On Transportability for Structural Causal Bandits

Min Woo Park[1] **Sanghack Lee***[1]

[1]Graduate School of Data Science, Seoul National University, Seoul, Republic of Korea

Abstract

The *structural causal bandit* (SCB) framework offers a graphical approach to identifying suboptimal actions by leveraging prior knowledge of the underlying causal structure. While theoretically appealing, there has been limited guidance on how to systematically transfer information across heterogeneous datasets from multiple environments. In this paper, we study the structural causal bandit under *transportability*, where prior knowledge from source environments is integrated to accelerate learning in a target deployment environment. We show that exploiting causal invariances across environments improves sample efficiency in identifying the optimal action.

1 INTRODUCTION

The multi-armed bandit (MAB) problem [Robbins, 1952, Lai and Robbins, 1985, Lattimore and Szepesvári, 2020] is a central topic in sequential decision-making, in which a learner aims to maximize cumulative rewards by repeatedly choosing actions based on observed rewards, balancing the exploration-exploitation trade-off.

Integrating causal knowledge into the decision-making process enables a learner to model decision problems with rich dependency structures [Zhang et al., 2020, Kumor et al., 2021, Zhang and Bareinboim, 2022, Ruan et al., 2024]. Structural causal models (SCMs) [Pearl, 2000] provide a principled representation of causal relationships among actions, rewards, and other relevant variables such as contexts or states. This approach enables learners to make informed decisions by considering how each action causally influences the reward through causal pathways [Bareinboim et al., 2024].

Existing work [Bareinboim et al., 2015, Lattimore et al., 2016, Forney et al., 2017] demonstrates that incorporating causal knowledge can significantly outperform approaches that ignore causal dependencies. Subsequent work has explored various specialized settings by introducing additional structural assumptions [Lu et al., 2020, Bilodeau et al., 2022, Feng and Chen, 2023, Varici et al., 2023]. Specifically, Lee and Bareinboim [2018] introduced the *structural causal bandit* (SCB) framework without parametric assumptions, and Lee and Bareinboim [2019] extended it to settings with non-manipulable variables.

While SCB effectively prunes suboptimal actions, substantial exploration may still be required, which can be costly in practice. A complementary strategy to reduce exploration cost is to leverage data collected from related source environments. Informative prior data may help narrow reward distributions and mitigate the *cold-start* problem, enabling faster identification of the optimal action. However, environmental discrepancies can be substantial, and naively transferring data may degrade performance in the target deployment environment. In the causality literature, this problem falls under the rubric of *transportability theory* [Pearl and Bareinboim, 2011, Bareinboim and Pearl, 2012b, 2016], which provides methods for determining when and how a causal effect can be computed across different environments.

Zhang and Bareinboim [2017] studied the transfer of prior observations in settings with specific graph structures, e.g., bow graphs and instrumental-variable (IV) graphs. Bellot et al. [2023] and Deng et al. [2025] extended this line of work to general causal diagrams, although their focus was limited to single-node interventions.

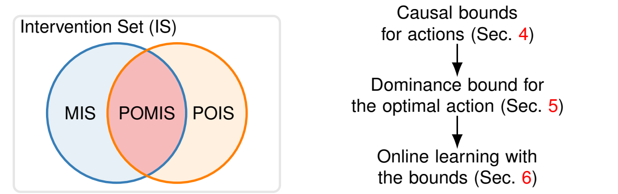
Contributions. We extend the SCB framework to leverage heterogeneous prior data while accounting for structural discrepancies across environments.

After presenting the preliminaries and problem formulation in Secs. 2 and 3, we introduce a method for estimating *causal bounds* by systematically combining source datasets

*Corresponding author: sanghack@snu.ac.kr.

in Sec. 4. Furthermore, we establish hierarchical relations among actions and derive a corresponding *dominance bound* on the optimal action, as detailed in Sec. 5. Finally, we integrate these bounds into the MAB problem to guide exploration and improve sample efficiency, as corroborated by the experimental results in Sec. 6.

We defer the proofs of all theoretical results to Appendix C.



(a) Euler diagram (b) Overall flow

Figure 1: (a) Relationship among intervention sets; (b) Tightening causal bounds via dominance relations (dominance bound) and integrating them into online learning.

2 PRELIMINARIES

We introduce notation and review relevant prior work. Following conventions, we use a capital letter, such as X, to represent a variable, with its corresponding lowercase letter, x, denoting a realization of the variable. Boldface letters denote sets of variables or values, written as $\mathbf{X}$ or $\mathbf{x}$. The domain of X is denoted by Ω_X and $\Omega_{\mathbf{X}} = \times_{X \in \mathbf{X}} \Omega_X$. We use calligraphic letters for graphs and models such as $\mathcal{G}$ and $\mathcal{M}$. The distribution over variables $\mathbf{X}$ is denoted by $P(\mathbf{X})$. We consistently use $P(\mathbf{x})$ as shorthand for $P(\mathbf{X} = \mathbf{x})$. We denote by $\mathbb{I}\{\mathbf{X} = \mathbf{x}\}$ the indicator function. We denote by $\mathbf{x} \setminus \mathbf{Z}$ the value of $\mathbf{X} \setminus \mathbf{Z}$ consistent with $\mathbf{x}$ and by $\mathbf{x} \cap \mathbf{Z}$ the subset of $\mathbf{x}$ corresponding to variables in $\mathbf{Z}$.

Structural Causal Models. We use a *structural causal model* (SCM) [Pearl, 2000] as the semantic framework to represent the underlying environment in which a decision maker is deployed. An SCM $\mathcal{M}$ is a quadruple $\langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{u}) \rangle$, where $\mathbf{U}$ is a set of exogenous variables determined by factors outside the model and distributed according to a joint distribution $P(\mathbf{u})$, and $\mathbf{V}$ is a set of endogenous variables whose values are determined by a collection of functions $\mathcal{F} = \{f_V\}_{V \in \mathbf{V}}$ such that $v \leftarrow f_V(\mathbf{pa}_V, \mathbf{u}_V)$ where $\mathbf{Pa}_V \subseteq \mathbf{V} \setminus \{V\}$ and $\mathbf{U}_V \subseteq \mathbf{U}$. The observational probability $P(\mathbf{v})$ is defined as $\int_{\mathbf{u}} \prod_{V \in \mathbf{V}} \mathbb{I}\{f_V(\mathbf{pa}_V, \mathbf{u}_V) = v\} dP(\mathbf{u})$. Every SCM $\mathcal{M}$ is associated with a *causal diagram* (also called a semi-Markovian graph) $\mathcal{G} = \langle \mathbf{V}, \mathbf{E} \rangle$ where a directed edge $V_i \rightarrow V_j \in \mathbf{E}$ is present if $V_i \in \mathbf{Pa}_{V_j}$, and a bidirected edge is present between V_i and V_j if $\mathbf{U}_{V_i}$ and $\mathbf{U}_{V_j}$ are dependent. The probability of $\mathbf{V} = \mathbf{v}$ when $\mathbf{X}$ is intervened upon to take the value $\mathbf{x}$ is denoted by $P(\mathbf{v} \mid do(\mathbf{x}))$ or $P_{\mathbf{x}}(\mathbf{v})$, and the submodel induced by the intervention is denoted by $\mathcal{M}_{\mathbf{x}}$.

Graphical notations. We use kinship notation for graphical relationships: parents ($\mathtt{Pa}$), children ($\mathtt{Ch}$), descendants ($\mathtt{De}$), and ancestors ($\mathtt{An}$). Ancestors and descendants include the variable itself. For a set of variables, we define the ancestral set as $\mathtt{An}(\mathbf{X})_{\mathcal{G}} = \bigcup_{X \in \mathbf{X}} \mathtt{An}(X)_{\mathcal{G}}$, and similarly for other relationships. $\mathcal{G}_{\overline{\mathbf{W}}\underline{\mathbf{X}}}$ is the result of removing edges coming into variables in $\mathbf{W}$ and going out from variables in $\mathbf{X}$. We denote the set of variables in $\mathcal{G}$ by $\mathbf{V}(\mathcal{G})$. A subgraph $\mathcal{G}[\mathbf{V}']$, where $\mathbf{V}' \subseteq \mathbf{V}(\mathcal{G})$, is defined as a vertex-induced subgraph in which all edges among the vertices in $\mathbf{V}'$ are preserved. We define $\mathcal{G} \setminus \mathbf{X}$ as $\mathcal{G}[\mathbf{V}(\mathcal{G}) \setminus \mathbf{X}]$ for $\mathbf{X} \subseteq \mathbf{V}(\mathcal{G})$. We denote by $\mathcal{G}\langle \mathbf{X} \rangle$ the latent projection of $\mathcal{G}$ onto $\mathbf{X}$.

3 STRUCTURAL CAUSAL BANDITS WITH TRANSPORTABILITY

We introduce the structural causal bandit (SCB) problem in which a learner interacts with a target system modeled by a structural causal model (SCM) $\mathcal{M}^*$ that includes a reward variable $Y \in \mathbf{V}$. We assume that $Y \in [0, 1]$.

In this setting, pulling each arm corresponds to an intervention $do(\mathbf{x})$ on some intervention set (IS) $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\} \setminus \mathbf{N}$, where $\mathbf{N}$ denotes non-manipulable variables. When $\mathbf{N} = \emptyset$, we refer to the setting as *unconstrained*. Otherwise, it is *constrained*, as interventions are restricted by the presence of non-manipulable variables. We use the terms *arm*, *action*, and *intervention* interchangeably, depending on the context. The learner does not have direct access to the target SCM $\mathcal{M}^*$, but can observe $\mathbf{V}$ through online interactions by pulling arms.

