
Structural Causal Bandits: Where to Intervene?

Sanghack Lee

Department of Computer Science
Purdue University
West Lafayette, IN 47907

Elias Bareinboim

Department of Computer Science
Purdue University
West Lafayette, IN 47907

Abstract

We study the problem of identifying the best action in a sequential decision-making setting when the reward distributions of the arms exhibit non-trivial dependencies, which are governed by the underlying causal structure of the domain where the agent is deployed. In this setting, playing an arm corresponds to intervening on a set of variables and setting them to specific values. In this paper, we start by showing that whenever the causal model relating the arms is unknown, the strategy of simultaneously intervening in all variables can, in general, lead to a sub-optimal policy (regardless of the number of iterations performed in the environment). We then derive structural properties implied by the given causal model, which is assumed to be known, albeit without parametrization. We further propose an algorithm that takes as input the causal structure and finds a minimal, sound, and complete set of qualified arms that the agent can play so as to maximize its reward. We empirically demonstrate that this algorithm leads to optimal, order of magnitude faster convergence rates when compared with its causal-insensitive counterparts.

1 Introduction

The multi-armed bandit (MAB) problem is one of the prototypical settings studied in the sequential decision-making literature [Lai and Robbins, 1985, Even-Dar et al., 2006, Bubeck et al., 2012]. At each time step, an agent needs to decide which arm to pull, and then it receives a corresponding reward, with the goal of maximizing its cumulative reward. The challenge is the inherent trade-off between exploiting known arms versus exploring new reward opportunities [Sutton and Barto, 1998].

There is a wide range of assumptions underlying MABs, but in most of the traditional settings, the arms' rewards are assumed to be independent, which means that knowing the reward distribution of one arm has no implication to the reward of the others. Many strategies were developed to solve this problem, including classic algorithms such as ϵ -greedy, UCB (Auer et al., 2002, Garivier and Cappé, 2011, Cappé et al., 2013), and Thompson sampling [Thompson, 1933, 1935]. Recently, the existence of non-trivial dependencies among arms has been acknowledged and explored under the rubric of *structured bandits*, which includes settings such as linear [Dani et al., 2008], unimodal [Combes and Proutiere, 2014], and Lipschitz [Magureanu et al., 2014], just to name a few. For example, a Lipschitz bandit imposes that the difference between the mean rewards (μ_i and μ_j) for two real-valued arms A_i and A_j is bounded, i.e., $|\mu_i - \mu_j| \leq L \cdot |A_i - A_j|$, where L is a positive constant. In this case, an *index-based* MAB algorithm, oblivious to the structural properties, can be suboptimal.

On another line of investigation, rich environments with complex dependency structures are modeled explicitly through the use of causal graphs, where nodes represent covariates, decision, and outcome variables, and direct edges represent direct influence of one variable to another (also called structural relations [Pearl, 2000, Ch. 7]). Despite the apparent connection between MABs and causality, only recently the use of causal reasoning has been incorporated into the design of MAB algorithms. For instance, Bareinboim et al. [2015] first explored the connection between causal models with unobserved confounders and reinforcement learning, where latent factors affect both the reward

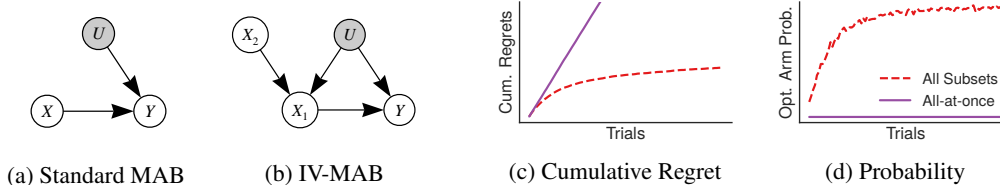


Figure 1: MAB problems as directed acyclic graphs where U is an unobserved variable. Plots of cumulative regrets and probability selecting an optimal arm when MAB algorithm intervenes X_1 and X_2 simultaneously (All-at-once) or all subsets of $\{X_1, X_2\}$ for IV-MAB.

distributions and the player’s intuition. This work was later extended to handle counterfactual distributions of higher dimensionality by Forney et al. [2017]. Lattimore et al. [2016] and Sen et al. [2017] studied the problem of best arm identification through importance weighting, where information on how playing arms influences the direct causes (parents, in causal terminology) of a reward variable is available. Zhang and Bareinboim [2017] leveraged causal graphs to address the transfer of samples when agents have different sensory and actuation capabilities.¹

In this paper, we focus on the challenge of identifying the best action in MABs where the arms correspond to interventions on an arbitrary causal graph, including when latent variables confound the observed relations (i.e., Semi-Markovian Causal models). To understand this challenge, we first note that a standard MAB can be seen as the simple causal model shown in Fig. 1a, where X represents an arm (with K different values), Y the reward variable, and U the unobserved variable that generates the randomness of Y .² After a sufficiently large number of pulls of X (chosen by the specific algorithm), Y ’s average reward can be determined with high confidence. Whenever unobserved confounders (UCs) affect more than one observed variable, however, novel, non-trivial challenges arise. To witness, consider the more involved MAB structure shown in Fig. 1b, where an unobserved confounder U affects both the action variable X_1 and the reward Y . A naive approach for an algorithm to play such a bandit would be to pull arms in a combinatorial manner, i.e., combining both arms ($X_1 \times X_2$) such that $D(X_1) \times D(X_2)$, where $D(X)$ is the domain of X . One may surmise that this is a valid strategy, albeit not the most efficient one. Somewhat unexpectedly, however, Fig. 1c shows that this is not the case — the optimal action comes from pulling X_2 and ignoring X_1 , while pulling $\{X_1, X_2\}$ together would lead to subpar cumulative rewards (regardless of the number of iterations) since it simply cannot pull the optimal arm (Fig. 1d).

After all, if one is oblivious to the causal structure and decides to take all intervenable variables as one (in this case, $X_1 \times X_2$), indiscriminately, she may get doomed to learn a suboptimal policy. In this paper, we investigate this phenomenon, and more broadly, causal MABs with non-trivial dependency structure between the arms. More specifically, our contributions are as follows:

1. We first formulate SCM-MABs, which is a structured multi-armed bandit instance within the causal framework. We then derive structural properties of a SCM-MAB, which can be computed given any causal model – i.e., arms’ equivalence based on *do*-calculus [Pearl, 1995], and partial orderedness among variables associated with arms in regards to the maximum rewards achievable.
2. We characterize a special set of variables called POMIS (possibly-optimal minimal intervention set), which are worth of intervening based on the aforementioned partial orders. We then introduce an algorithm that identifies a complete set of POMISs so that only a subset of arms associated with them can be explored in a MAB algorithm. Simulations corroborate with our findings.

Preliminaries: notations and structural causal model

We follow the notations used in the causal inference literature. A capital letter is used for a variable or a mathematical object. The domain of X is denoted by $D(X)$. A bold capital letter is for a set of variables, e.g., $\mathbf{X} = \{X_i\}_{i=1}^n$, while a lowercase letter $x \in D(X)$ is a value assigned to X , and

¹ On another line of research, Ortega and Braun [2014] investigated a generalized version of Thompson sampling applied to the problem of adaptive control.

²In causal notation, $Y \leftarrow f_Y(U, X)$, which means that Y ’s value is determined by the choice of X and the realization of the latent variable U . If f_Y is linear, we would have a linear bandit. Our results do not constraint the types of structural functions, which is usually within non-parametric causal inference [Pearl, 2000, Ch. 7].

$\mathbf{x} \in D(\mathbf{X}) = \times_{X \in \mathbf{X}} (D(X))$. We denote by $\mathbf{x}[\mathbf{W}]$, values of \mathbf{x} corresponding to $\mathbf{W} \cap \mathbf{X}$. A graph $G = \langle \mathbf{V}, \mathbf{E} \rangle$ is a pair of vertices \mathbf{V} and edges \mathbf{E} . We adopt family relationships — *pa*, *ch*, *an*, and *de* to represent parents, children, ancestors, and descendants of a given variable; *Pa*, *Ch*, *An*, and *De* extends *pa*, *ch*, *an*, and *de* by including the argument as the result, e.g., $Pa(X)_G = pa(X)_G \cup \{X\}$. With a set of variables as argument, $pa(\mathbf{X})_G = \bigcup_{X \in \mathbf{X}} pa(X)_G$ and others are similarly defined. We denote by $\mathbf{V}(G)$ the set of variables in G . $G[\mathbf{V}']$ for $\mathbf{V}' \subseteq \mathbf{V}(G)$ is a vertex-induced subgraph where all edges among \mathbf{V}' are preserved. We define $G \setminus \mathbf{X}$ as $G[\mathbf{V}(G) \setminus \mathbf{X}]$ for $\mathbf{X} \subseteq \mathbf{V}(G)$.

We adopt the language of Structural Causal Models (SCM) [Pearl, 2000]. An SCM M is a tuple $\langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$, where \mathbf{U} is a set of exogenous (unobserved or latent) variables and \mathbf{V} is a set of endogenous (observed) variables. \mathbf{F} is a set of functions $\mathbf{F} = \{f_i\}$, where f_i determines the value of $V_i \in \mathbf{V}$ based on parents $\mathbf{PA}_i \subseteq \mathbf{V} \setminus \{V_i\}$ and $\mathbf{U}^i \subseteq \mathbf{U}$. A causal diagram $G = \langle \mathbf{V}, \mathbf{E} \rangle$, associated with M , is a tuple of vertices \mathbf{V} (the endogenous variables) and edges \mathbf{E} , where a directed edge $V_i \rightarrow V_j \in \mathbf{E}$ if $V_i \in \mathbf{PA}_j$, and a bidirected edge between V_i and V_j if there exists an unobserved confounder, i.e., $\mathbf{U}^i \cap \mathbf{U}^j \neq \emptyset$. Probability of $Y = y$ when \mathbf{X} is held fixed at \mathbf{x} (i.e., intervened) is denoted by $P(y|do(\mathbf{x}))$, where intervention on \mathbf{X} is graphically represented by $G_{\overline{\mathbf{X}}}$, the graph G with incoming edges onto \mathbf{X} removed so that We denote by $CC(X)_G$ the c-component of G that contains X where a *c-component* is a maximal set of vertices connected with bidirected edges [Tian and Pearl, 2002]. We define $CC(\mathbf{X})_G = \bigcup_{X \in \mathbf{X}} CC(X)_G$. For a detailed discussion on the properties of SCMs, we refer readers to [Pearl, 2000, Ch. 7]. All proofs can be found in Appendix.

