_1(Z_k(a_{H'\cap A})|W'(a_{H'\cap A}))p_1(Z_{k+1}(a_{H'\cap A})|Z_k(a_{H'\cap A}))$$

$$p_2(Z_{k+1}(a_{H'\cap A})|W'(a_{H'\cap A})) = \sum_{Z_k} p_2(Z_k(a_{H'\cap A})|W'(a_{H'\cap A}))p_2(Z_{k+1}(a_{H'\cap A})|Z_k(a_{H'\cap A})).$$

Let

$$p_1(Z_{k+1}(a_{H'\cap A})|Z_k(a_{H'\cap A})) = p_2(Z_{k+1}(a_{H'\cap A})|Z_k(a_{H'\cap A})).$$

Then we have

$$p_1(Z_{k+1}(a_{H'\cap A})|W'(a_{H'\cap A})) \neq p_2(Z_{k+1}(a_{H'\cap A})|W'(a_{H'\cap A}))$$

if and only if

$$p_1(Z_k(a_{H'\cap A})|W'(a_{H'\cap A})) \neq p_2(Z_k(a_{H'\cap A})|W'(a_{H'\cap A})).$$

These latter distributions are not equal in $p_1(\mathbb{H}')$ and $p_2(\mathbb{H}')$ by the inductive hypothesis.

If Z_{k+1} is a sibling of Z_k, we repeat the above two cases, since this case may be rephrased without loss of generality in terms of an unobserved variable H_k that is a parent of both Z_{k+1} and Z_k.

If Z_{k+1} and Z_k are both parents of a variable C_k which is an ancestor of element W_k in W', we have:

$$p_1(Z_{k+1}(a_{H'\cap A})|W_k(a_{H'\cap A}), W'(a_{H'\cap A})) = \sum_{Z_k} p_1(Z_{k+1}|W_k, Z_k)\times \frac{p_1(W_k|Z_k)}{\sum_{Z_k} p_1(W_k|Z_k)p_1(Z_k|W'(a_{H'\cap A}))} p_1(Z_k|W'(a_{H'\cap A}))$$

$$p_2(Z_{k+1}(a_{H'\cap A})|W_k(a_{H'\cap A}), W'(a_{H'\cap A})) = \sum_{Z_k} p_2(Z_{k+1}|W_k, Z_k)\times \frac{p_2(W_k|Z_k)}{\sum_{Z_k} p_2(W_k|Z_k)p_2(Z_k|W'(a_{H'\cap A}))} p_2(Z_k|W'(a_{H'\cap A}))$$

Assume $W_k = C_k$. We must choose

$$p_1(Z_{k+1}, W_k \mid Z_k) = p_2(Z_{k+1}, W_k \mid Z_k)$$

such that

$$p_1(Z_{k+1}(a_{H'\cap A})|W_k(a_{H'\cap A}), W'(a_{H'\cap A})) \neq p_2(Z_{k+1}(a_{H'\cap A})|W_k(a_{H'\cap A}), W'(a_{H'\cap A}))$$

if

$$p_1(Z_k|W'(a_{H'\cap A})) \neq p_2(Z_k|W'(a_{H'\cap A})),$$

(which is true by the inductive hypothesis).

For a fixed W_k, we have 5 degrees of freedom: $p(Z_{k+1})$, $p(W_k|Z_{k+1}, Z_k)$, $p(W_k|Z_{k+1}, 1 - Z_k)$, $p(W_k|1 - Z_{k+1}, Z_k)$, and $p(W_k|1 - Z_{k+1}, 1 - Z_k)$.

It suffices to specify these in such a way that the linear mapping induced by

$$p(Z_{k+1} \mid W_k, Z_k) = \frac{p(Z_{k+1})p(W_k \mid Z_{k+1}, Z_k)}{\sum_{Z_{k+1}} p(Z_{k+1})p(W_k \mid Z_{k+1}, Z_k)}$$

is a one-to-one mapping, and for some W_k, $c = p_1(W_k \mid Z_k) = p_2(W_k \mid Z_k)$, and $k = p_1(W_k \mid 1 - Z_k) = p_2(W_k \mid 1 - Z_k)$ are chosen such that

$$\frac{p_1(Z_k|W'(a_{H'\cap A}))}{p_2(Z_k|W'(a_{H'\cap A}))} \neq \frac{k + p_1(Z_k|W'(a_{H'\cap A}))(c-k)}{k + p_2(Z_k|W'(a_{H'\cap A}))(c-k)},$$

for some $p_1(Z_k|W'(a_{H'\cap A})) \neq p_2(Z_k|W'(a_{H'\cap A}))$. But there are sufficient degrees of freedom to satisfy both properties. In particular, we can choose $p_1(Z_k|W'(a_{H'\cap A}))$ and $p_2(Z_k|W'(a_{H'\cap A}))$ to be distinct one-to-one mappings (since these are 2 by 2 matrices, and almost all such matrices are full column rank) and $c \neq k$ to obtain the above inequality.