Action space refinement. Given a causal diagram, restricting attention to a *possibly-optimal* action space can lead to improved performance. Let $\mu_{\mathbf{x}}$ denote the mean reward of $do(\mathbf{x})$. We denote by $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \Omega_{\mathbf{X}}} \mu_{\mathbf{x}}$ the best arm within $\Omega_{\mathbf{X}}$.

Definition 1 (Minimality [Lee and Bareinboim, 2018])**.** Let $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\} \setminus \mathbf{N}$ be a set for which no strict subset $\mathbf{X}' \subsetneq \mathbf{X}$ satisfies $\mu_{\mathbf{x}} = \mu_{\mathbf{x} \cap \mathbf{X}'}$[1]. Then, we refer to $\mathbf{X}$ as a *minimal* intervention set (MIS).

Intuitively, minimality implies that every variable $X \in \mathbf{X}$ affects the reward variable without passing through $\mathbf{X} \setminus \{X\}$ in $\mathcal{G}$. An MIS $\mathbf{X}$ is considered a possibly-optimal MIS (POMIS) [Lee and Bareinboim, 2018, 2019] if some intervention $\mathbf{x} \in \Omega_{\mathbf{X}}$ can achieve the optimal expected reward in some SCM consistent with a causal diagram $\mathcal{G}$.

We denote by $\mathbb{M}^{\mathbf{N}}_{\mathcal{G},Y}$ the collection of MISs and by $\mathbb{P}^{\mathbf{N}}_{\mathcal{G},Y}$ the collection of POMISs with respect to $\langle \mathcal{G}, Y, \mathbf{N} \rangle$. Moreover,

[1]We refer to $\mathbf{X}$ and $\mathbf{X}'$ as *equivalent* if the equality holds.

we denote by $\mathcal{A}^{\mathbf{N}}_{\mathcal{G},Y} = \{\mathbf{x} \in \Omega_{\mathbf{X}} \mid \mathbf{X} \in \mathbb{P}^{\mathbf{N}}_{\mathcal{G},Y}\}$ the corresponding action set[2]. Hence, given $\langle \mathcal{G}, Y, \mathbf{N} \rangle$, the optimal action can be expressed as:

$$\mathbf{x}^{\star} \triangleq \arg \max_{\mathbf{x} \in \mathcal{A}^{\mathbf{N}}_{\mathcal{G},Y}} \mu_{\mathbf{x}} \tag{1}$$

which restricts the search space to POMISs. Accordingly, a learner informed by the POMIS structure should confine exploration to actions consistent with these sets. Furthermore, if $\mathbf{X}$ is equivalent to a POMIS but not minimal, we refer to it as a POIS, as illustrated in Fig. 1a. Related graphical characterizations are provided in Appendix B.

Structural causal bandits with transportability. The learner has access to data from one or more related source environments $\Pi = \{\pi^1, \pi^2, \cdots, \pi^n\}$, each associated with SCMs $\mathcal{M}^1, \mathcal{M}^2, \cdots, \mathcal{M}^n$. The distributions associated with π^i under $do(\mathbf{x})$ are denoted by $P^i_{\mathbf{x}}$. We use the superscript $*$ throughout this paper to consistently denote the target environment, e.g., $\mathcal{M}^*$ and $P^*_{\mathbf{x}}$.

Let $\mathbf{x}_t$ denote the action selected at round t. The cumulative regret is defined as:

$$R_T \triangleq T\mu^*_{\mathbf{x}^\star} - \sum_{t=1}^{T} \mu^*_{\mathbf{x}_t} = \sum_{\mathbf{x} \in \mathcal{A}^{\mathbf{N}}_{\mathcal{G},Y}} \Delta_{\mathbf{x}} N_T(\mathbf{x}) \tag{2}$$

which compares the cumulative reward of the optimal arm in π^* with that obtained by the selected actions. Here, $\Delta_{\mathbf{x}} \triangleq \mu^*_{\mathbf{x}^\star} - \mu^*_{\mathbf{x}}$ denotes the suboptimality gap, and $N_T(\mathbf{x})$ denotes the number of times $do(\mathbf{x})$ is selected up to round T. The objective in the SCB setting is to minimize the expected cumulative regret in the target environment π^*, i.e., $\mathbb{E}R_T$.

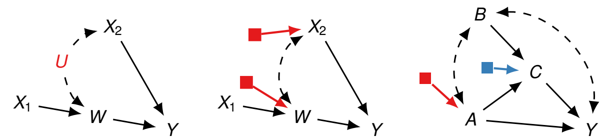
Graphical encoding of environmental discrepancies. To account for environmental shifts, we introduce a *selection diagram* [Bareinboim and Pearl, 2012b], which explicitly represents discrepancies across environments.

Definition 2 (Environment discrepancy)**.** Let π^i and π^j be environments associated with SCMs $\mathcal{M}^i$ and $\mathcal{M}^j$ conforming to a causal diagram $\mathcal{G}$. We denote by $\Delta^{i,j}$ a set of variables such that, for every $V \in \Delta^{i,j}$, there is a discrepancy; either $f^i_V \neq f^j_V$ or $P^i(\mathbf{U}_V) \neq P^j(\mathbf{U}_V)$[3].

Definition 3 (Selection diagram)**.** Given a collection of discrepancies $\boldsymbol{\Delta} = \{\Delta^{*,i}\}^n_{i=1}$ with regard to $\mathcal{G} = \langle \mathbf{V}, \mathbf{E} \rangle$, let $\mathbf{S}^i = \{S_V \mid V \in \Delta^{*,i}\}$ be *selection nodes*. The graph $\mathcal{G}^{\Delta^{*,i}} = \langle \mathbf{V} \cup \mathbf{S}^i, \mathbf{E} \cup \{S_V \to V\}_{S_V \in \mathbf{S}^i} \rangle$ is called a *selection diagram*. Let $\mathbf{S} = \bigcup^n_{i=1} \mathbf{S}^i$. The *collective selection diagram* $\mathcal{G}^{\boldsymbol{\Delta}}$ is defined as $\langle \mathbf{V} \cup \mathbf{S}, \mathbf{E} \cup \{S_V \to V\}_{S_V \in \mathbf{S}} \rangle$.

[2]We sometimes omit $\mathbf{N}$ when $\mathbf{N} = \emptyset$, e.g., $\mathbb{P}_{\mathcal{G},Y}$ and $\langle \mathcal{G}, Y \rangle$. Strictly speaking, actions are interventions $do(\mathbf{x})$ rather than assignments $\mathbf{x}$; we nevertheless use $\mathbf{x}$ to denote the corresponding action when no ambiguity arises.

[3]Superscripts (e.g., Δ^i) indicate environment discrepancies, while subscripts (e.g., $\Delta_{\mathbf{x}}$) denote suboptimality gaps.



(a) (b) $\mathcal{G}^{\boldsymbol{\Delta}}$ (c) Multiple sources

Figure 2: (a) Difference in U between the source and target populations; (b) collective selection diagram for (a); (c) collective selection diagram with $\Delta^1 = \{A\}$ and $\Delta^2 = \{C\}$.

We shorten $\Delta^{*,i}$ as Δ^i. The collective selection diagram $\mathcal{G}^{\boldsymbol{\Delta}}$ encodes qualitative information about $\mathcal{M}^*$ and discrepancies $\boldsymbol{\Delta}$. The absence of a selection node pointing to a variable indicates that the causal mechanism responsible for assigning values to that variable is identical across all environments corresponding to π^* and $\Pi = \{\pi^i\}^n_{i=1}$.

One can view the selection nodes $\mathbf{S}$ as *switches* controlling environment shifts, and the collective selection diagram $\mathcal{G}^{\boldsymbol{\Delta}}$ as the causal diagram for a unified SCM representing heterogeneous SCMs satisfying the following:

$$P^{\boldsymbol{\Delta}}_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w}, \mathbf{S}^i = \mathbf{i}, \mathbf{S}^{-i} = \mathbf{0}) = P^i_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$$

and

$$P^{\boldsymbol{\Delta}}_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w}, \mathbf{S} = \mathbf{0}) = P^*_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$$

where $\mathbf{S}^{-i} = \mathbf{S} \setminus \mathbf{S}^i$[4].

This representation enables probabilistic operations across the target environment π^* and sources Π.

Example (Cardiovascular Disease Treatment). Consider a scenario in which Y represents *cardiovascular disease*, W denotes *blood pressure*, X_1 corresponds to the intake of an *antihypertensive drug*, X_2 to the use of an *anti-diabetic drug*, and U captures unobserved factors (e.g., physical activity levels and dietary patterns; Ferrannini and Cushman [2012]).

Suppose that prior data from a source population are available to design a population-level treatment strategy for patients with cardiovascular disease in a target population, with the goal of determining appropriate medications. Fig. 2a graphically illustrates this scenario. Although such data can be informative, they must be handled with care, particularly when the source and target populations are suspected to differ, since a strategy that is optimal in the source population may become suboptimal in the target due to differences in the distribution of U.

We illustrate a collective selection diagram $\mathcal{G}^{\boldsymbol{\Delta}}$ corresponding to this example in Fig. 2b. In this scenario, the distributions of W and X_2 may differ between $\mathcal{M}^*$ and $\mathcal{M}^1$ due to differences in the distribution of the unobserved confounder

[4]$\mathbf{S}^i = \mathbf{i}$ means $S_V = i$ for all $V \in \Delta^i$.

U, which influences both variables. This is consistent with $\mathbf{\Delta} = \{\Delta^1 = \{W, X_2\}\}$ and $\mathbf{S} = \mathbf{S}^1 = \{S_W, S_{X_2}\}$. In the diagram, the selection nodes S_W and S_{X_2} are represented as red squares.