2 Multi-armed bandits with structural causal model

We recall that MABs consider a sequential setting where there are K arms and pulling an arm at each round gives a player a stochastic reward from an unknown distribution associated with the arm. The goal is to minimize (maximize) cumulative regret (reward) after T rounds. The mean reward of arm a is denoted by μ_a and the maximal reward is $\mu^* = \max_{1 \leq a \leq K} \mu_a$. We focus on a cumulative regret, $\text{Reg}_T = T\mu^* - \sum_{t=1}^T \mathbb{E}[Y_{A_t}] = \sum_{a=1}^K \Delta_a \mathbb{E}[T_a(T)]$, where A_t is the arm played at time t , $T_a(t)$ is the number of arm a has been played after t rounds, and $\Delta_a = \mu^* - \mu_a$.

We now can connect a MAB instance to its SCM counterpart. Let M be a SCM $\langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$ and $Y \in \mathbf{V}$ be a reward variable, where $D(Y) \subseteq \mathbb{R}$. The bandit contains arms $\{\mathbf{x} \in D(\mathbf{X}) \mid \mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}\}$, a set of all possible interventions on endogenous variables except the reward variable. Each arm $A_{\mathbf{x}}$ (or simply \mathbf{x}) associates with a reward distribution $P(Y|do(\mathbf{x}))$ where its mean reward $\mu_{\mathbf{x}}$ is $\mathbb{E}[Y|do(\mathbf{x})]$. We call this setting a SCM-MAB, which can be simply represented as a pair $\langle M, Y \rangle$. Throughout this paper, we assume that the causal graph G of M is fully accessible to the agent³, although its parametrization is unknown: that is, an agent facing a SCM-MAB $\langle M, Y \rangle$ plays arms with knowledge of G and Y . For simplicity, we denote information provided to an agent playing a SCM-MAB by $\llbracket G, Y \rrbracket$. We now investigate structural properties that follow from the causal structure G of the SCM-MAB.

Property 1. Equivalence among arms

We start by noting that *do*-calculus [Pearl, 1995] provides rules to evaluate invariances in the interventional space. In particular, we focus here on the third rule, which ascertains the condition such that a set of interventions can be deleted or inserted, namely, $P(y|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(y|do(\mathbf{x}), \mathbf{w})$. Since arms correspond to interventions (including the *null* intervention), we consider the test $P(y|do(\mathbf{x}), do(\mathbf{z})) = P(y|do(\mathbf{x}))$, which implies $\mu_{\mathbf{x}, \mathbf{z}} = \mu_{\mathbf{x}}$, if $Y \perp\!\!\!\perp \mathbf{Z} \mid \mathbf{X}$ in $G_{\overline{\mathbf{X} \cup \mathbf{Z}}}$. If valid, this condition ascertains that it is sufficient to play with only one arm among arms in the equivalent class.

Definition 1 (Minimal Intervention Set (MIS)). *Given information $\llbracket G, Y \rrbracket$, a set of variables $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}$ is said to be a minimal intervention set if there is no $\mathbf{X}' \subset \mathbf{X}$ such that $\mu_{\mathbf{x}[\mathbf{X}']} = \mu_{\mathbf{x}}$ for every SCM conforming to G .*

For instance, the MISs corresponding to all of the causal graphs in Fig. 2 are $\{\emptyset, \{X\}, \{Z\}\}$ without $\{X, Z\}$ since $\mu_x = \mu_{x,z}$. MISs are determined without regards to the unobserved confounders in a causal graph. The empty set and all singletons in $an(Y)_G$ are MISs for G with respect to Y . The task of finding a best arm among all possible arms can be reduced to the ones within MISs.

³In settings where this is not the case, one can spend the first iterations with the environment to learn the causal graph G from observational [Spirites et al., 2001] and experimental data [Kocaoglu et al., 2017].

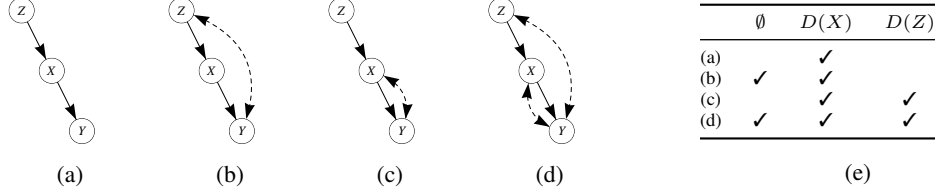


Figure 2: Causal graphs such that $\mu_x = \mu_{x,z}$, and non-dominated arms

Proposition 1 (Minimality). *A set of variables $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}$ is a minimal intervention set for G with respect to Y if and only if $\mathbf{X} \subseteq \text{an}(Y)_{G_{\overline{\mathbf{X}}}}$.*

All the MISs given $\llbracket G, Y \rrbracket$ can be determined without explicitly enumerating $2^{\mathbf{V} \setminus \{Y\}}$ while checking the condition in Prop. 1. We construct a recursive, efficient algorithm to enumerate the complete set of MISs given G and Y (Appendix A) with $O(mn^2)$, where m is the size of the set of MISs.

Property 2. Partial-orders among arms

We now explore the partial-orders among subsets of $\mathbf{V} \setminus \{Y\}$. One may wonder whether all the MIS arms are worth exploring when the goal is to maximize the cumulative reward. Given the causal diagram G , it is possible that intervening on some variables is *always* as good as intervening on another set of variables (regardless of the parametrization of the underlying model). Formally, there can be two different sets of variables $\mathbf{W}, \mathbf{Z} \subseteq \mathbf{V} \setminus \{Y\}$ such that

$$\max_{\mathbf{w} \in D(\mathbf{W})} \mu_{\mathbf{w}} \leq \max_{\mathbf{z} \in D(\mathbf{Z})} \mu_{\mathbf{z}}$$

in every possible SCM conforming to the given causal graph G . If that’s the case, it would be unnecessary (and possibly harmful in terms of data efficiency) to play arms $D(\mathbf{W})$. We next define Possibly-Optimal MIS, which incorporates the partial-orderedness among subsets of $\mathbf{V} \setminus \{Y\}$ into MIS denoting the optimal value for a given $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}$ by \mathbf{x}^* .

Definition 2 (Possibly-Optimal Minimal Intervention Set (POMIS)). *Given information $\llbracket G, Y \rrbracket$, let \mathbf{X} be an MIS. If there exists a SCM conforming to G such that $\mu_{\mathbf{x}^*} > \forall_{\mathbf{Z} \in \mathbb{Z} \setminus \{\mathbf{x}\}} \mu_{\mathbf{z}^*}$, where \mathbb{Z} is the set of MISs with respect to G and Y , then \mathbf{X} is a possibly-optimal minimal intervention set with respect to the information $\llbracket G, Y \rrbracket$.*

Intuitively, one may believe that the best action will be to intervene on the direct causes (parents) of the reward variable Y , since this would entail a higher degree of “controllability” of Y within the system. This, in fact, holds true if Y is not confounded with any of its ancestors, which includes the case where no unobserved confounders are present at all in the system (Markovian models).

Proposition 2. *Given information $\llbracket G, Y \rrbracket$, if Y is not confounded with $\text{an}(Y)_G$ via unobserved confounders, then $\text{pa}(Y)_G$ is an only POMIS.*

Corollary 3 (Markovian POMIS). *Given $\llbracket G, Y \rrbracket$, if G is Markovian, then $\text{pa}(Y)_G$ is an only POMIS.*

For instance, in Fig. 2a, $\{\{X\}\}$ is the set of POMISs. Whenever unobserved confounders (UCs) are present ⁴, on the other hand, the analysis becomes more involved. To witness, let us analyze the maximum achievable rewards of the MISs with the causal diagrams in Fig. 2. We start with Fig. 2b where $\mu_{z^*} \leq \mu_{x^*}$ can be derived: $\mu_{z^*} = \sum_x \mu_x P(x | \text{do}(z^*)) \leq \sum_x \mu_x^* P(x | \text{do}(z^*)) = \mu_{x^*}$. However, μ_\emptyset is not comparable to μ_{x^*} . For a concrete example, consider a SCM where the domains of variables are $\{0, 1\}$. Let U be the UC between Y and Z where $P(U = 1) = 0.5$. Let $f_Z(u) = 1 - u$, $f_X(z) = z$, and $f_Y(x, u) = x \oplus u$, where \oplus is the exclusive-or function. If X is not intervened, x will be $1 - u$ yielding $y = 1$ for both cases $u = 0$ or $u = 1$ so that $\mu_\emptyset = 1$. However, if X is intervened to either 0 or 1, y will be 1 only half the time since $P(U = 1) = 0.5$ resulting $\mu_{x^*} = 0.5$. We also provide in Appendix A a SCM such that $\mu_\emptyset < \mu_{x^*}$ holds true. This model ($\mu_\emptyset > \mu_{x^*}$) illustrates an interesting phenomenon — allowing an UC to affect Y freely may lead to higher rewards, which may be broken upon interventions. We now consider a different confounding structure shown in Fig. 2c (similar to Fig. 1b), where the variable Z lies outside of the influence of the UC

⁴Recall that unobserved confounders are represented in the graph as bidirected dashed edges.