If $W_k \neq C_k$, the above construction may be trivially extended by letting all variables on the directed path from C_k to W_k be identity functions of their parents.

Assume there exists a recanting district D in $\mathcal{D}(\mathcal{G}_{W^*})$, where $W^* \equiv \mathrm{An}_{\mathcal{G}_{V\setminus A}}(W)$. Further, find the minimal subset W' of W such that the set of all childless vertices in D is in $\mathrm{An}_{\mathcal{G}_{V\setminus A}}(W')$. Let H be the smallest set of vertices that contains W', D, an element $A_i \in A \cap \mathrm{Pa}_{\mathcal{G}}(D)$ with a conflicting treatment assignment, and such that the set of childless vertices in D is in $\mathrm{An}_{\mathcal{G}_H}(W')$.

Consider any edge subgraph of $\mathcal{G}_H$ such that each vertex has at most one child. We construct elements $p_1(\mathbb{H})$ and $p_2(\mathbb{H})$ in the causal model in $\mathcal{G}_H$ as follows. In $p_1(\mathbb{H})$ each structural equation is a bit parity function of the parents, and each bidirected arc between V_i and V_j corresponds to a binary latent common parent ϵ_{ij} where each such latent is involved in precisely two functions. Moreover $p(\epsilon_{ij})$ is a uniform distribution. In $p_2(\mathbb{H})$ the same is true, except A_i is not involved in the structural equation for any element in D. It has been shown in [3] that $p_1(H) = p_2(H)$, but $p_1(W'(\pi, a_i, a'_i)) \neq p_2(W'(\pi, a_i, a'_i))$.

As before, consider a *backdoor path* π in $\mathcal{G}(a)$ from some element W_i in W' to an element Y_j in Y, such that W_i is m-connected to Y given W, and the edge on the path adjacent to W_i has an arrowhead into W_i. Such a path must exist by construction of W. In addition, consider the smallest subset W'' of W such that W_i is m-connected to Y_j given W'' in $\mathcal{G}_{V\setminus A}$. Pick the smallest set H' containing H such that the above m-connection statement holds in $\mathcal{G}_{H'}$. We now extend $p_1(\mathbb{H})$ and $p_2(\mathbb{H})$ to $p_1(\mathbb{H}')$ and $p_2(\mathbb{H}')$ to show $p(Y_j(a_{H'\cap A}) \mid W'(a_{H'\cap A}))$ is not identified.

We have three base cases. The first case assumes the first node Z_j on π not in H is a parent of an element Z_i. In this case, we let Z_i be the bit parity function of all its parents in $\mathcal{G}_{H'}$, including Z_j in both $p_1(\mathbb{H}')$ and $p_2(\mathbb{H}')$. By reasoning analogous to the hedge case, this implies $p_1(H') = p_2(H')$, but $p_1(Z_j(a_{H'\cap A}) = 0 \mid W'(a_{H'\cap A}) = 0) < 1$, while $p_2(Z_j(a_{H'\cap A}) = 0 \mid W'(a_{H'\cap A}) = 0) = 1$.

The second case assumes the first node Z_j on π not in H is a child of an element Z_i in H. The third case, which we also consider here, assumes $Y \in H$, in which case we let $Y = Z_i$. If $p(Z_j(\pi, a_i, a'_i) \mid W'(\pi, a_i, a'_i))$ (or $p(Y(\pi, a_i, a'_i) \mid W'(\pi, a_i, a'_i)))$ is not identified, we are done. Otherwise, we assume $p(Z_j(\pi, a_i, a'_i) \mid W'(\pi, a_i, a'_i))$ is identified. Consider the edge subgraph $\mathcal{G}'_{H'}$ of $\mathcal{G}_{H'}$ that lacks the outgoing directed edges from Z_i within H.

If Z_i is not in D, by reasoning analogous to reasoning in the hedge case, $\mathcal{G}_{H'}$ contains the recanting district structure for $p(Z_j(\pi, a, a'), W''(\pi, a, a'))$, where $W'' = W' \setminus \{W_i\}$ and W_i is $W' \cap \mathrm{De}_{\mathcal{G}_{H'}}(Z_j)$. If $p(W''(\pi, a, a'))$ is identified, we are done, since we established the base case where $p(Z_j(\pi, a, a') \mid W''(\pi, a, a'))$ is not identified. If $p(W''(\pi, a, a'))$ is not identified, note that W'' is a strictly smaller set then W', and we restart the base case argument, finding either a hedge or a recanting district for this smaller set, constructing a new set H, and a new backdoor path to an element in Y. Since the new subset of W is strictly smaller, we can only do this a finite number of times before encountering another base case.