Fig. 2c further illustrates a collective selection diagram in which the distribution of A may differ between $\mathcal{M}^*$ and $\mathcal{M}^1$, and the distribution of C may differ between $\mathcal{M}^*$ and $\mathcal{M}^2$. This corresponds to $\mathbf{\Delta} = \{\Delta^1 = \{A\}, \Delta^2 = \{C\}\}$ and $\mathbf{S} = \{S_A, S_C\}$. The selection nodes S_A and S_C are represented as red and blue squares respectively.

Outline. The remainder of the paper is organized as follows. We first obtain causal bounds $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$ for $\mu^*_{\mathbf{x}}$ from source environments whenever permitted. Next, we derive a dominance bound $[\ell^\star, u^\star]$ for $\mu^*_{\mathbf{x}^\star}$ based on hierarchical knowledge. Finally, we incorporate these bounds into a MAB algorithm. Fig. 1b illustrates the overall flow.

4 PARTIAL TRANSPORTABILITY AND CAUSAL BOUNDS

Transportability. Let $\mathbb{Z} = \{\mathbb{Z}^i\}_{i=1}^n$ be a collection of interventions $\mathbb{Z}^i$ conducted in the source environment π^i. Then, we say that $\mu^*_{\mathbf{x}}$ is *transportable* with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z}\rangle$ if $\mu^*_{\mathbf{x}}$ is uniquely computable from source distributions $\mathcal{P}^{\Pi}_{\mathbb{Z}} = \{P^i_{\mathbf{z}} \mid \mathbf{z} \in \Omega_{\mathbf{Z}}, \mathbf{Z} \in \mathbb{Z}^i \in \mathbb{Z}\}$ in any collection of models consistent with $\mathcal{G}^{\mathbf{\Delta}}$ [Lee et al., 2020].

As an intuition pump, consider the collective selection diagram $\mathcal{G}^{\mathbf{\Delta}}$ in Fig. 2c, and suppose that the first dataset from π^1 is collected under $\mathbb{Z}^1 = \{\emptyset\}$ while the dataset from π^2 is obtained under $\mathbb{Z}^2 = \{\{A, C\}\}$. The target environment π^* is assumed to be subject to the constraint $\mathbf{N}^* = \{C\}$. Our objective is to identify transportable quantities to minimize unnecessary exploration by leveraging prior information from π^1 and π^2. As a representative instance, consider $a \in \mathcal{A}^{\mathbf{N}^*}_{\mathcal{G},Y}$. Then, the following holds:

$$\begin{aligned}
P^*_a(y) &= \textstyle\sum_{b,c} P^*_a(b) P^*_a(c \mid b) P^*_a(y \mid b, c) \\
&\overset{\text{(a)}}{=} \textstyle\sum_{b,c} P^*_{a,c}(b) P^*_{\emptyset}(c \mid a, b) P^*_{a,c}(y \mid b) \\
&= \textstyle\sum_{b,c} P^*_{\emptyset}(c \mid a, b) P^*_{a,c}(b, y) \\
&\overset{\text{(b)}}{=} \textstyle\sum_{b,c} P^1_{\emptyset}(c \mid a, b) P^*_{a,c}(b, y) \\
&\overset{\text{(c)}}{=} \textstyle\sum_{b,c} P^1_{\emptyset}(c \mid a, b) P^2_{a,c}(b, y).
\end{aligned} \tag{3}$$

The equality (a) follows from applications of Rule 2 (converting conditioning on c to $do(c)$ and $do(a)$ to conditioning on a) and Rule 3 (adding $do(c)$) of do-calculus [Pearl, 1995].

Equality (b) holds since the discrepancy between π^* and π^1 is irrelevant due to $(S_A \perp\!\!\!\perp_d C \mid A, B)_{\mathcal{G}^{\mathbf{\Delta}}}$.

The final equality (c) follows from the invariance of the mechanisms f_B and f_Y between π^* and π^2. We observe that $P^*_a(y)$ can be expressed in terms of $P^1_{\emptyset}$ and $P^2_{a,c}$, implying that $\mu^*_a = \sum_y y P^*_a(y)$ can be estimated from the sources Π. Thus, we say μ^*_a is transportable relative to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z}\rangle$.

4.1 PARTIAL TRANSPORTABILITY

When $\mu^*_{\mathbf{x}}$ is *not* transportable, it may seem that a learner cannot benefit from sources and must estimate rewards entirely from scratch through *cold* exploration. However, the learner may still leverage partial knowledge from the prior to improve estimates within a feasible interval.

Formally, we say that $\mu^*_{\mathbf{x}}$ is *partially transportable* with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z}\rangle$ if it can be bounded by an informative interval $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$ that remains valid over $\mathcal{P}^{\Pi}_{\mathbb{Z}}$ across all models inducing $\mathcal{G}^{\mathbf{\Delta}}$. To construct these bounds, we first introduce the *Q-decomposition* [Tian and Pearl, 2003], which decomposes the target query $\mu^*_{\mathbf{x}}$ into finer-grained components.

Q-decomposition. Let $\mathsf{cc}_{\mathcal{G}} = \{\mathbf{C}_q\}_{q=1}^m$ be the collection of *c-components* of $\mathcal{G}$.

Definition 4 (C-component [Tian and Pearl, 2003])**.** A *c-component* is a maximal bidirected-connected set of vertices in a semi-Markovian graph.

For $\mathbf{C} \subseteq \mathbf{V}$, we define the quantity $Q[\mathbf{C}](\mathbf{v}) = P_{\mathbf{v}\setminus\mathbf{C}}(\mathbf{c})$, which corresponds to the post-intervention distribution and is referred to as a *c-factor*. For convenience, we omit the input $\mathbf{v}$ and write $Q[\mathbf{C}]$.

We denote the quantities for the target environment π^* as:

$$Q^*[\mathbf{C}] = P^*_{\mathbf{v}\setminus\mathbf{C}}(\mathbf{c}) = P^{\mathbf{\Delta}}_{\mathbf{v}\setminus\mathbf{C}}(\mathbf{c} \mid \mathbf{S} = \mathbf{0}) \tag{4}$$

and for source environments π^i as:

$$Q^i[\mathbf{C}] = P^i_{\mathbf{v}\setminus\mathbf{C}}(\mathbf{c}) = P^{\mathbf{\Delta}}_{\mathbf{v}\setminus\mathbf{C}}(\mathbf{c} \mid \mathbf{S}^i = \mathbf{i}, \mathbf{S}^{-i} = \mathbf{0}). \tag{5}$$

We denote by $\mathbf{Y}^+$ the set of ancestors of Y in $\mathcal{G}_{\underline{\mathbf{X}}}$. Then $\mu^*_{\mathbf{x}}$ admits a unique factorization in terms of m c-factors $\mathbf{C}_q \in \mathsf{cc}_{\mathcal{G}[\mathbf{Y}^+]}$ as follows:

$$\mu^*_{\mathbf{x}} = \sum_y y P^*_{\mathbf{x}}(y) = \sum_{\mathbf{y}^+} y \prod_{q=1}^m Q^*[\mathbf{C}_q]. \tag{6}$$

When $\mathbf{C}$ satisfies $\mathbf{C} \cap \Delta^i = \emptyset$ and there exists a c-component $\mathbf{C} \subseteq \mathbf{C}'$ such that $Q^i[\mathbf{C}]$ is *identifiable* from $\mathcal{G}[\mathbf{C}']$[5], we refer to $Q^*[\mathbf{C}]$ as being *transportable* from π^i [Lee et al., 2020]. Furthermore, $\mu^*_{\mathbf{x}}$ is transportable with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z}\rangle$ *if and only if* all c-factors $Q^*[\mathbf{C}_q]$ in the right-hand side of Eq. (6) are transportable from some source environment $\pi^i \in \Pi$.

An intuitive metaphor is the jigsaw puzzle in Fig. 3. The query and the underlying causal diagram define the puzzle, while each source distribution provides individual pieces.

[5] This means that $Q^i[\mathbf{C}]$ is uniquely computable from $\mathcal{G}[\mathbf{C}']$.

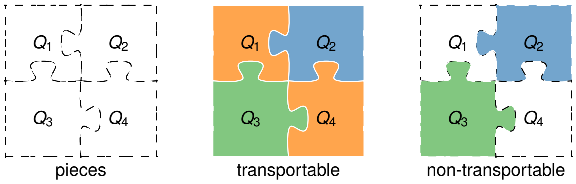


Figure 3: A jigsaw abstraction. The target $\mu^*_{\mathbf{x}}$ is decomposed into c-factor pieces, corresponding to the right-hand side of Eq. (6). Each color represents a source environment that provides a subset of these pieces. If all pieces can be filled regardless of color, then $\mu^*_{\mathbf{x}}$ is transportable; otherwise, it is non-transportable.

When suitable pieces are unavailable, partial transportability replaces them with surrogate pieces corresponding to lower and upper bounds.

Partial transportability. We revisit the collective selection diagram $\mathcal{G}^{\mathbf{\Delta}}$ in Fig. 2c in a more challenging setting. Suppose $\mathbb{Z}^1 = \{\emptyset, \{C\}\}$ and $\mathbb{Z}^2 = \{\{B\}\}$. We reformulate Eq. (3) in terms of c-factors, following Eq. (6):

$$\mu^*_a = \sum_y y P^*_a(y) = \sum_{b,c,y} y Q^*[C] Q^*[B, Y]. \tag{7}$$

In this setting, $Q^*[C]$ is identifiable (i.e., $= P^1_{\emptyset}(c \mid a, b)$). However, $Q^1[B, Y]$ is not identifiable from $\mathcal{G}[\mathbf{V} \setminus \mathbf{Z}]$ for any $\mathbf{Z} \in \mathbb{Z}^1$. Consequently, $Q^*[B, Y]$ is not transportable with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$. Nevertheless, we still have:

$$Q^*[B, Y] = Q^1[B, Y] \tag{8}$$

since $\{B, Y\} \cap \Delta^1 = \emptyset$. Therefore, any valid bound for $Q^1[B, Y]$ immediately applies to $Q^*[B, Y]$. The *natural bound* [Manski, 1990, Robins, 1989] provides the following:

$$P(\mathbf{x}, \mathbf{y}) \leq P_{\mathbf{x}}(\mathbf{y}) \leq P(\mathbf{x}, \mathbf{y}) + 1 - P(\mathbf{x}). \tag{9}$$

We generalize the bound to heterogeneous c-factors.