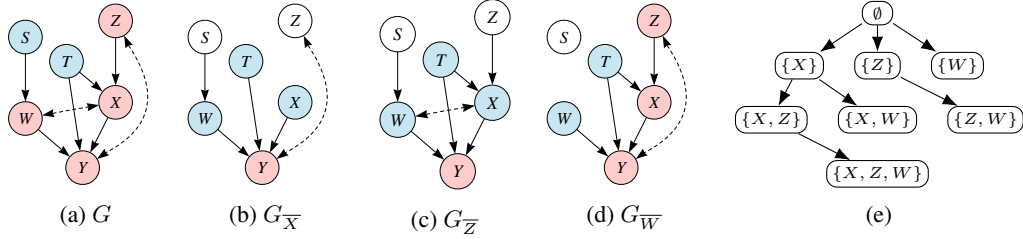


Figure 3: Causal graphs where pink and blue nodes are MUCT and IB, respectively. (Right most) A schematic showing an exploration order of subsets of variables

associated with Y . In this case, intervening on Z leads to a higher reward, $\mu_{z^*} \geq \mu_\emptyset$. To witness, note that $\mu_\emptyset = \sum_z \mathbb{E}[Y|z]P(z) = \sum_z \mu_z P(z) \leq \sum_z \mu_{z^*} P(z) = \mu_{z^*}$. However, μ_{z^*} and μ_{x^*} are incomparable, which is shown through two models provided in Appendix A. Finally, we can add the confounders of the two previous models, which is shown in Fig. 2d. In this case, all three μ_{x^*} , μ_{z^*} , and μ_\emptyset are incomparable. One can imagine scenarios where the influence of the UCs are weak enough so that corresponding models produce results similar to Figs. 2a to 2c.

It's clear that the interplay between the location of the intervened variable, the outcome variable, and the UCs entails non-trivial interactions and consequences in terms of the reward. The table in Fig. 2e highlights the arms that are contenders to generate the highest rewards in each model (i.e., each arm intervenes a POMIS to specific values). Note that X , the only parent of Y , is not dominated by any other arms in any scenario. In words, this means that the intuition that controlling variables closer to Y is not entirely lost even when UCs are present, they are not the only POMIS, but certainly one of them. Given that more complex mechanisms cannot be, in general, ruled out performing experiments would be required to identify the best arm. Still, the results of the table suggest that the search can be refined so that MAB solvers can discard arms that cannot lead to profitable outcomes, and converge faster to playing the optimal arm.

3 Graphical characterization of POMIS

Our goal in this section is to graphically characterize POMISs. We'll leverage the discussion in the previous section and note that UCs connected to a reward variable affect the reward distributions in a way that intervening on a variable outside the coverage of such UCs (including no UC) can be optimal — e.g., $\{X\}$ for Fig. 2a, \emptyset for Figs. 2b and 2d, and $\{Z\}$ for Fig. 2c. We introduce two graphical concepts to help capturing these subtleties.

Definition 3 (Unobserved-Confounders' Territory). *Given information $\llbracket G, Y \rrbracket$, let H be $G[An(Y)_G]$. A set of variables $\mathbf{T} \subseteq \mathbf{V}(H)$ containing Y is called an UC-territory on G with respect to Y if $De(\mathbf{T})_H = \mathbf{T}$ and $CC(\mathbf{T})_H = \mathbf{T}$.*

UC-territory \mathbf{T} is said to be *minimal* if no $\mathbf{T}' \subset \mathbf{T}$ is UC-territory. A minimal LC-Territory (MUCT) for G and Y can be constructed by extending a set of variables, starting from $\{Y\}$, alternatively updating the set with the c-component and descendants of the set.

Definition 4 (Interventional Border). *Let \mathbf{T} be a minimal UC-territory on G with respect to Y . Then, $\mathbf{X} = pa(\mathbf{T})_G \setminus \mathbf{T}$ is called an interventional border for G with respect to Y .*

The interventional border (IB) encompasses essentially the parents of the UC-territory. For concreteness, note that $\{W, X, Y, Z\}$ is the MUCT for the causal diagram in Fig. 3a with respect to Y and the IB associated with the MUCT is $\{S, T\}$ (marked in pink and blue, respectively). As its name suggests, MUCT is a set of endogenous variables governed by a set of UCs where at least one of them is related to a reward variable. The reward is determined by values of: (1) the set of UCs governing the MUCT; (2) a set of unobserved variables where each affects an endogenous variable in the MUCT; and (3) the IB. In other words, there is no UC interplaying across MUCT and its outside so that $\mu_{\mathbf{x}} = \mathbb{E}[Y|\mathbf{x}]$ where \mathbf{x} is a value assigned to $IB(G, Y)$. We now relate MUCT and IB to POMIS. Let $MUCT(G, Y)$ and $IB(G, Y)$ be, respectively, the MUCT and IB given $\llbracket G, Y \rrbracket$.

Proposition 4. *$IB(G, Y)$ is a POMIS given $\llbracket G, Y \rrbracket$.*

The main strategy of the proof is to construct a SCM M where intervening on any variable in $\text{MUCT}(G, Y)$ causes significant loss of reward. It seems that MUCT and IB can only identify a single POMIS given $\llbracket G, Y \rrbracket$. However, they, in fact, serve as basic units to identify all POMISs.

Proposition 5. *Given $\llbracket G, Y \rrbracket$, $\text{IB}(G_{\overline{\mathbf{W}}}, Y)$ is a POMIS, for any $\mathbf{W} \subseteq \mathbf{V} \setminus \{Y\}$.*

Prop. 5 generalizes Prop. 4, which is a special case when $\mathbf{W} = \emptyset$, while taking care of UCs across $\text{MUCT}(G_{\overline{\mathbf{W}}}, Y)$ and its outside in the original causal graph G . See Fig. 3d, for an instance, where $\text{IB}(G_{\overline{\mathbf{W}}}, Y) = \{W, T\}$. Intervening W cuts the influence of S and the UC between W and X , while still allowing the confounder affect X .⁵ Similarly, one can observe that $\text{IB}(G_{\overline{\mathbf{X}}}, Y) = \{T, W, X\}$ where intervening on X lets Y be an only element of MUCT making its parents an interventional border, hence, a POMIS. Note that $pa(Y)_G$ is always a POMIS since $\text{MUCT}(G_{\overline{pa(Y)_G}}, Y) = \{Y\}$ and $\text{IB}(G_{\overline{pa(Y)_G}}, Y) = pa(Y)_G$. With Prop. 5, one can enumerate the POMISs given $\llbracket G, Y \rrbracket$ considering all subsets of $\mathbf{V} \setminus \{Y\}$. We show in the sequel that this strategy encompasses all the POMISs.

Theorem 6. *Given $\llbracket G, Y \rrbracket$, $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}$ is a POMIS if and only if $\text{IB}(G_{\overline{\mathbf{X}}}, Y) = \mathbf{X}$.*

Thm. 6 provides a graphical necessary and sufficient condition for a set of variables being a POMIS given $\llbracket G, Y \rrbracket$. This characterization allows one to determine all possible arms in a SCM-MAB that are intervening on POMISs being free from pulling unnecessary arms.

4 Algorithmic characterization of POMIS

Although the graphical characterization provides a means to enumerate the complete set of POMISs given $\llbracket G, Y \rrbracket$, a naively implemented algorithm requires time exponential in $|\mathbf{V}|$. We construct an efficient algorithm (Algorithm 1) that enumerates all the POMISs based on propositions (7 and 8) below and the graphical characterization introduced in the previous section (Thm. 6).

Proposition 7 (Ignorability). *Let \mathbf{T} and \mathbf{X} be $\text{MUCT}(G_{\overline{\mathbf{W}}}, Y)$ and $\text{IB}(G_{\overline{\mathbf{W}}}, Y)$, respectively. Then, $\text{MUCT}(G_{\overline{\mathbf{X} \cup \mathbf{Z}}}, Y) = \mathbf{T}$ and $\text{IB}(G_{\overline{\mathbf{X} \cup \mathbf{Z}}}, Y) = \mathbf{X}$ for any $\mathbf{Z} \subseteq \mathbf{V} \setminus \mathbf{T}$.*

Proposition 8. *Let $H = G_{\overline{\mathbf{X}}}[\mathbf{T} \cup \mathbf{X}]$ where \mathbf{T} and \mathbf{X} are MUCT and IB given $\llbracket G_{\overline{\mathbf{W}}}, Y \rrbracket$, respectively. Then, for any $\mathbf{W}' \subseteq \mathbf{T} \setminus \{Y\}$, $H_{\overline{\mathbf{W}'}}$ and $G_{\overline{\mathbf{W}' \cup \mathbf{W}'}}$ yield the same MUCT and IB with respect to Y .*

Prop. 7 provides a means to avoid enumerating $G_{\overline{\mathbf{W}'}}$ for every $\mathbf{W}' \subseteq \mathbf{V} \setminus \{Y\}$. Prop. 8 characterizes recursive nature of MUCT and IB where identification of POMISs can be delegated to subgraphs. Based on these, we design a recursive algorithm (Algorithm 1), which explores subsets of $\mathbf{V} \setminus \{Y\}$ with a certain order. See Fig. 3e for an example where subsets of $\{X, Z, W\}$ are connected based on set inclusion relationship and an order of variables, e.g., (X, Z, W) . That is, there exists a directed edge between two sets if (i) one set is larger than the other by a variable and (ii) the variable's index (as in the order) is larger than other variable's index in the smaller set. The diagram traces how the algorithm will explore the subsets following the edges, while effectively skipping nodes.

Given an order of variables, which we choose the reverse topological order (Line 3), it examines POMISs by intervening on a single variable against the given graph (Line 6–9). If the IB (\mathbf{X} in Line 7) of such an intervened graph intersects with \mathbf{O}' (a set of variables that should be considered in other branch), then no subsequent call is made (Line 8). Otherwise, a subsequent subPOMISs call will take as arguments an MUCT-IB induced subgraph (Prop. 8), a refined order, and a set of variables not to be intervened in the given branch. For clarity, we provide a detailed working example in Appendix C with Fig. 3a where the algorithm explores only four intervened graphs (G , $G_{\overline{\{X\}}}$, $G_{\overline{\{Z\}}}$, $G_{\overline{\{W\}}}$) and generates the complete set of POMISs $\{\{S, T\}, \{T, W\}, \{T, W, X\}\}$.