If Z_i is in D, then the resulting graph $\mathcal{G}'_{H'}$ contains a recanting district structure for $p(Z_j(\pi, a, a'), W'(\pi, a, a'))$ with the set of childless vertices of the previous district and also Z_i (since it is now childless in H). Given the recanting district construction, $p_1(Z_j(\pi, a, a') = 0 \mid W'(\pi, a, a') = 0) < 1$, while $p_2(Z_j(\pi, a, a') = 0 \mid W'(\pi, a, a') = 0) = 1$, and we are done.

Since we now established bases for the induction for the recanting district case, we can apply the inductive argument for the hedge case to conclude $p(Y(\pi, a_i, a'_i) \mid W'(\pi, a_i, a'_i))$ is not identified, as above. Having established that $p(Y(\pi, a_i, a'_i) \mid W'(\pi, a_i, a'_i))$ is not identified in $\mathcal{G}_{H'}$ or $\mathcal{G}'_{H'}$, it is trivial to extend $p_1(\mathbb{H}')$ and $p_2(\mathbb{H}')$ to $p_1(\mathbb{V})$ and $p_2(\mathbb{V})$ for $\mathcal{G}(V)$.

Finally, our conclusion is established for $\mathcal{G}^e(V, A^{\mathrm{Ch}})$ and $p(Y(a^\pi) \mid W(a^\pi))$ by Proposition 9. □

A Weaker Causal Model

We phrased all our discussion in terms of the functional causal model, defined by the restriction (3). A weaker causal model called the *finest fully randomized causally interpretable structured tree graph* (FFRCISTG) suffices for many causal inference tasks. This model asserts that the variables,

$$\{V_i(\mathrm{pa}_i) \mid i \in \{1, \ldots, k\}\}, \tag{10}$$

are mutually independent for every $v \in \mathfrak{X}_V$, where pa_i is the subset of v associated with Pa_i. Note that the set of independences asserted by (10) is a subset of the set of independences asserted by (3). In particular, (10) only asserts independences among a set of potential outcomes associated with a globally consistent intervention operation, while (3) may allow independences among potential outcomes with inconsistent interventions. For example, a model defined by (3) may assert that $Y(a, m) \perp\!\!\!\perp M(a')$, while (10) never asserts such an independence if $a \neq a'$.

Since the SWIG global Markov property only asserts independences on random variables associated with a globally consistent intervention operation, it is implied not only by (3) but also the weaker model represented by (10) [2]. Potential outcomes like $p(Y(a, M(a')))$ that arise in mediation analysis are not identified under (10), but are sometimes identified under (3); see [7] for details. Note, however, that our rephrasing of edge-consistent counterfactuals $p(V_i(\pi, a, a'))$ in the causal model for $\mathcal{G}(V)$ in terms of an intervention $p(V_i(a^\pi))$ in the extended causal model for $\mathcal{G}^e(V \cup A^{\text{Ch}})$ leads to an identification theory for which model (10) for the variables in $\mathbb{V}^e$ is sufficient. The reason that counterfactuals $p(V_i(\pi, a, a'))$ requiring the stronger set of assumptions (3) may be rephrased as counterfactuals $p(V_i(a^\pi))$ only requiring the weaker set of assumptions (10) has to do with the specific way in which $\mathcal{G}^e$ was constructed. Specifically, $\mathcal{G}^e$ implicitly imposed strong restrictions on the associated FFRCISTG, having to do with deterministic relationships between A_i and A_i^j as well as absences of edges between any element A_i^j in A^{Ch} and any element in $\text{Ch}_i^{\mathcal{G}}$ other than V_j. Had these edges not been absent in $\mathcal{G}^e$, identification would no longer be possible. In some sense, $\mathcal{G}^e$ is the graph corresponding to the "weakest" FFRCISTG that encodes assumptions associated with the functional model on $\mathcal{G}$. These assumptions may be viewed informally as stating that a treatment variable A_i in A may be decomposed into components that only influence particular children (immediate effects) of A_i, and no other children of A_i.

References

[1] 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.

[2] 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.

[3] Ilya Shpitser. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cognitive Science (Rumelhart Special Issue)*, 37:1011–1035, 2013.

[4] 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.

[5] 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.

[6] 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.

[7] Ilya Shpitser and Eric J. Tchetgen Tchetgen. Causal inference with a graphical hierarchy of interventions. *Annals of Statistics*, 44(6):2433–2466, 2016.