Proposition 1. *Let $Q^i[\mathbf{D}]$ be an identifiable c-factor from $\pi^i \in \Pi$ such that $\mathbf{C} \cap \Delta^i = \emptyset$ and $\mathbf{C} \subsetneq \mathbf{D}$. Let $\mathbf{C}' = \mathtt{An}(\mathbf{C})_{\mathcal{G}[\mathbf{D}]}$. Suppose that $\mathbf{C} = \mathtt{An}(\mathbf{C})_{\mathcal{G}[\mathbf{C}']}$. Then $Q^*[\mathbf{C}]$ is bounded as follows. If $Q^i[\mathbf{C}']$ is not identifiable in π^i, then:*

$$\sum_{\mathbf{c}' \setminus \mathbf{C}} Q^i[\mathbf{D}] \leq Q^*[\mathbf{C}] \tag{10}$$

$$\leq \sum_{\mathbf{c}' \setminus \mathbf{C}} \{Q^i[\mathbf{D}] + 1 - \sum_{\mathbf{c}'} Q^i[\mathbf{D}]\}. \tag{11}$$

Otherwise, $Q^*[\mathbf{C}] = \sum_{\mathbf{c}' \setminus \mathbf{C}} Q^i[\mathbf{C}']$ (i.e., transportable).

According to Prop. 1, since $Q^1[A, B, Y]$ is identifiable from π^1, we can establish the following partially transportable bounds for μ^*_a:

$$\sum_{b,c,y} y P^1_{\emptyset}(c \mid a, b) \{P^1_c(a, b, y)\} \leq \mu^*_a \tag{12}$$

Algorithm 1: Partial transportability (PATR)

Input: $\mathbf{y}$; $\mathbf{x}$; $\mathcal{G}$; $\mathbb{Z}$ (global); $\mathbf{\Delta}$; type $\in \{\ell, u\}$: type of bound

1 **function** PATR($\mathbf{y}, \mathbf{x}, \mathcal{G}, \mathbf{\Delta}$, type)
2   **if** $\exists \mathbf{Z} \in \mathbb{Z}^i \in \mathbb{Z}$ such that $(\mathbf{X} = \mathbf{Z} \cap \mathbf{V}) \wedge (\mathbf{S}^i \perp\!\!\!\perp_d \mathbf{Y})_{\mathcal{G}^{\mathbf{\Delta}} \setminus \mathbf{X}}$ **then yield** $P^i_{\mathbf{z} \setminus \mathbf{V}, \mathbf{x} \cap \mathbf{Z}}(\mathbf{y})$
3   **if** $\mathbf{V}' := \mathbf{V} \setminus \mathtt{An}(\mathbf{Y})_{\mathcal{G}} \neq \emptyset$ **then yield** PATR($\mathbf{y}, \mathbf{x} \setminus \mathbf{V}', \mathcal{G} \setminus \mathbf{V}', \{\Delta^i \setminus \mathbf{V}' \mid \Delta^i \in \mathbf{\Delta}\}$, type)
4   **if** $\mathbf{V}' := (\mathbf{V} \setminus \mathbf{X}) \setminus \mathtt{An}(\mathbf{Y})_{\mathcal{G} \setminus \mathbf{X}} \neq \emptyset$ **then yield** PATR($\mathbf{y}, \mathbf{x} \cup \mathbf{V}', \mathcal{G}, \mathbf{\Delta}$, type)
5   **if** $|\mathsf{cc}_{\mathcal{G} \setminus \mathbf{X}}| > 1$ **then yield** $\sum_{\mathbf{v} \setminus (\mathbf{X} \cup \mathbf{Y})} \prod_{\mathbf{C} \in \mathsf{cc}_{\mathcal{G} \setminus \mathbf{X}}}$ PATR($\mathbf{c}, \mathbf{v} \setminus \mathbf{c}, \mathcal{G}, \mathbf{\Delta}$, type)
6   **for** $\pi^i \in \Pi$ such that $\Delta^i \cap (\mathbf{V} \setminus \mathbf{X}) = \emptyset$, **for** $\mathbf{Z} \in \mathbb{Z}^i$ such that $\mathbf{Z} \cap \mathbf{V} \subseteq \mathbf{X}$ **do**
7   **yield** PAID($\mathbf{y}, \mathbf{x} \setminus \mathbf{Z}, P^i_{\mathbf{z} \setminus \mathbf{V}, \mathbf{x} \cap \mathbf{Z}}, \mathcal{G} \setminus (\mathbf{Z} \cap \mathbf{X})$, type)

8 **function** PAID($\mathbf{y}, \mathbf{x}, P, \mathcal{G}$, type)
9     $\{\mathbf{C}\} \leftarrow \mathsf{cc}_{\mathcal{G} \setminus \mathbf{X}}$
10   **if** $\mathbf{X} = \emptyset$ **then yield** $\sum_{\mathbf{v} \setminus \mathbf{Y}} P(\mathbf{v})$
11   **if** $\mathbf{V} \neq \mathtt{An}(\mathbf{Y})_{\mathcal{G}}$ **then yield** PAID($\mathbf{y}, \mathbf{x} \cap \mathtt{An}(\mathbf{Y})_{\mathcal{G}}, \sum_{\mathbf{v} \setminus \mathtt{An}(\mathbf{Y})_{\mathcal{G}}} P(\mathbf{v}), \mathcal{G}[\mathtt{An}(\mathbf{Y})_{\mathcal{G}}]$, type)
12   **if** $\mathsf{cc}_{\mathcal{G}} = \{\mathbf{V}\}$ **then**
13   **if** type $= \ell$ **then yield** $\sum_{\mathbf{v} \setminus (\mathbf{X} \cup \mathbf{Y})} P(\mathbf{v})$ **else yield** $\sum_{\mathbf{v} \setminus (\mathbf{X} \cup \mathbf{Y})} \{P(\mathbf{v}) + 1 - \sum_{\mathbf{v} \setminus \mathbf{X}} P(\mathbf{v})\}$
14   **if** $\mathbf{C} \in \mathsf{cc}_{\mathcal{G}}$ **then yield** $\sum_{\mathbf{c} \setminus \mathbf{Y}} \prod_{V_i \in \mathbf{C}} P(v_i \mid \mathbf{v}^{(i-1)}_{\pi})$
15   **if** $\mathbf{C} \subsetneq \mathbf{C}' \in \mathsf{cc}_{\mathcal{G}}$ **then yield** PAID($\mathbf{y}, \mathbf{x} \cap \mathbf{C}', \prod_{V_i \in \mathbf{C}'} P(V_i \mid \mathbf{V}^{(i-1)}_{\pi} \cap \mathbf{C}', \mathbf{v}^{(i-1)}_{\pi} \setminus \mathbf{C}'), \mathcal{G}[\mathbf{C}']$, type)

$$\leq \sum_{b,c,y} y P^1_{\emptyset}(c \mid a, b) \{P^1_c(a, b, y) + 1 - P^1_c(a)\}. \tag{13}$$

Equipped with this bounding mechanism, we are now ready to construct general valid bounds for $\mu^*_{\mathbf{x}}$ algorithmically.

Algorithmic partial transportability. We propose the partial transportability algorithm PATR (Alg. 1), which returns bounding expressions for the target query. In the algorithm, $\mathbb{Z}$ is defined globally and does not change with the specific invocation of the algorithm. In contrast, variables $\mathbf{V}$ and selection variables $\mathbf{S}$ reflect the graph $\mathcal{G}$ and discrepancies $\mathbf{\Delta}$, respectively, relative to the arguments passed to the current execution of the procedure.

The algorithm enumerates valid expressions via the subroutine PAID, which outputs either a lower or an upper bound depending on type $\in \{\ell, u\}$. The design of PATR follows that of GTRU [Lee et al., 2020], with two key differences.

First, whereas GTRU returns a single expression for the target query, PATR yields multiple valid expressions. Second, when the target query is non-transportable, PAID instead returns a valid lower or upper bound (rather than GTRU's failure).