Theorem 9 (Soundness and Completeness). *Given information $\llbracket G, Y \rrbracket$, the algorithm POMISs returns all and only POMISs.*

The POMISs algorithm can be combined with an MAB algorithm, such as the kl-UCB, creating a simple yet effective SCM-MAB solver — see Algorithm 2. We assume f is a non-decreasing function and T is the horizon. kl-UCB satisfies $\limsup_{T \rightarrow \infty} \frac{\mathbb{E}[\text{Reg}_T]}{\log(n)} \leq \sum_{\mathbf{x}: \mu_{\mathbf{x}} < \mu^*} \frac{\mu^* - \mu_{\mathbf{x}}}{KL(\mu_{\mathbf{x}}, \mu^*)}$ where KL is Kullback-Leibler divergence between two Bernoulli distributions. It is clear that the reduction in the size of non-optimal arms will lower the upper bounds of cumulative regrets.

⁵Note that unobserved variables who are not UCs are not explicitly represented in the graph.

Algorithm 1 Algorithm enumerating all POMISs with $\llbracket G, Y \rrbracket$

```
1: function POMISs( $G, Y$ )
2:    $\mathbf{T}, \mathbf{X} = \text{MUCT}(G, Y), \text{IB}(G, Y); H = G_{\overline{\mathbf{X}}}[\mathbf{T} \cup \mathbf{X}]$ 
3:   return  $\{\mathbf{X}\} \cup \text{subPOMISs}(H, Y, \text{reversed}(\text{topological-sort}(H)) \cap (\mathbf{T} \setminus \{Y\}), \emptyset)$ 

4: function subPOMISs( $G, Y, \pi, \mathbf{O}$ )
5:    $\mathbf{P} = \emptyset$ 
6:   for  $\pi_i \in \pi$  do
7:      $\mathbf{T}, \mathbf{X}, \pi', \mathbf{O}' = \text{MUCT}(G_{\overline{\pi_i}}, Y), \text{IB}(G_{\overline{\pi_i}}, Y), \pi^{i+1:|\pi|} \cap \mathbf{T}, \mathbf{O} \cup \pi^{1:i-1}$ 
8:     if  $\mathbf{X} \cap \mathbf{O}' = \emptyset$  then
9:        $\mathbf{P} = \mathbf{P} \cup \{\mathbf{X}\} \cup (\text{subPOMISs}(G_{\overline{\mathbf{X}}}[\mathbf{T} \cup \mathbf{X}], Y, \pi', \mathbf{O}'))$  if  $\pi' \neq \emptyset$  else  $\emptyset$ 
10:  return  $\mathbf{P}$ 
```

Algorithm 2 POMIS-kl-UCB

```
1: function POMIS-KL-UCB( $B, G, Y, f, T$ )
2:   Input:  $B$ , a SCM-MAB,  $G$ , a causal diagram;  $Y$ , a reward variable
3:    $\mathbf{A} = \bigcup_{\mathbf{X} \in \text{POMISs}(G, Y)} D(\mathbf{X})$ 
4:   kl-UCB( $B, \mathbf{A}, f, T$ )
```

5 Experiments

In this section, we present empirical results demonstrating the use of arms based on POMISs makes standard MAB solvers converge faster to an optimal arm. We employ two popular MAB solvers, kl-UCB [Auer et al., 2002], which enjoys cumulative regret growing logarithmically with the number of rounds [Auer et al., 2002, Cappé et al., 2013], and Thompson sampling (TS, [Thompson, 1933]), which usually has strong empirical performance. We considered four strategies for arm selection, including POMISs, MISs, Brute-force, and All-at-once, where Brute-force evaluates all combinations of arms $\bigcup_{\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}} D(\mathbf{X})$, and All-at-once considers intervening in all variables simultaneously, $D(\mathbf{V} \setminus \{Y\})$, oblivious to the causal structure and any knowledge about the action space.

The performance of eight (4×2) algorithms are evaluated relative to three different SCM-MAB instances (the detailed parametrizations are provided in Appendix D). We set the horizon large enough so as to observe near convergence, and repeat the simulations 300 times. We plot (i) the average cumulative regrets (CR) along with their respective standard deviation and (ii) the probability of an optimal arm being selected averaged over the repeated tests (OAP).

Task 1: We start by analyzing the Markovian model. We note that by Corollary 3, searching for the arms within the parent set is sufficient in this case. The number of arms for POMISs, MISs, Brute-force, and All-at-once are 4, 49, 81, and 16, respectively. Note that there are 4 optimal arms within All-at-once arms — for instance, if the parent configuration is $X_1 = x_1, X_2 = x_2$, this strategy will also include combinations of $Z_1 = z_1, Z_2 = z_2, \forall z_1, z_2$. The simulated results are shown in Fig. 4a. CR at round 1000 with kl-UCB are 3.0, 48.0, 72, and 12 (in the order), and all strategies were able to find the optimal arms at this time. POMIS and All-at-once first reached 95% OAP at round 20 and 66, respectively. Two additional remarks here. First, at an early stage, OAP for MISs is smaller than Brute-force since it has only 1 optimal arm among 49 arms, while Brute-force has 9 among 81. The advantage of employing MIS over Brute-force is only observed after a sufficiently large number of plays. Similarly, POMIS and All-at-once both have the common optimal to non-optimal arm-ratio (1:3 versus 4:12). However, POMIS dominates All-at-once since the agent can learn better about the mean reward of the optimal arm while playing non-optimal arms less, which translates into less variability and additional certainty about the optimal arm.

Task 2: We consider the popular setting known as instrumental variable (IV), which was shown in Fig. 2c. The true (unknown by the agent) optimal arm is setting $Z = 0$. The number of arms for the four strategies is 4, 5, 9, and 4, respectively. The results are shown in Fig. 4b. Since the All-at-once strategy only considers non-optimal arms (i.e., pulling Z, X together), it incurs in a linear regret without ever selecting an optimal arm (0%). CR (and OAP) at round 1000 with TS are POMIS 16.1 (98.67%), MIS 21.4 (99.00%), Brute-force 42.9 (93.33%), and All-at-once 272.1 (0%). At round 5000, where Brute-force nearly converged, the ratio of CRs for POMIS and Brute-force is $\frac{54.2}{18.1} = 2.99 \gtrsim 2.67 = \frac{9-1}{4-1}$. POMIS, MIS, and Brute-force first hits 95% OAP at 172, 214, and 435.

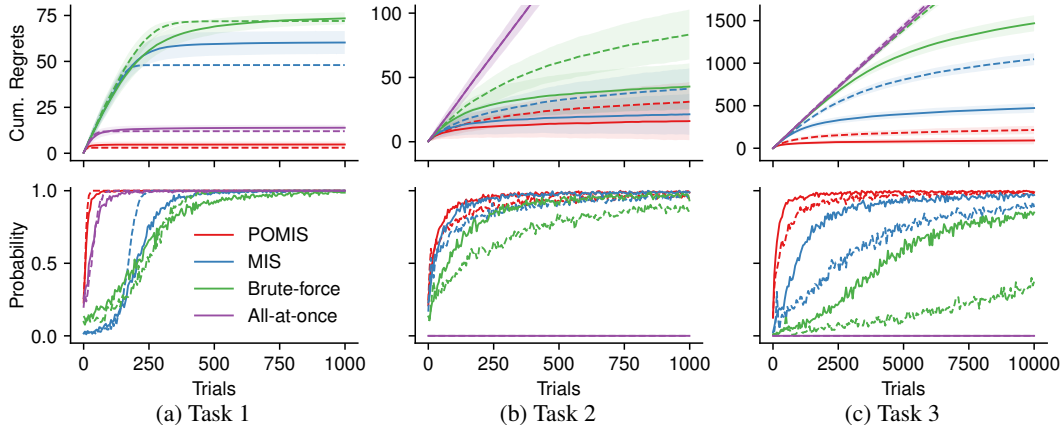


Figure 4: Comparisons across tasks (columns) with cumulative regrets (top) and optimal arm selection probability (bottom) with TS for solid and kl-UCB for dashed lines. Best viewed in color.

Task 3: We now study the non-trivial scenario discussed throughout the paper and shown in Fig. 3a. In this scenario, the optimal arm is intervening on $\{S, T\}$, which means that the system should follow its natural flow of UCs, which All-at-once is unable to “pull.” There are 16, 75, 243, and 32 arms for the strategies (in the order). The results are shown in Fig. 4c. The CR (and OAP) at round 10000 with TS are POMIS 91.4 (99.0%), MIS 472.4 (97.0%), Brute-force 1469.0 (85.0%), and All-at-once 2784.8 (0%). Similarly, the ratio (in round 10000) is $\frac{1469.0}{91.4} = 16.07 \approx 16.13 = \frac{243-1}{16-1}$ where we expect the ratio will keep increasing since Brute-force is not yet converged at the moment. Only POMIS and MIS achieved OAP of 95% first at 684 and 3544.

There are few notes following the experiments. First, the reduction in CRs is approximately proportional to the reduction in the number of non-optimal arms pulled by (PO)MIS. This makes the POMIS-based solver the clear winner throughout the simulations. Still, it’s not inconceivable that the number of arms examined by All-at-once is smaller than for POMIS in a specific SCM-MAB (associating $\llbracket G, Y \rrbracket$), which would entail a lower CR to the former. However, such a lower CR is not guaranteed since arms excluded from All-at-once but included in POMIS can be optimal in some SCM-MAB conforming to $\llbracket G, Y \rrbracket$; All-at-once is unable to find the optimal arm. Also, a POMIS-based strategy always dominates the corresponding MIS and brute force ones. These observations together suggest that, in practice, POMIS-based strategies should be preferred given that it will always converge and will usually be faster than their counterparts. Remarkably, there is an interesting trade-off between having knowledge of the causal structure versus not having the knowledge about the dependency among arms and potentially been incurring in linear regret (All-at-once) or exponential slow-down (Brute-force). In practice, for the cases in which the structure is unknown, the pull of the arms themselves can be used as experiments and could be coupled with efficient strategies to simultaneously learn the causal model [Kocaoglu et al., 2017].