In particular, Line 5 decomposes the query into sub-queries such that $\mathbf{Y}$ in each sub-query forms a c-component. Line 6

Algorithm 2: CAUSALBOUND

Input: y; $\mathbf{x}$; $\mathcal{G}$; $\mathbf{\Delta}$; $\mathbb{Z}$; $\mathcal{P}_{\mathbb{Z}}^{\Pi}$

1 Initialize a causal bound $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$ as $[0, 1]$.
2 **for** $P_{\mathbf{x}}^{\ell}(y) :=$ PATR$(\{y\}, \mathbf{x}, \mathcal{G}, \mathbf{\Delta}, \ell)$ **do**
3 $\quad \ell_{\mathbf{x}} = \max\{\ell_{\mathbf{x}}, \sum_y y \cdot P_{\mathbf{x}}^{\ell}(y)\}$
4 **for** $P_{\mathbf{x}}^{u}(y) :=$ PATR$(\{y\}, \mathbf{x}, \mathcal{G}, \mathbf{\Delta}, u)$ **do**
5 $\quad u_{\mathbf{x}} = \min\{u_{\mathbf{x}}, \sum_y y \cdot \min\{1, P_{\mathbf{x}}^{u}(y)\}\}$
6 **return** $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$

then checks whether there exists a source distribution $P_{\mathbf{z}}^{i}$ that can be used to identify the query. If so, PATR forwards the query to PAID with slight modifications to both the query and the graph. The condition in Prop. 1 is enforced within PAID in Line 12; that is, if the algorithm reaches this line, the current target:

$$\begin{aligned} P_{\mathbf{x}}^{*}(\mathbf{y}) &= \textstyle\sum_{\mathbf{v} \setminus (\mathbf{X} \cup \mathbf{Y})} P_{\mathbf{x}}^{*}(\mathbf{v} \setminus \mathbf{X}) \\ &= \textstyle\sum_{\mathbf{v} \setminus (\mathbf{X} \cup \mathbf{Y})} P_{\mathbf{x}}^{i}(\mathbf{v} \setminus \mathbf{X}) = \sum_{\mathbf{v} \setminus (\mathbf{X} \cup \mathbf{Y})} Q^{i}[\mathbf{V} \setminus \mathbf{X}] \end{aligned}$$

is non-identifiable from π^{i}. Since $\mathrm{cc}_{\mathcal{G}} = \{\mathbf{V}\}$, we can use $Q^{i}[\mathbf{V}]$. Hence, Prop. 1 yields the corresponding bound expression for $P_{\mathbf{x}}^{*}(\mathbf{y})$ in terms of $Q^{i}[\mathbf{V}]$.

PATR runs in $\mathcal{O}(zn^4)$, where z is the number of experiments and n is the number of vertices. PAID can be invoked at most z times per PATR call, and each PAID call takes $\mathcal{O}(n^3)$ time. Due to the factorization step in Line 5, PATR may be called up to $\mathcal{O}(n)$ times, resulting in $\mathcal{O}(zn)$ PAID calls overall.

Theorem 1. PATR *returns valid closed-form expressions for lower/upper bounds.*

Remark 1. PATR *is a standalone procedure for deriving valid bounding expressions beyond bandit settings.*

Causal bounds. Using the lower and upper bounding expressions derived by PATR, CAUSALBOUND (Alg. 2) computes their numerical values.

The algorithm initializes causal bounds to $[0, 1]$ so that subsequent updates refine them (Line 1). Given $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$ and the corresponding realized source distributions $\mathcal{P}_{\mathbb{Z}}^{\Pi} = \{P_{\mathbf{z}}^{i} \mid \mathbf{z} \in \Omega_{\mathbf{Z}}, \mathbf{Z} \in \mathbb{Z}^{i} \in \mathbb{Z}\}$, the causal bound $\ell_{\mathbf{x}}/u_{\mathbf{x}}$ is obtained as the maximum/minimum over the realized lower/upper bound expressions returned by PATR.

Since the upper bounds of $P_{\mathbf{x}}^{*}(y)$ may exceed one due to sum-product operations over non-transportable factors, we truncate the resulting value at one.

Corollary 1. *Let* $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$ *be a bound from* CAUSALBOUND. *Then* $\mu_{\mathbf{x}}^{*}$ *is partially transportable with respect to* $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$ *with the bound given by* $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$.

Thus far, we have focused on bounding expected rewards using source priors. These bounds ignore structural constraints present in the target environment. In the following section, we build upon these causal bounds by exploiting dominance relations among actions under distinct constraints, ultimately deriving a tighter bound for the optimal action.

5 DOMINANCE BOUNDS

We say that an action space *dominates* another if its maximum achievable expected reward is at least as large as that of the other action space.

Lower dominance bound. It is immediate from the definition of POMIS that all POMISs dominate any non-POMIS under the same constraint. That is,

"*POMISs dominate non-POMISs*".

Therefore, under the same constraint $\mathbf{N}^{*}$, we define a lower dominance bound $\ell^{\star}$ as follows:

$$\ell^{\star} \triangleq \max_{\mathbf{w} \in \Omega_{\mathbf{W}}, \mathbf{W} \in \mathbb{M}_{\mathcal{G},Y}^{\mathbf{N}^{*}}} \ell_{\mathbf{w}}. \tag{14}$$

Indeed, we have:

$$\ell^{\star} = \ell_{\mathbf{w}^{\dagger}} \leq \mu_{\mathbf{w}^{\dagger}}^{*} \leq \mu_{\mathbf{x}^{\star}}^{*}. \tag{15}$$

where $\mathbf{w}^{\dagger} = \arg\max_{\mathbf{w} \in \Omega_{\mathbf{W}}, \mathbf{W} \in \mathbb{M}_{\mathcal{G},Y}^{\mathbf{N}^{*}}} \ell_{\mathbf{w}}$. Thus, the lower dominance bound $\ell^{\star}$ provides a valid lower bound.

Upper dominance bound. Beyond the trivial case, we now turn our attention to more nuanced dominance relations that arise between constrained and unconstrained POMISs.

Proposition 2. *A POMIS with respect to* $\langle \mathcal{G}, Y, \mathbf{N} \rangle$ *is minimal with respect to* $\langle \mathcal{G}, Y \rangle$.

The proposition states that a constrained POMIS is an MIS in the unconstrained setting, but it may not be a POMIS without constraints. This observation leads to the following statement.

"*Unconstrained POMISs dominate constrained POMISs*".

To see this, consider the collective selection diagram in Fig. 2b. Suppose the constraint $\mathbf{N}^{*} = \{W\}$. The POMISs are then given by $\mathbb{P}_{\mathcal{G},Y}^{\mathbf{N}^{*}} = \{\{X_1\}, \{X_1, X_2\}\}$, which implies that the optimal action must be consistent with $do(x_1^{*})$ or $do(\{x_1, x_2\}^{*})$ and not with $do(\emptyset)$ or $do(x_2^{*})$.

For concreteness, consider an unconstrained POMIS $\mathbf{R} = \{W, X_2\}$. The expected reward under $do(x_1^{*})$ can be decomposed as:

$$\mu_{x_1^{*}}^{*} = \sum_{\mathbf{r}} \mathbb{E}_{P_{x_1^{*}}^{*}}[Y \mid \mathbf{r}] P_{x_1^{*}}^{*}(\mathbf{r}) \overset{\text{(a)}}{=} \sum_{\mathbf{r}} \mathbb{E}_{P_{x_1^{*},\mathbf{r}}^{*}}[Y] P_{x_1^{*}}^{*}(\mathbf{r})$$

Algorithm 3: Upper dominance bound (UDB)

Input: $\mathcal{G}$; Y; $\mathbf{N}$; $\mathbb{L}$, $\mathbb{U}$: collections of lower/upper causal bounds

```
1  function UDB(G, Y, N, L, U)
2      Initialize the upper dominance bound u* = 0.
3      Let H = G⟨V \ N⟩.
4      Let A_{H,Y} = {r ∈ Ω_R | R ∈ P_{H,Y}}.
5      Set (u*, r*) = (max_{r∈A_{H,Y}} u_r, arg max_{r∈A_{H,Y}} u_r).
6      if u* < 1 and u_{r*} = ℓ_{r*} (i.e., transportable) then
         return u*
7      if N = ∅ then return u*
8      for all N' ⊂ N such that |N'| = |N| − 1 do
9          u* = min{u*, UDB(G, Y, N', L, U)}
10     return u*
```

$$\stackrel{\text{(b)}}{=} \sum_{\mathbf{r}} \mathbb{E}_{P^*_{\mathbf{r}}}[Y] P^*_{x_1^*}(\mathbf{r}) \leq \mu^*_{\mathbf{r}^*} \tag{16}$$

$$\leq \mu^*_{\mathbf{r}^\star} \tag{17}$$

where (a) follows from Rule 2 and (b) follows from Rule 3 of do-calculus. A similar argument applied to $do(\{x_1, x_2\}^*)$ yields:

$$\mu^*_{\{x_1,x_2\}^*} \leq \mu^*_{\mathbf{r}^\star}. \tag{18}$$

Therefore, we conclude that

$$\mu^*_{\mathbf{x}^\star} \leq \mu^*_{\mathbf{r}^\star}. \tag{19}$$

Intuitively, under the dominance relation, the POMISs under the constraint $\mathbf{N}^*$—namely $do(x_1)$ and $do(x_1, x_2)$—serve as the *best alternative* plans. This dominance observation leads to the following theorem.

Theorem 2 (Upper dominance)**.** *Let $\mathbf{r}^\star$ be the optimal action with respect to $\langle \mathcal{G}, Y, \mathbf{N} \rangle$ where $\mathbf{N}$ is a proper subset of $\mathbf{N}^*$. Then $\mu^*_{\mathbf{x}^\star}$ is upper bounded by $\mu^*_{\mathbf{r}^\star}$.*

That is, POMISs are dominated by POMISs under a *weaker* constraint $\mathbf{N} \subsetneq \mathbf{N}^*$. This can be viewed as a generalization of the standard "constrained vs. unconstrained" dominance relation to a relation between "constrained vs. relatively weaker constrained" one. Therefore, for each weaker constraint $\mathbf{N} \subsetneq \mathbf{N}^*$, any maximizer $\mathbf{r}^\star = \arg\max_{\mathbf{r} \in \mathcal{A}^{\mathbf{N}}_{\mathcal{G},Y}} \mu^*_{\mathbf{r}}$ provides a valid upper bound on $\mu^*_{\mathbf{x}^\star}$.

Furthermore, although $\mu^*_{\mathbf{r}^\star}$ may not be transportable, one can instead leverage the corresponding upper causal bound $u_{\mathbf{r}^\star}$. Consequently, an upper bound on the target optimal value can be obtained from either $\mu^*_{\mathbf{r}^\star}$ or $u_{\mathbf{r}^\star}$, by taking the minimum over *all* constraints strictly weaker than $\mathbf{N}^*$.

Hierarchical dominance relationships. One might be concerned that evaluating POMISs for *all* weaker constraints is computationally expensive. However, it is worth noting that any bound associated with a strictly weaker constraint is *looser*. We illustrate this in Fig. 4, where $\mathbf{N}^* = \{A, C\}$ for the setting in Fig. 2c.

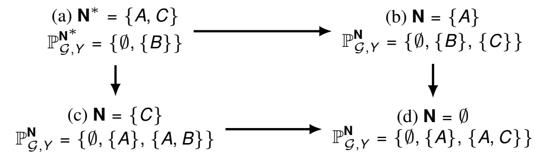


Figure 4: Hierarchical relationships among POMISs under different constraints. Edges indicate the direction of dominance, i.e., $\mathbb{P}_1 \to \mathbb{P}_2$ means that $\mathbb{P}_2$ dominates $\mathbb{P}_1$.

In this setting, each optimal action under weaker constraints, (b–d), dominates (a); furthermore, (d) dominates both (b) and (c). This implies that (d) is not tighter than either (b) or (c); thus if (b) or (c) is known, estimating (d) is unnecessary to upper bound (a).

In other words, the following hierarchical dominance relations hold:

$$\mu^*_{\mathbf{x}^\star} \leq \mu^*_{\mathbf{r}_1^\star} \leq \cdots \leq \mu^*_{\mathbf{r}_k^\star} \tag{20}$$

where $\mathbf{x}^\star \in \mathcal{A}^{\mathbf{N}^*}_{\mathcal{G},Y}$, and $\mathbf{r}_k^\star \in \mathcal{A}^{\mathbf{N}_k}_{\mathcal{G},Y}$ with

$$\mathbf{N}^* \supsetneq \mathbf{N}_1 \supsetneq \cdots \supsetneq \mathbf{N}_{k-1} \supsetneq \mathbf{N}_k = \emptyset. \tag{21}$$

Equipped with this knowledge, we can compute an upper dominance bound $u^\star$ efficiently. The corresponding algorithm (Alg. 3) hierarchically traverses the space of constrained POMISs and stops *early* once a transportable $\mu^*_{\mathbf{r}^\star}$ is found, thereby avoiding unnecessary computation.

In Line 2, the algorithm initializes $u^\star = 0$ so that subsequent updates refine it. In Lines 3–5, the algorithm attempts to compute $u_{\mathbf{r}^\star}$ over $\mathbf{r} \in \mathcal{A}_{\mathcal{H},Y}$ where $\mathcal{A}_{\mathcal{H},Y} = \mathcal{A}^{\mathbf{N}}_{\mathcal{G},Y}$ holds [Lee and Bareinboim, 2019].

If the algorithm reaches Line 8, this implies that either:

1. there exists at least one action whose upper causal bound is not partially transportable (i.e., $u^\star = 1$); or
2. the current $u^\star$ corresponds to the upper causal bound of a *non*-transportable action (i.e., $u^\star = u_{\mathbf{r}^\star} \neq \mu^*_{\mathbf{r}^\star}$).

Note that $\mu^*_{\mathbf{r}^\star_{k-1}} \leq \mu^*_{\mathbf{r}^\star_k}$ does not imply $u_{\mathbf{r}^\star_{k-1}} \leq u_{\mathbf{r}^\star_k}$. Therefore, in either case, the hierarchy must be further traversed. In such cases, the algorithm recursively explores one-level weaker constraints $\mathbf{N}'$ and returns the tightest bound among the results of recursive calls.

Conversely, if $u^\star < 1$ and the upper dominance bound corresponds to the expected reward of a transportable action (**early stop**; Line 6), no further recursive calls are needed, and the algorithm terminates with $u^\star$.

Synergy of bounds. Causal bounds $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$ provide pointwise constraints that bound the reward of each action using

Algorithm 4: TRUCB

1 Initialize the action space $\mathcal{A}^* = \mathcal{A}^{\mathbf{N}^*}_{\mathcal{G},Y}$.
2 Set causal bounds $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$ using CAUSALBOUND (Alg. 2).
3 Compute $u^\star$ using UDB (Alg. 3).
4 Compute $\ell^\star = \max_{\mathbf{w}\in\Omega_{\mathbf{W}},\, \mathbf{W}\in\mathbb{M}^{\mathbf{N}^*}_{\mathcal{G},Y}} \ell_{\mathbf{w}}$ (Eq. (14)).
5 Remove all actions $\mathbf{x} \in \mathcal{A}^*$ such that $u_{\mathbf{x}} < \ell^\star$.
6 Pull all actions once in $\mathcal{A}^*$.
7 **for** each round $t \le T$ **do**
8 Choose an arm $\mathbf{x}_t = \operatorname{argmax}_{\mathbf{x}\in\mathcal{A}^*} \bar{U}_{\mathbf{x}}(t)$ where $\bar{U}_{\mathbf{x}}(t) = \min\{U_{\mathbf{x}}(t), \min\{u_{\mathbf{x}}, u^\star\}\}$.
9 Intervene $do(\mathbf{x}_t)$ for round t and receive $Y_t \sim P^*_{\mathbf{x}_t}$.

source data, whereas the upper dominance bound $u^\star$ exploits the hierarchical transportability of non-manipulable variables to enforce relationships *across* actions. These two components are complementary in practice. In some environments, the offline causal bounds are tight enough to directly prune suboptimal arms, i.e., when $u_{\mathbf{x}} < \ell^\star$. In other environments, where the causal bounds overlap, the dominance bound can further tighten the feasible reward range and guide exploration among the remaining actions. See Appendix D for detailed worked examples.

6 UCB WITH TRANSFERRED BOUNDS

In this section, we incorporate the obtained bounds into a MAB problem. To this end, we adopt the well-known *clipping* approach [Zhang and Bareinboim, 2017, Gong et al., 2023] as our index strategy within the UCB [Auer et al., 2002, Audibert et al., 2009].

TRUCB (Alg. 4) begins by initializing the target action space $\mathcal{A}^*$ with POMISs $\mathcal{A}^{\mathbf{N}^*}_{\mathcal{G},Y}$. Next, in Lines 2–4, the algorithm computes causal bounds and a dominance bound. In Line 5, we can safely eliminate actions $\mathbf{x}$ whose upper causal bound is lower than the lower dominance bound. In the final online part (Lines 6–9), following Gong et al. [2023], the algorithm enters the interaction phase with the final action space $\mathcal{A}^*$.

We denote the empirical reward estimate at round t for each arm in the target environment π^* as:

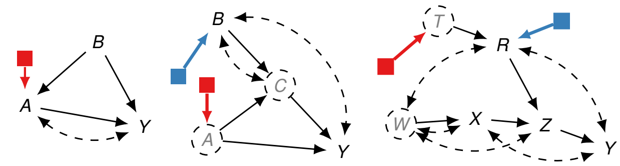
$$\hat{\mu}^*_{\mathbf{x},t} = \frac{1}{N_t(\mathbf{x})} \sum_{t'=1}^{t} Y_{t'} \mathbb{I}\{\mathbf{X}_{t'} = \mathbf{x}\}. \tag{22}$$

The UCB index is defined as:

$$U_{\mathbf{x}}(t) = \hat{\mu}^*_{\mathbf{x},t} + \sqrt{\frac{2\sigma^2_{\mathbf{x}} \log(2t/\delta)}{N_t(\mathbf{x})}} \tag{23}$$

where δ denotes a confidence parameter, and $\sigma^2_{\mathbf{x}}$ denotes the maximal reward variance for $\mathbf{x} \in \mathcal{A}^*$, defined as:

$$\sigma^2_{\mathbf{x}} = \max\{\mu(1-\mu) \mid \mu \in [\ell^\star, \min\{u_{\mathbf{x}}, u^\star\}]\}. \tag{24}$$



(a) Task 1 (b) Task 2 (c) Task 3

Figure 5: Collective selection diagrams for main experiments. The constrained variables $\mathbf{N}^*$ are shown with dashed circles.

At each round t, TRUCB computes the clipped $\bar{U}_{\mathbf{x}}(t)$ for every arm by combining the empirical rewards collected up to round t with $\min\{u_{\mathbf{x}}, u^\star\}$. The learner then selects $\mathbf{x}_t$ with the highest index and receives the corresponding reward Y_t.

Taking $\delta = \frac{1}{T}$, the expected number of pulls of any suboptimal arm $\mathbf{x} \neq \mathbf{x}^\star$ satisfies the following bound, up to universal constant factors [Gong et al., 2023]:

$$\mathbb{E}N_T(\mathbf{x}) \le \begin{cases} 0 & \text{if } u_{\mathbf{x}} < \ell^\star \ (\mathbf{x} \notin \mathcal{A}^*); \\ 1 & \text{if } \mathbf{x} \in \mathcal{A}^* \text{ and } u_{\mathbf{x}} < \mu^*_{\mathbf{x}^\star}; \\ 1+8\sigma^2_{\mathbf{x}} \frac{\log(T)}{\Delta^2_{\mathbf{x}}} & \text{if } \mathbf{x} \in \mathcal{A}^* \text{ and } u_{\mathbf{x}} \ge \mu^*_{\mathbf{x}^\star} \end{cases} \tag{25}$$

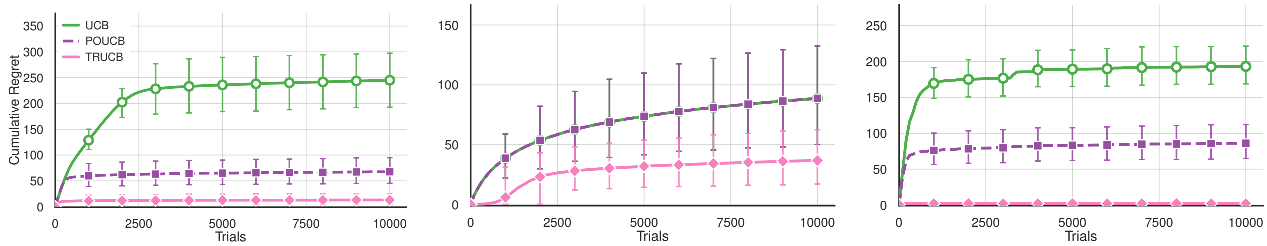
Consequently, since $\mathbb{E}R_T = \sum_{\mathbf{x}} \Delta_{\mathbf{x}} \mathbb{E}N_T(\mathbf{x})$, the expected regret admit the bound $\mathcal{O}\Big(\sum_{\substack{\mathbf{x}\in\mathcal{A}^*:\\ u_{\mathbf{x}} \ge \mu^*_{\mathbf{x}^\star},\, \Delta_{\mathbf{x}}>0}} \frac{\log T}{\Delta_{\mathbf{x}}}\Big)$[6].