6 Conclusions

We studied the problem of deciding whether to perform or not a causal intervention and, if so, which variables should be intervened upon. The problem was formalized as a special type of structural causal model (SCMs) that could be coupled with MAB machinery. We noted that there are subtleties in settings where latent variables are present, which could lead to non-convergence of standard MAB algorithms (i.e., without access to the underlying structure). Our strategy to tackle this problem was based on the do-calculus [Pearl, 2000] (which allowed the removal of redundant arms) and properties of the partial-orders among the sets of variables in the causal model (which allowed to establish the maximum reward achievable when they are intervened on). We developed an algorithm based on the possibly-optimal minimal intervention sets (called POMIS) that decides whether (and if so, where) interventions should be performed. We showed by simulation that this strategy performs more efficiently and more robustly than their non-causal counterparts.

Acknowledgments

This research is supported in parts by grants from IBM Research, Adobe Research, NSF IIS-1704352 and IIS1750807 (CAREER).

References

- Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2/3):235–256, 2002.
- Elias Bareinboim, Andrew Forney, and Judea Pearl. Bandits with unobserved confounders: A causal approach. In *Advances in Neural Information Processing Systems 28*, pages 1342–1350. 2015.
- Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.
- Olivier Cappé, Aurélien Garivier, Odalric-Ambrym Maillard, Rémi Munos, Gilles Stoltz, et al. Kullback–Leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, 41(3):1516–1541, 2013.
- Richard Combes and Alexandre Proutiere. Unimodal bandits: Regret lower bounds and optimal algorithms. In *International Conference on Machine Learning*, pages 521–529, 2014.
- Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *Proceedings of Conference On Learning Theory (COLT)*, pages 355–366, 2008.
- Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(Jun):1079–1105, 2006.
- Andrew Forney, Judea Pearl, and Elias Bareinboim. Counterfactual data-fusion for online reinforcement learners. In *International Conference on Machine Learning*, pages 1156–1164, 2017.
- Aurélien Garivier and Olivier Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. In *Proceedings of the 24th annual Conference On Learning Theory*, pages 359–376, 2011.
- Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Experimental design for learning causal graphs with latent variables. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, pages 7021–7031, 2017.
- Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.
- Finnian Lattimore, Tor Lattimore, and Mark D Reid. Causal bandits: Learning good interventions via causal inference. In *Advances in Neural Information Processing Systems 29*, pages 1181–1189. 2016.
- Stefan Magureanu, Richard Combes, and Alexandre Proutiere. Lipschitz bandits: Regret lower bound and optimal algorithms. In Maria Florina Balcan, Vitaly Feldman, and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pages 975–999, Barcelona, Spain, 13–15 Jun 2014. PMLR.
- Pedro A. Ortega and Daniel A. Braun. Generalized thompson sampling for sequential decision-making and causal inference. *Complex Adaptive Systems Modeling*, 2(2), 2014.
- Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.
- Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000. Second ed., 2009.
- Rajat Sen, Karthikeyan Shanmugam, Alexandros G Dimakis, and Sanjay Shakkottai. Identifying best interventions through online importance sampling. In *International Conference on Machine Learning*, pages 3057–3066, 2017.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. A Bradford Book, 2001.

Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: An introduction*. MIT press, 1998.

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

William R. Thompson. On the theory of apportionment. *American Journal of Mathematics*, 57(2): 450–456, 1935.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 567–573, 2002.

Junzhe Zhang and Elias Bareinboim. Transfer learning in multi-armed bandits: A causal approach. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 1340–1346, 2017.

A Supplementary Material for “Structural Causal Bandits: Where to Intervene?”

Appendix A Multi-armed bandits with structural causal model

Proposition 1 (Minimality). *A set of variables $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}$ is a minimal intervention set for G with respect to Y if and only if $\mathbf{X} \subseteq an(Y)_{G_{\overline{\mathbf{X}}}}$.*

Proof. (If) Assume that there exists $\mathbf{X}' \subset \mathbf{X}$ such that $\mu_{\mathbf{x}'} = \mu_{\mathbf{x}}$ with $\mathbf{x}' = \mathbf{x}[\mathbf{X}']$ so that \mathbf{X} is not minimal. Create a SCM with all variables real-valued, where each variable $V_i \in \mathbf{V}$ associates with its own binary exogenous variable $P(u_i = 1) = 0.5$. Let the function of an endogenous variable be the sum of values of its parents. For the sake of contradiction assume that $\mathbf{X} \subseteq an(Y)_{G_{\overline{\mathbf{X}}}}$. Then, there exists directed paths from $\mathbf{X} \setminus \mathbf{X}'$ to Y without passing \mathbf{X}' . Hence, setting $\mathbf{W} = \mathbf{X} \setminus \mathbf{X}'$ to $\mathbb{E}[\mathbf{w} | do(\mathbf{x}')] + 1$ will yield a larger outcome, i.e., $\mu_{\mathbf{w}, \mathbf{x}'} > \mu_{\mathbf{x}'}$, breaking the equality, which contradicts the assumption.

(Only if) Let $\mathbf{X} \not\subseteq an(Y)_{G_{\overline{\mathbf{X}}}}$. Let $\mathbf{Z} = \mathbf{X} \setminus an(Y)_{G_{\overline{\mathbf{X}}}}$, which is a non-empty set and $\mathbf{X}' = \mathbf{X} \setminus \mathbf{Z}$. By Rule 3 of *do*-calculus, $\mu_{\mathbf{x}', \mathbf{z}} = \mu_{\mathbf{x}'}$, which violates the definition of MIS. \square

We present an algorithm (Algorithm 3) for enumerating all the MISs given a causal diagram G and a reward variable Y . The algorithm builds a set of MISs by adding a variable to a previously obtained MIS so that the resulting set is a MIS.

Algorithm 3 Minimal Intervention Set Enumeration

```

1: function MISs( $G, Y$ )
2:   Input:  $G$  a causal diagram;  $Y$  an outcome variable
3:    $H = G[An(Y)_G]$ 
4:   return subMISs( $H, Y, \emptyset$ , reversed(topological-sort( $H$ ))  $\cap \{\mathbf{V} \setminus \{Y\}\}$ )

5: function subMISs( $G, Y, \mathbf{X}, \mathbf{W}$ )
6:    $\mathbb{X} = \{\mathbf{X}\}$ 
7:   for  $W_i \in \mathbf{W}$  do
8:      $H = G_{\overline{W_i}}[An(Y)_{G_{\overline{W_i}}}]$ 
9:      $\mathbb{X} = \mathbb{X} \cup \text{subMISs}(H, Y, \mathbf{X} \cup \{W_i\}, \mathbf{W}^{i+1} \cap \mathbf{V}(H))$ 
10:  return  $\mathbb{X}$ 

```

We prove Prop. 2 below with the following observation — given two different MISs \mathbf{X} and \mathbf{Z} , if $\mu_{\mathbf{x}^*} \leq \mu_{\mathbf{z}^*}$ for every SCM conforming to a given causal diagram G , then there exists a SCM $\mu_{\mathbf{x}^*} < \mu_{\mathbf{z}^*}$.

Proposition 2. *Given information $\llbracket G, Y \rrbracket$, if Y is not confounded with $an(Y)_G$ via unobserved confounders, then $pa(Y)_G$ is an only POMIS.*

Proof. Let \mathbf{X} be a MIS. Let $\mathbf{X}' = \mathbf{X} \setminus pa(Y)_G$ and $\mathbf{Z} = pa(Y)_G \setminus \mathbf{X}$.

$$\begin{aligned}
\mu_{\mathbf{x}} &= \sum_{\mathbf{z}} \mathbb{E}[Y \mid do(\mathbf{x}), \mathbf{z}] P(\mathbf{z} \mid do(\mathbf{x})) \\
&= \sum_{\mathbf{z}} \mathbb{E}[Y \mid do(\mathbf{x}, \mathbf{z})] P(\mathbf{z} \mid do(\mathbf{x})) && \because \text{Rule 2} \\
&= \sum_{\mathbf{z}} \mathbb{E}[Y \mid do(\mathbf{x}[pa(Y)_G], \mathbf{z})] P(\mathbf{z} \mid do(\mathbf{x})) \\
&\leq \sum_{\mathbf{z}} \mu_{pa_Y^*} P(\mathbf{z} \mid do(\mathbf{x})) \\
&= \mu_{pa_Y^*}
\end{aligned}$$

□

We describe two models for Fig. 2b showing $\mu_{\emptyset} > \mu_{x^*}$ and $\mu_{\emptyset} < \mu_{x^*}$, respectively. Let \oplus represent the exclusive-or function.

- $\mu_{\emptyset} > \mu_{x^*}$: Let the domains of U , X , and Z be $\{0, 1\}$ and let $0 < P(U = 1) = \alpha < 1$. \mathbf{F} consists of $f_Z(u) = 1 - u$, $f_X(z) = z$, and $f_Y(z, u) = z \oplus u$. Then, $\mu_{x^*} = \mu_{z^*} = \max(\alpha, 1 - \alpha)$, which is smaller than $\mu_{\emptyset} = 1$:

$$\begin{aligned}
\mu_{z^*} &= \sum_{u, x, y} y \cdot P(y \mid x, u) P(x \mid z^*) P(u) \\
&= \sum_{u, x} P(y = 1 \mid x, u) P(x \mid z^*) P(u) \\
&= \sum_u P(y = 1 \mid X = z^*, u) P(u) \\
&= \alpha P(y = 1 \mid X = z^*, 1) + (1 - \alpha) P(y = 1 \mid X = z^*, 0) \\
&= \alpha \delta_{z^*, 0} + (1 - \alpha) \delta_{z^*, 1} \\
&= \max(\alpha, 1 - \alpha)
\end{aligned}$$

Since $\mu_{\emptyset} = 1$, observation is strictly better than intervening on either Z or X .

- $\mu_{\emptyset} < \mu_{x^*}$: Changing $f_Y(x, u)$ to $x + u$, we observe $\mu_{x^*} = \mu_{z^*} = 1 + \alpha > \mu_{\emptyset} = 1$.

Deterministic relations can be modified to a probabilistic one by introducing binary unobserved variables U_X , U_Y , and U_Z and modifying functions for X , Y , and Z to exclusive-or with U_X , U_Y , and U_Z , respectively. By setting probability of them being 1 small enough, one can keep the orders $\mu_{\emptyset} > \mu_{x^*}$ or $\mu_{\emptyset} < \mu_{x^*}$.

We devise two models where $\mu_{z^*} > \mu_{x^*}$ and $\mu_{z^*} < \mu_{x^*}$, respectively, for Fig. 2c. Let U_X , U_Y , and U_Z be variable-specific exogenous variables affecting X , Y , and Z , respectively. Let U be the unobserved confounder between X and Y .

- $\mu_{z^*} > \mu_{x^*}$: Let $P(U = 1) = 0.5$ with $\forall U_V \in \{U_X, U_Y, U_Z\} P(U_V = 1) = \epsilon \approx 0$. Let $f_Y(x, u, u_Y) = x \oplus (1 - u) \oplus u_Y$, $f_X(z, u, u_X) = z \oplus u_X \oplus u$, and $f_Z(u_Z) = u_Z$. Then, $\mu_x = (1 - \epsilon) \cdot P(U_X = 1 - x) + \epsilon \cdot P(U_X = x) \approx P(U_X = 1 - x) = 0.5$ while $\mu_{z^*} = 1 - 2\epsilon + \epsilon^2 \approx 1$.
- $\mu_{z^*} < \mu_{x^*}$: Let probabilities of exogenous variables being 1 be 0.5. Let $f_Y(x, u, u_Y) = x + u + u_Y$, $f_X(z, u, u_X) = z \oplus u_X \oplus u$, and $f_Z(u_Z) = u_Z$. Then, $\mu_x = x + 0.5 + 0.5 = x + 1$ and $\mu_z = P(x = 1 \mid do(z)) + 0.5 + 0.5$. Therefore, $\mu_{z^*} = 1.5 < 2 = \mu_{x^*}$.

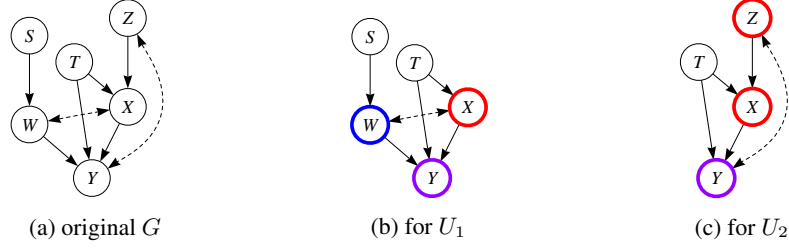


Figure 5: A causal diagram and colored graphs for unobserved confounders.

Appendix B Graphical characterization of POMIS

We present an algorithm (Algorithm 4) retrieving a MUCT given $\llbracket G, Y \rrbracket$.

Algorithm 4 Minimal Latent Confounders' Territory

```

1: function MUCT( $G, Y$ )
2:    $H = G \upharpoonright_{An(Y)}$ 
3:    $\mathbf{Q} = \{Y\}; \mathbf{T} = \{Y\}$ 
4:   while  $\mathbf{Q} \neq \emptyset$  do
5:     remove an element  $Q_1$  from  $\mathbf{Q}$ 
6:      $\mathbf{W} = \text{CC}(Q_1)_H; \mathbf{T} = \mathbf{T} \cup \mathbf{W}; \mathbf{Q} = (\mathbf{Q} \cup \text{de}(\mathbf{W})_H) \setminus \mathbf{T}$ 
7:   return  $\mathbf{T}$ 

```

The following proposition and corollary will be used partly to prove propositions and theorems in the main text.

Proposition 10 (Subsumption). *Let \mathbf{T} and \mathbf{X} be the MUCT and IB for G with respect to Y , respectively. Then, for any $\mathbf{Z} \subseteq \mathbf{V} \setminus \{Y\}$, $\mu_{\mathbf{S}} \geq \mu_{\mathbf{Z}^*}$ where $\mathbf{S} = (\mathbf{T} \cap \mathbf{Z}) \cup \mathbf{X}$.*

Proof. (Case $\mathbf{Z} \supseteq \mathbf{X}$) By definition of IB and Rule 3 of *do*-calculus, $\mu_{\mathbf{Z}^*} = \mathbb{E}[Y | \text{do}(\mathbf{z}^* [\mathbf{T} \cup \mathbf{X}])]$. Since $\mathbf{Z} \cap (\mathbf{T} \cup \mathbf{X}) = \mathbf{S}$, $\mu_{\mathbf{Z}^*} = \mu_{\mathbf{S}^*}$.

(Otherwise) Let $\mathbf{X}' = \mathbf{X} \setminus \mathbf{Z}$. Then,

$$\mu_{\mathbf{Z}^*} = \sum_{\mathbf{x}'} \mathbb{E}[Y | \text{do}(\mathbf{z}^*), \mathbf{x}'] P(\mathbf{x}' | \text{do}(\mathbf{z}^*))$$

The first term becomes

$$\begin{aligned} \mathbb{E}[Y | \text{do}(\mathbf{z}^*), \mathbf{x}'] &= \mathbb{E}[Y | \text{do}(\mathbf{z}^*), \text{do}(\mathbf{x}')] && \because \text{Rule 2 } (Y \perp\!\!\!\perp \mathbf{X}' | \mathbf{Z})_{G_{\mathbf{Z}\mathbf{X}'}} \\ &= \mathbb{E}[Y | \text{do}(\mathbf{z}^* [\mathbf{T} \cup \mathbf{X}]), \text{do}(\mathbf{x}')] && \because \text{Rule 3} \\ &\leq \mu_{\mathbf{S}} && \because (\mathbf{Z} \cap (\mathbf{T} \cup \mathbf{X})) \cup \mathbf{X}' = \mathbf{S} \end{aligned}$$

Finally, $\mu_{\mathbf{Z}^*} \leq \mu_{\mathbf{S}^*}$ because $\sum_{\mathbf{x}'} \mu_{\mathbf{S}^*} P(\mathbf{x}' | \text{do}(\mathbf{z}^*)) = \mu_{\mathbf{S}^*}$. \square

Corollary 11. *Given $\llbracket G, Y \rrbracket$, no POMIS intersects with an $(\mathbf{X})_G \setminus \mathbf{X}$ where $\mathbf{X} = \text{IB}(G, Y)$.*

The proposition says that rewards of arms related to intervening on \mathbf{Z} cannot be better than intervening on \mathbf{Z} and the border \mathbf{X} together. Further, since intervening outside of the territory and the border is ineffective, one can intervene only \mathbf{Z} that are inside the territory altogether with the border.

Proposition 4. *$\text{IB}(G, Y)$ is a POMIS given $\llbracket G, Y \rrbracket$.*

Proof. Let \mathbf{X} be $\text{IB}(G, Y)$. In this proof, every unobserved variables \mathbf{U} is a binary variable with its domain being $\{0, 1\}$. An easy case is when $\mathbf{T} = \{Y\}$ where \mathbf{X} is the parents of Y in G . We construct a SCM such that

1. Each endogenous variable V associates with an unobserved variable U_V ;

2. $f_Y = 1 - (\bigvee \mathbf{u}^Y \oplus (\bigvee \mathbf{x}))$ with $P(\mathbf{u}^Y = 0) \approx 1$;
3. $f_V = (\bigoplus \mathbf{u}^V) \oplus (\bigoplus \mathbf{pa}_V)$ for $V \in \mathbf{V} \setminus \{Y\}$ with $P(u_j) = 0.5$ for every $U_j \in \mathbf{U} \setminus \mathbf{U}^Y$.

Then, $\mathbb{E}[Y|do(\mathbf{X} = 0)] = P(\mathbf{u}^Y = 0) \approx 1$ while all others yield expectations less than or equal to 0.5, otherwise.

Now, we consider a general case where $\mathbf{T} \supset \{Y\}$, that is, there exists at least one unobserved confounder between Y and its ancestors. As a first step, we prove the existence of a SCM M , conforming to $H = G[\mathbf{T} \cup \mathbf{X}]$, which yields the maximum outcome *only* through $do(\mathbf{X} = 0)$. To do so, we construct a SCM for each unobserved confounder in $H[\mathbf{T}]$. Let $\mathbf{U}' = \{U_j\}_{j=1}^m$ be unobserved confounders in $H[\mathbf{T}]$. Then, those m individual SCMs $\{M_i\}_{i=1}^m$ will be integrated into a single SCM M so that any intervention other than $\mathbf{x} = 0$ negatively affects the outcome of Y .

We proceed to describe M_i for $U_i \in \mathbf{U}'$. Let $B^{(i)}$ and $R^{(i)}$ be two children of U_i . Let

$$H_i = H \left[De \left(\left\{ B^{(i)}, R^{(i)} \right\} \right)_H \cup \left(\mathbf{X} \cap pa \left(De \left(\left\{ B^{(i)}, R^{(i)} \right\} \right)_H \right) \right) \right]$$

with all bidirected edges removed except U_i . Functions for variables in $H[\mathbf{T}]$ will be described below. We label (i.e., color code) vertices in $De(B^{(i)})_H \setminus De(R^{(i)})_H$ as **blue** and $De(R^{(i)})_H \setminus De(B^{(i)})_H$ as **red**, and $De(B^{(i)})_H \cap De(R^{(i)})_H$ as **purple**. Each of $B^{(i)}$ and $R^{(i)}$ perceives that U_i is a parent colored as **blue** with value u_i and **red** with value $1 - u_i$, respectively. Those **blue**, **red**, **purple** variables are assigned 3 if any of their parents in \mathbf{X} is not 0. Otherwise, their values are determined as follows. For every **blue** and **red** vertex, its corresponding function returns the common value of its parents of the same color and returns 3 if colored parents' values are not homogeneous. For every **purple** vertex, its corresponding function returns 2 if every **blue**, **red**, and **purple** parent is 0, 1, and 2, respectively, and returns 1 if 1, 0, and 1, respectively. For other cases, the function returns 3. One can view the value 3 as a parity propagated to Y .