Numerical experiments. We compare TRUCB with two baselines: standard UCB over all arm combinations $\mathcal{A}^*_{\text{UCB}}$ and UCB restricted to POMISs $\mathcal{A}^*_{\text{POUCB}}$ $(= \mathcal{A}^{\mathbf{N}^*}_{\mathcal{G},Y})$. The number of trials is set to 10k, which is sufficient to observe the performance differences, as shown in Fig. 6.

Task 1. We start with the simple problem represented in Fig. 5a, where $\mathbf{N}^* = \emptyset$. The learner has access to $\mathbb{Z}^1 = \{\{B\}\}$ from an environment with $\Delta^1 = \{A\}$. The action space without prior information corresponds to $\mathbb{P}^{\mathbf{N}^*}_{\mathcal{G},Y} = \{\{B\}, \{A, B\}\}$. In this setting, two actions $do(A{=}0, B{=}1)$ and $do(A{=}1, B{=}1)$ are eliminated from $\mathcal{A}^*$ by $u_{\mathbf{x}} < \ell^\star$. We observe that the mean cumulative regret at the final trial is 244.94, <u>67.54</u> and **12.91** for UCB, POUCB and TRUCB, respectively[7].

Task 2. We consider the setting in Fig. 5b where $\Delta^1 = \{A\}$ and $\Delta^2 = \{B\}$ with priors $\mathbb{Z}^1 = \{\emptyset\}$ and $\mathbb{Z}^2 = \{\{C\}\}$ and $\mathbf{N}^* = \{A, C\}$. The action space without prior information corresponds to $\mathbb{P}^{\mathbf{N}^*}_{\mathcal{G},Y} = \{\emptyset, \{B\}\}$. Here, $2^{\mathbf{V}\setminus\{Y\}\setminus\mathbf{N}^*} = \{\emptyset, \{B\}\}$, implying $\mathcal{A}^* = \mathcal{A}^*_{\text{UCB}}$. Furthermore, there is no action $\mathbf{x} \in \mathcal{A}^{\mathbf{N}^*}_{\mathcal{G},Y}$ such that $u_{\mathbf{x}} < \ell^\star$, implying that the size of the action space remains unchanged (i.e.,

[6] Since $\mu^*_{\mathbf{x}^\star} \le u^\star$ always holds, $u^\star$ does not appear.
[7] Best results are in **bold** and the runners-up are <u>underlined</u>.



(a) Task 1 (b) Task 2 (c) Task 3

Figure 6: Cumulative regrets of **TRUCB** compared with standard UCB over all combinations of arms (**UCB**) and over POMISs (**POUCB**). Each simulation is repeated 1,000 times to ensure consistency. The thick line with markers indicates the mean cumulative regret averaged over repetitions and the vertical lines indicate the 2.5th and 97.5th percentiles of the empirical cumulative regrets.

$\mathcal{A}^* = \mathcal{A}^*_{\text{UCB}} = \mathcal{A}^*_{\text{POUCB}}$). A POMIS action $do(A{=}1, C{=}1)$ from weaker constraint $\mathbf{N} = \emptyset$ provides $u^\star$, which in turn yields tighter upper bounds for $\mathbf{x} \in \mathcal{A}^*$. Even though no actions are eliminated, the mean cumulative regret at T is **36.84** for TRUCB and 88.69 for both UCB and POUCB.

Task 3. We consider a more involved scenario (Fig. 5c) to validate our results. Let $\Delta^1 = \{T\}$ and $\Delta^2 = \{R\}$, with priors $\mathbb{Z}^1 = \{\emptyset, \{Z\}\}$ and $\mathbb{Z}^2 = \{\{Z\}\}$, and $\mathbf{N}^* = \{T, W\}$. The action space without prior information corresponds to $\mathbb{P}^{\mathbf{N}^*}_{\mathcal{G},Y} = \{\emptyset, \{R\}, \{X\}, \{Z\}\}$. In this setting, three actions $do(\emptyset)$, $do(X{=}0)$ and $do(Z{=}0)$ are eliminated from $\mathcal{A}^*$ by $u_{\mathbf{x}} < \ell^\star$. The mean cumulative regret at T is 193.21, 86.13 and **2.02** for UCB, POUCB and TRUCB, respectively.

Additional details on the illustrative example and the experimental setup are provided in Appendix D. These results demonstrate performance improvements when transferred causal knowledge from sources is taken into account.

7 LIMITATIONS

Sufficient prior data from source environments. Our method relies on source data to construct dominance and causal bounds that guide exploration. For improved sample efficiency, reward data collected under an action and its minimal representative may be treated as equivalent for estimating the expected reward, since both interventions induce the same interventional distribution over the reward. This perspective parallels the relationship between a POMIS and its corresponding POISs.

However, even when leveraging such sample-efficient aggregation of source data, the resulting bounds may be inaccurate if the prior data are insufficient or biased. In particular, inaccurate bounds may fail to contain the true expected reward of certain actions. To address this issue, Gong et al. [2023] propose a method for incorporating "noisy" bounds into clipping UCB, assuming that the noise level is known. Extending this line of work to more general settings with potentially unknown noise levels remains an important direction for future research.

Tightness of causal bounds. We do not guarantee the *tightness* of our causal bounds; that is, we cannot ensure that there always exists an SCM under which the expected reward exactly equals either $\ell_{\mathbf{x}}$ or $u_{\mathbf{x}}$. Characterizing conditions under which lower and upper bounds are attainable by some SCM would enhance the interpretability and reliability of our framework. Investigating tighter query-specific bounds and formally establishing their tightness from the available information remain important future work.

Misspecification of collective selection diagrams. While collective selection diagrams provide a principled and interpretable framework for analyzing environment shifts and transferring information across environments, they restrict the analysis to settings in which the common causal diagram and invariant variables are explicitly specified. Nevertheless, some degree of misspecification can be tolerated without invalidating the performance guarantees, particularly when the assumed collective selection diagram forms a *super-model* of the true environment [Bellot et al., 2023].

Selection nodes indicate *potential* discrepancies between environments. Thus, when uncertain whether a mechanism differs across environments, a researcher can conservatively assume such a discrepancy. This preserves the validity of transportability guarantees, though it may increase the number of POMISs, reduce informativeness, and loosen bounds. By contrast, failing to model a true discrepancy can lead to incorrect inference.

8 CONCLUSION

We investigated a structured causal bandit strategy that can utilize prior data from related heterogeneous environments. Since source environments may differ, some knowledge of the underlying structure and potential discrepancies is necessary to enable consistent extrapolation. To address this, we proposed a strategy that exploits transportable causal knowledge by incorporating bounds derived from dominance relations and causal structure.

Acknowledgements

We thank the anonymous reviewers for constructive comments to improve the manuscript. This work was supported by the IITP (RS-2022-II220953/50%) and NRF (RS-2023-00211904/50%) grants funded by the Korean government.

References

Virginia Aglietti, Xiaoyu Lu, Andrei Paleyes, and Javier González. Causal bayesian optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 3155–3164. PMLR, 2020.

Jean-Yves Audibert, Rémi Munos, and Csaba Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902, 2009.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256, 2002.

Alexander Balke and Judea Pearl. Counterfactuals and policy analysis in structural models. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 11–18, 1995.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American statistical Association*, 92(439):1171–1176, 1997.

Elias Bareinboim and Judea Pearl. Causal inference by surrogate experiments: z-identifiability. In *Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence*, pages 113–120. AUAI Press, 2012a.

Elias Bareinboim and Judea Pearl. Transportability of causal effects: Completeness results. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pages 698–704, 2012b.

Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

Elias Bareinboim, Sanghack Lee, Vasant Honavar, and Judea Pearl. Transportability from multiple environments with limited experiments. *Advances in Neural Information Processing Systems*, 26, 2013.

Elias Bareinboim, Andrew Forney, and Judea Pearl. Bandits with unobserved confounders: A causal approach. In *Proceedings of the 28th Annual Conference on Neural Information Processing Systems*, pages 1342–1350, 2015.

Elias Bareinboim, Junzhe Zhang, and Sanghack Lee. An introduction to causal reinforcement learning. Technical Report R-65, Causal Artificial Intelligence Lab, Columbia University, Dec 2024. URL https://causalai.net/r65.pdf.

Alexis Bellot and Silvia Chiappa. Towards estimating bounds on the effect of policies under unobserved confounding. *Advances in Neural Information Processing Systems*, 37:104556–104594, 2024.

Alexis Bellot, Alan Malek, and Silvia Chiappa. Transportability for bandits with data from different environments. *Advances in Neural Information Processing Systems*, 36:44356–44381, 2023.