Now, we integrate m SCMs into one. In M_i , two bits are sufficient to represent every variable. Then, we build a unified SCM where each variable in \mathbf{T} is represented with $2m$ bits where a SCM for U_i will take $2i - 1^{\text{th}}$ and $2i^{\text{th}}$ bit (n.b. the right most digit representing $2^0 = 1$ corresponds to the first bit). We then binarize Y by setting 1 if $2i - 1^{\text{th}}$ and $2i^{\text{th}}$ bits are 01 or 10 for every $1 \leq i \leq m$ and 0 otherwise. Let $P(u_i = 1) = 0.5$ for $U_i \in \mathbf{U}'$. This unified SCM M provides a core mechanism to output $Y = 1$ if $do(\mathbf{X} = 0)$ and $Y = 0$ if $do(\mathbf{X} \neq 0)$. If any of variable in \mathbf{T} is intervened, then at least one sub-SCM among m sub-SCMs will be disrupted yielding an expectation smaller than or equal to 0.5.

We now extend the core SCM for $H[\mathbf{T} \cup \mathbf{X}]$ to a SCM for G . However, we can ignore joint probability distributions for any exogenous variables only affecting endogenous variables the outside of H . In addition, functions for endogenous variables lying outside H is irrelevant to the reward distribution. We define a function for $V \in An(Y)_G \setminus \mathbf{T}$ and joint distributions for unobserved variables except \mathbf{U}' . Let $f_V = \bigoplus \mathbf{u}^V \oplus \bigoplus \mathbf{pa}_V$ and $P(u_i = 0) = 0.5$ for $U_i \in \mathbf{U}$ whose child(ren) disjoint to \mathbf{T} and $P(u_j = 0) \approx 1$ for $U_j \in \mathbf{U}$ whose child(ren) intersects with \mathbf{T} . This ensures that the core mechanism will only be randomly disturbed with a small probability close to 0 preserving the best arm being $do(\mathbf{X} = 0)$. \square

We provide an example in Fig. 5 illustrating how sub-SCMs are constructed. Further, values of variables for M_1 , M_2 and a unified M are shown in Table 1 with $s = t = 0$.

Proposition 5. *Given $\llbracket G, Y \rrbracket$, $\text{IB}(G_{\overline{\mathbf{W}}}, Y)$ is a POMIS, for any $\mathbf{W} \subseteq \mathbf{V} \setminus \{Y\}$.*

Proof. Let $\mathbf{T} = \text{MUCT}(G_{\overline{\mathbf{W}}}, Y)$, $\mathbf{X} = \text{IB}(G_{\overline{\mathbf{W}}}, Y)$, and $\mathbf{T}_0 = \text{MUCT}(G, Y)$. We adopt the strategy used in Prop. 4 where a SCM is constructed so that an optimal arm is $do(\text{IB}(G, Y) = 0)$. We first similarly build a SCM for $G[\mathbf{T} \cup \mathbf{X}]$ while ignoring, for now, *dangling* unobserved confounders between \mathbf{T} and $\mathbf{T}_0 \setminus \mathbf{T}$. Let \mathbf{U}' be such unobserved confounders. Then, we modify the SCM so that a dangling unobserved confounder $U_i \in \mathbf{U}'$ flips (i.e., $0 \leftrightarrow 1$) the value of its endogenous child in \mathbf{T} when $u_i = 1$. Let $P(\mathbf{u}' \neq 0)$ be close to 0 so that $\mathbb{E}[Y|do(\mathbf{X} = 0)]$ is close to 1. However, intervening on $\mathbf{X} \neq 0$ or on \mathbf{Z} where $\mathbf{Z} \cap \mathbf{T} \neq \emptyset$ will make corresponding expectations around 0.5 or below. \square

Theorem 6. *Given $\llbracket G, Y \rrbracket$, $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}$ is a POMIS if and only if $\text{IB}(G_{\overline{\mathbf{X}}}, Y) = \mathbf{X}$.*

U_1	U_2	M_1			M_2			M				
		$w^{(1)}$	$x^{(1)}$	$y^{(1)}$	$z^{(2)}$	$x^{(1)}$	$y^{(1)}$	w	z	x	y'	y
0	0	1	0	2	0	0	2	00 01	00 00	00 00	10 10	1
	1				1	1	1		01 00	01 00	01 10	
1	0	0	1	1	0	0	2	00 00	00 00	00 01	10 01	
	1				1	1	1		01 00	01 01	01 01	

Table 1: Values with $s = t = 0$ where values for M are shown as binary. y' represents $4y^{(2)} + y^{(1)}$ before the value is binarized.

Proof. (If part) A special case of Prop. 5 where $\mathbf{W} = \mathbf{X}$.

(Only if part) Let $\mathbf{W} \subseteq \mathbf{V} \setminus \{Y\}$. Let \mathbf{T} and \mathbf{X} be $\text{MUCT}(G_{\overline{\mathbf{W}}}, Y)$ and $\text{IB}(G_{\overline{\mathbf{W}}}, Y)$ and \mathbf{T}_0 and \mathbf{X}_0 be $\text{MUCT}(G, Y)$ and $\text{IB}(G, Y)$, respectively. We will prove that \mathbf{W} is not a POMIS when $\mathbf{W} \neq \mathbf{X}$. We can limit $\mathbf{W} \subseteq \mathbf{X}_0 \cup \mathbf{T}_0 \setminus \{Y\}$ due to Prop. 10. Let $\mathbf{X}' = \mathbf{X} \setminus \mathbf{W}$. First, we can observe that $\mathbf{W} \subseteq \text{An}(\mathbf{X})_G$ since otherwise $\mathbf{W} \cap \mathbf{T} \neq \emptyset$, which contradicts that \mathbf{W} is neither a descendant of some variable nor confounded in $G_{\overline{\mathbf{W}}}$. Then,

$$\begin{aligned}
\mu_{\mathbf{w}^*} &= \sum_{\mathbf{x}'} \mathbb{E}[Y | do(\mathbf{w}^*), \mathbf{x}'] P(\mathbf{x}' | do(\mathbf{w}^*)) && \because \text{Basic algebra} \\
&= \sum_{\mathbf{x}'} \mathbb{E}[Y | do(\mathbf{w}^*), do(\mathbf{x}')] P(\mathbf{x}' | do(\mathbf{w}^*)) && \because \text{Rule 2 } (Y \perp\!\!\!\perp \mathbf{X}' | \mathbf{W})_{G_{\overline{\mathbf{W}\mathbf{X}'}}} \\
&= \sum_{\mathbf{x}'} \mathbb{E}[Y | do(\mathbf{w}^*[\mathbf{X}]), do(\mathbf{x}')] P(\mathbf{x}' | do(\mathbf{w}^*)) && \because \text{Rule 3 } (Y \perp\!\!\!\perp (\mathbf{W} \setminus \mathbf{X}) | \mathbf{X})_{G_{\overline{\mathbf{X}, \mathbf{W} \setminus \mathbf{X}}}} \\
&\leq \sum_{\mathbf{x}'} \mu_{\mathbf{x}^*} P(\mathbf{x}' | do(\mathbf{w}^*)) \\
&= \mu_{\mathbf{x}^*}
\end{aligned}$$

Therefore, \mathbf{W} is not a POMIS. \square

Appendix C Algorithmic characterization of POMIS

Proposition 7 (Ignorability). *Let \mathbf{T} and \mathbf{X} be $\text{MUCT}(G_{\overline{\mathbf{W}}}, Y)$ and $\text{IB}(G_{\overline{\mathbf{W}}}, Y)$, respectively. Then, $\text{MUCT}(G_{\overline{\mathbf{X}\cup\mathbf{Z}}}, Y) = \mathbf{T}$ and $\text{IB}(G_{\overline{\mathbf{X}\cup\mathbf{Z}}}, Y) = \mathbf{X}$ for any $\mathbf{Z} \subseteq \mathbf{V} \setminus \mathbf{T}$.*

Proof. Given that \mathbf{T} is a MUCT for G being intervened on \mathbf{X} , additional intervention outside \mathbf{T} does not affect \mathbf{T} being the MUCT and \mathbf{X} being the IB. \square

Proposition 8. *Let $H = G_{\overline{\mathbf{X}}}[\mathbf{T} \cup \mathbf{X}]$ where \mathbf{T} and \mathbf{X} are MUCT and IB given $\llbracket G_{\overline{\mathbf{W}}}, Y \rrbracket$, respectively. Then, for any $\mathbf{W}' \subseteq \mathbf{T} \setminus \{Y\}$, $H_{\overline{\mathbf{W}'}}$ and $G_{\overline{\mathbf{W}' \cup \mathbf{W}'}}$ yield the same MUCT and IB with respect to Y .*

Proof. $G_{\overline{\mathbf{W}' \cup \mathbf{W}'}}$ and $H_{\overline{\mathbf{W}'}}$ share the same edges among $\mathbf{T} \cup \mathbf{X}$ except the fact that \mathbf{X} has no parent in $H_{\overline{\mathbf{W}'}}$, which is irrelevant to identifying MUCT \mathbf{T} and IB \mathbf{X} in both diagrams. \square

Proposition 12. *If \mathbf{W} is a POMIS, for any $\mathbf{W}' \subset \mathbf{W}$, $\mathbf{W} \setminus \mathbf{W}' \subset \text{MUCT}(G_{\overline{\mathbf{W}'}}), Y) \cup \text{IB}(G_{\overline{\mathbf{W}'}}), Y)$.*

Proof. Otherwise, \mathbf{W} is not a MIS since intervening on $\text{IB}(G_{\overline{\mathbf{W}'}}), Y)$ is preferred to intervening on \mathbf{W} . \square

We illustrate how the algorithm works with a causal graph G in Fig. 3a. The graph and its manipulated graphs are shown in Fig. 3. Given G and Y , POMISS obtains a MUCT-and-IB induced subgraph with the IB intervened, $G_{\overline{\{S, T\}}}[\{W, X, Y, Z\} \cup \{S, T\}]$ (Line 2), which is the same as G in this example. While recording the first POMIS, the IB for $G_{\overline{\emptyset}}$, POMISS calls subPOMISS (Line 3) with $(Y, X, Z, T, W, S) \cap \{W, X, Z\} = (X, Z, W)$ for the parameter π assuming (Y, X, Z, T, W, S) is acquired among many reversed topological orders of H (Line 3). This corresponds to requesting

subPOMISs to compute the IBs for the passed causal graph with each variable in (X, Z, W) intervened (Lines 6–10). With $G_{\overline{X}}$, its corresponding MUCT and IB are $\{Y\}$ and $\{T, W, X\}$ (Fig. 3b). The IB will be recorded as a POMIS (Line 9) but no subsequent call for subPOMISs will be made since $(X, Z, W)^{i+1} \cap \{Y\} = \emptyset$. Given $G_{\overline{Z}}$, the same set of MUCT and IB is obtained as $G_{\overline{X}}$ (Fig. 3c). It is unnecessary to record the IB $\{T, W, X\}$ since it contains X , which precedes Z in the given order (Line 7, 8). The MUCT and IB given $G_{\overline{W}}$ are $\{X, Y, Z\}$ and $\{T, W\}$ (Fig. 3d), respectively. It will record the IB $\{T, W\}$, which is disjoint to $\{X, Z\}$. With $\{X, Y, Z\}$ as MUCT, the algorithm checks whether subsequent calls are necessary. However, since both X and Z precede W in the order, no recursive call is made. After all, the complete set of POMISs $\{\{S, T\}, \{T, W\}, \{T, W, X\}\}$ given $\llbracket G, Y \rrbracket$ will be returned only examining 4 IBs compared to $2^5=32$. If an order such that W precedes Z is considered, then a recursive call for subPOMISs will be made for $G_{\overline{\{W, Z\}}}$ after examining $G_{\overline{W}}$. It is unclear at this moment how different ordering affects the number of recursive calls.

Theorem 9 (Soundness and Completeness). *Given information $\llbracket G, Y \rrbracket$, the algorithm POMISs returns all and only POMISs.*

Proof. Let $\mathbb{P}_{G, Y}$ be all POMISs for $\llbracket G, Y \rrbracket$. Let π be an arbitrary sequence of \mathbf{V} such that $\mathbf{V} = \{\pi(1), \pi(2), \dots, \pi(n)\}$ where $n = |\mathbf{V}|$. Let $A \prec \mathbf{B}$ if $\forall B \in \mathbf{B} \pi^{-1}(A) < \pi^{-1}(B)$ and $A \preceq \mathbf{B}$ be similarly defined. For readability, let $\mathbf{T}^{(Q)} = \text{MUCT}(G_{\overline{\{Q\}}}, Y)$ and $\mathbf{X}^{(Q)} = \text{IB}(G_{\overline{\{Q\}}}, Y)$. We first show

$$\begin{aligned} \mathbb{P}_{G, Y} &= \{\text{IB}(G_{\overline{\mathbf{Z}}}, Y)\}_{\mathbf{Z} \subseteq \mathbf{T} \setminus \{Y\}} \\ &= \{\text{IB}(G, Y)\} \cup \{\text{IB}(G_{\overline{\mathbf{Z}}}, Y)\}_{\emptyset \neq \mathbf{Z} \subseteq \mathbf{T} \setminus \{Y\}} \\ &= \{\text{IB}(G, Y)\} \cup \bigcup_{Q \in \mathbf{T} \setminus \{Y\}} \{\text{IB}(G_{\overline{\{Q, \mathbf{Z}\}}}, Y)\}_{\mathbf{Z} \subseteq \mathbf{T} \setminus \{Q, Y\}: Q \prec \mathbf{Z}} \\ &= \{\text{IB}(G, Y)\} \cup \bigcup_{Q \in \mathbf{T} \setminus \{Y\}: Q \preceq \mathbf{X}^{(Q)}} \{\text{IB}(G_{\overline{\mathbf{X}^{(Q)}}}[\mathbf{T}^{(Q)} \cup \mathbf{X}^{(Q)}]_{\overline{\mathbf{Z}}}, Y)\}_{\mathbf{Z} \subseteq \mathbf{T} \setminus \{Y\}: Q \prec \mathbf{Z}} \end{aligned}$$

The third line partitions $2_{\neq \emptyset}^{\mathbf{T} \setminus \{Y\}}$, the power set of $\mathbf{T} \setminus \{Y\}$ excluding an empty set, so that sets of variables in each partition include the same variable, e.g., Q , with no variable smaller than Q with respect to π . The fourth line follows from Prop. 7 (change for \mathbf{Z}) and Prop. 8 (change of arguments for IB). An additional constraint for Q being $Q \preceq \mathbf{X}^{(Q)}$ avoids redundant computations since two different variables, e.g., Q' and Q'' , can both yield the same MUCT and IB, $\mathbf{T}^{(Q')} = \mathbf{T}^{(Q')}$ and $\mathbf{X}^{(Q')} = \mathbf{X}^{(Q')}$. Since

$$\{\text{IB}(G_{\overline{\mathbf{X}^{(Q)}}}[\mathbf{T}^{(Q)} \cup \mathbf{X}^{(Q)}]_{\overline{\mathbf{Z}}}, Y)\}_{\mathbf{Z} \subseteq \mathbf{T} \setminus \{Y\}} = \mathbb{P}_{G_{\overline{\mathbf{X}^{(Q)}}}[\mathbf{T}^{(Q)} \cup \mathbf{X}^{(Q)}], Y},$$

we can rewrite $\mathbb{P}_{G, Y}$ as (abusing notations)

$$\mathbb{P}_{G, Y} = \{\text{IB}(G, Y)\} \cup \bigcup_{Q \in \mathbf{T} \setminus \{Y\}: Q \preceq \mathbf{X}^{(Q)}} \mathbb{P}_{G_{\overline{\mathbf{X}^{(Q)}}}[\mathbf{T}^{(Q)} \cup \mathbf{X}^{(Q)}], Y}^{Q \prec \mathbf{Z}}$$

where the superscript $Q \prec \mathbf{Z}$ serves as an additional constraint for efficiency.

The algorithm implements the above equality where parameter π in subPOMISs carries the constraint $Q \prec \mathbf{Z}$ and parameter \mathbf{O} conveys the constraint $Q \preceq \mathbf{X}^{(Q)}$. Let \mathbf{W} be an arbitrary POMIS in $\mathbb{P}_{G, Y}$. We can index its element (i.e., variable) $\mathbf{W} = \{W_i\}_{i=1}^{|\mathbf{W}|}$ with respect to π so as to $\pi^{-1}(W_i) < \pi^{-1}(W_j)$ if $i < j$. Then, there exists a sequence of recursive calls of subPOMISs where variables in \mathbf{W} are sequentially determined to be intervened skipping the already determined to do so (i.e., $\mathbf{W} \cap \text{IB}(G', Y)$ where G' is the first argument of subPOMISs). Therefore, the algorithm completely enumerates all POMISs effectively avoiding redundant computations. \square

Appendix D Experiments

Task 1: $P(U_{X_1} = 1) = 0.54$, $P(U_{X_2} = 1) = 0.67$, $P(U_Y = 1) = 0.58$, $P(U_{Z_1} = 1) = 0.54$, and $P(U_{Z_2} = 1) = 0.44$, and functions:

$$\begin{aligned} f_{Z_1}(u_{Z_1}) &= u_{Z_1} \\ f_{Z_2}(u_{Z_2}) &= u_{Z_2} \\ f_{X_1}(z_1, z_2, u_{X_1}) &= z_1 \oplus z_2 \oplus u_{X_1} \\ f_{X_2}(z_1, z_2, u_{X_2}) &= 1 \oplus z_1 \oplus z_2 \oplus u_{X_2} \\ f_Y(x_1, x_2, u_Y) &= (x_1 \& x_2) | u_Y \end{aligned}$$

Task 2: $P(U_X = 1) = 0.11$, $P(U_Y = 1) = 0.15$, $P(U_Z = 1) = 0.6$, and $P(U_{XY} = 1) = 0.51$ and functions:

$$\begin{aligned} f_Z(u_Z) &= u_Z \\ f_X(z, u_X, u_{XY}) &= u_X \oplus u_{XY} \oplus z \\ f_Y(x, u_Y, u_{XY}) &= 1 \oplus u_Y \oplus u_{XY} \oplus x \end{aligned}$$

Task 3: $P(U_S = 1) = 0.45$, $P(U_T = 1) = 0.81$, $P(U_W = 1) = 0.07$, $P(U_X = 1) = 0.06$, $P(U_Y = 1) = 0.06$, $P(U_Z = 1) = 0.05$, $P(U_{WX} = 1) = 0.51$, and $P(U_{YZ} = 1) = 0.54$, and functions:

$$\begin{aligned} f_S(u_S) &= u_S \\ f_T(u_T) &= u_T \\ f_W(s, u_W, u_{WX}) &= u_W \oplus u_{WX} \oplus s \\ f_Z(u_Z, u_{YZ}) &= u_Z \oplus u_{YZ} \\ f_X(t, z, u_X, u_{WX}) &= 1 \oplus t \oplus z \oplus u_X \oplus u_{WX} \\ f_Y(t, w, x, u_Y, u_{YZ}) &= t \oplus w \oplus x \oplus u_Y \oplus u_{YZ} \end{aligned}$$