Shriya Bhatija, Paul-David Zuercher, Jakob Thumm, and Thomas Bohné. Multi-objective causal Bayesian optimization. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pages 4172–4195. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/bhatija25a.html.

Blair Bilodeau, Linbo Wang, and Daniel M. Roy. Adaptively exploiting d-separators with causal bandits. *Advances in Neural Information Processing Systems*, 35:20381–20392, 2022.

Ryan Carey, Sanghack Lee, and Robin J Evans. Toward a complete criterion for value of information in insoluble decision problems. *Transactions on Machine Learning Research*, 2024.

Juan Correa and Elias Bareinboim. General transportability of soft interventions: Completeness results. *Advances in Neural Information Processing Systems*, 33:10902–10912, 2020.

Juan D Correa and Elias Bareinboim. Counterfactual graphical models: Constraints and inference. In *Forty-second International Conference on Machine Learning*, 2025.

Juan D Correa, Sanghack Lee, and Elias Bareinboim. Counterfactual transportability: a formal approach. In *International Conference on Machine Learning*, pages 4370–4390. PMLR, 2022.

Mingwei Deng, Ville Kyrki, and Dominik Baumann. Transfer learning in latent contextual bandits with covariate shift through causal transportability. In *Proceedings of the Fourth Conference on Causal Learning and Reasoning (CLeaR)*, volume 275 of *Proceedings of Machine Learning Research*, pages 731–756. PMLR, 2025.

Wen-Bo Du, Tian Qin, Tian-Zuo Wang, and Zhi-Hua Zhou. Avoiding undesired future with minimal cost in non-stationary environments. *Advances in Neural Information Processing Systems*, 37:135741–135769, 2024.

Wen-Bo Du, Hao-Yi Lei, Lue Tao, Tian-Zuo Wang, and Zhi-Hua Zhou. Enabling optimal decisions in rehearsal learning under CARE condition. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=jh0Fss7LA0.

Muhammad Qasim Elahi, Mahsa Ghasemi, and Murat Kocaoglu. Partial structure discovery is sufficient for no-regret learning in causal bandits. *Advances in Neural Information Processing Systems*, 37:109066–109100, 2024.

Tom Everitt, Ryan Carey, Eric D Langlois, Pedro A Ortega, and Shane Legg. Agent incentives: A causal perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 11487–11495, 2021.

Shi Feng and Wei Chen. Combinatorial causal bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 7550–7558, 2023.

Ele Ferrannini and William C Cushman. Diabetes and hypertension: the bad companions. *The Lancet*, 380(9841): 601–610, 2012.

Andrew Forney, Judea Pearl, and Elias Bareinboim. Counterfactual data-fusion for online reinforcement learners. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1156–1164. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/forney17a.html.

Xueping Gong, Wei You, and Jiheng Zhang. Efficient transfer learning via causal bounds, 2023. URL https://arxiv.org/abs/2308.03572.

Yimin Huang and Marco Valtorta. Pearl's calculus of intervention is complete. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, pages 217–224. AUAI Press, 2006.

Yimin Huang and Marco Valtorta. On the completeness of an identifiability algorithm for semi-Markovian models. *Annals of Mathematics and Artificial Intelligence*, 54(4): 363–408, 2008.

Inwoo Hwang, Yunhyeok Kwak, Suhyung Choi, Byoung-Tak Zhang, and Sanghack Lee. Fine-grained causal dynamics learning with quantization for improving robustness in reinforcement learning. In *International Conference on Machine Learning*, pages 20842–20870. PMLR, 2024.

Kasra Jalaldoust and Elias Bareinboim. Transportable representations for domain generalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 12790–12800, 2024.

Kasra Jalaldoust, Alexis Bellot, and Elias Bareinboim. Partial transportability for domain generalization. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Shalmali Joshi, Junzhe Zhang, and Elias Bareinboim. Towards safe policy learning under partial identifiability: A causal approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 13004–13012, 2024.

Yonghan Jung and Alexis Bellot. Efficient policy evaluation across multiple different experimental datasets. *Advances in Neural Information Processing Systems*, 37:136361–136392, 2024.

Yonghan Jung, Min Woo Park, and Sanghack Lee. Complete graphical criterion for sequential covariate adjustment in causal inference. *Advances in Neural Information Processing Systems*, 37:19813–19838, 2024.

Yaroslav Kivva, Ehsan Mokhtarian, Jalal Etesami, and Negar Kiyavash. Revisiting the general identifiability problem. In *The 38th Conference on Uncertainty in Artificial Intelligence*, 2022.

Daniel Kumor, Junzhe Zhang, and Elias Bareinboim. Sequential causal imitation learning with unobserved confounders. *Advances in Neural Information Processing Systems*, 34:14669–14680, 2021.

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. *Advances in neural information processing systems*, 29, 2016.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? *Advances in Neural Information Processing Systems*, 31, 2018.

Sanghack Lee and Elias Bareinboim. Structural causal bandits with non-manipulable variables. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4164–4172, 2019.

Sanghack Lee and Elias Bareinboim. Characterizing optimal mixed policies: Where to intervene and what to observe. *Advances in Neural Information Processing Systems*, 33: 8565–8576, 2020a.

Sanghack Lee and Elias Bareinboim. Causal effect identifiability under partial-observability. In *Proceedings of the 37th International Conference on Machine Learning*, 2020b.

Sanghack Lee, Juan D Correa, and Elias Bareinboim. General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2019.

Sanghack Lee, Juan Correa, and Elias Bareinboim. General transportability–synthesizing observations and experiments from heterogeneous domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10210–10217, 2020.

Yangyi Lu, Amirhossein Meisami, Ambuj Tewari, and William Yan. Regret analysis of bandit problems with causal background knowledge. In *Conference on Uncertainty in Artificial Intelligence*, pages 141–150. PMLR, 2020.

Charles F Manski. Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323, 1990.

Min Woo Park and Sanghack Lee. Counterfactual structural causal bandits. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=gjvTNxVd2f.

Min Woo Park, Andy Arditi, Elias Bareinboim, and Sanghack Lee. Structural causal bandits under Markov equivalence. *Advances in Neural Information Processing Systems*, 38, 2025.

Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–710, 1995.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000. 2nd edition, 2009.

Judea Pearl and Elias Bareinboim. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 25, pages 247–254, 2011.

Judea Pearl and James Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 444–453, 1995.

Tian Qin, Tian-Zuo Wang, and Zhi-Hua Zhou. Rehearsal learning for avoiding undesired future. *Advances in Neural Information Processing Systems*, 36:80517–80542, 2023.

Tian Qin, Tian-Zuo Wang, and Zhi-Hua Zhou. Gradient-based nonlinear rehearsal learning with multivariate alterations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 26859–26867, 2025.

Herbert Robbins. Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.*, 58(5):527–535, 1952.

James M Robins. The analysis of randomized and non-randomized aids treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS*, pages 113–159, 1989.

Kangrui Ruan, Junzhe Zhang, Xuan Di, and Elias Bareinboim. Causal imitation for Markov decision processes: A partial identification approach. *Advances in Neural Information Processing Systems*, 37:87592–87620, 2024.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of the 21st AAAI Conference on Artificial Intelligence*, page 1219. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

Lue Tao, Tian-Zuo Wang, Yuan Jiang, and Zhi-Hua Zhou. Avoiding undesired future with sequential decisions. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 6245–6253, 09 2025. doi: 10.24963/ijcai.2025/695.

Lue Tao, Tian-Zuo Wang, Yuan Jiang, and Zhi-Hua Zhou. On measuring influence in avoiding undesired future. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=VHdF91MvJq.

Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Proceedings of the 18th National Conference on Artificial Intelligence*, pages 567–573, 2002.

Jin Tian and Judea Pearl. On the identification of causal effects. Technical Report R-290-L, Cognitive Systems Laboratory, UCLA, 2003.

Burak Varici, Karthikeyan Shanmugam, Prasanna Sattigeri, and Ali Tajer. Causal bandits for linear structural equation models. *Journal of Machine Learning Research*, 24(297): 1–59, 2023.

Lai Wei, Muhammad Qasim Elahi, Mahsa Ghasemi, and Murat Kocaoglu. Approximate allocation matching for structural causal bandits with unobserved confounders. *Advances in Neural Information Processing Systems*, 36: 68810–68832, 2023.

Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection: Expressiveness, learnability, and inference. *Advances in Neural Information Processing Systems*, 34, 2021.

Junzhe Zhang and Elias Bareinboim. Transfer learning in multi-armed bandit: a causal approach. In *Proceedings of the 16th Conference on Autonomous Agents and Multi-Agent Systems*, pages 1778–1780, 2017.

Junzhe Zhang and Elias Bareinboim. Near-optimal reinforcement learning in dynamic treatment regimes. *Advances in Neural Information Processing Systems*, 32, 2019.

Junzhe Zhang and Elias Bareinboim. Designing optimal dynamic treatment regimes: A causal reinforcement learning approach. In *International conference on machine learning*, pages 11012–11022. PMLR, 2020.

Junzhe Zhang and Elias Bareinboim. Bounding causal effects on continuous outcome. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12207–12215, 2021.

Junzhe Zhang and Elias Bareinboim. Online reinforcement learning for mixed policy scopes. *Advances in Neural Information Processing Systems*, 35:3191–3202, 2022.

Junzhe Zhang and Elias Bareinboim. Causal eligibility traces for confounding robust off-policy evaluation. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 286 of *Proceedings of Machine Learning Research*, pages 4933–4942. PMLR, 2025.

Junzhe Zhang, Daniel Kumor, and Elias Bareinboim. Causal imitation learning with unobserved confounders. *Advances in Neural Information Processing Systems*, 33: 12263–12274, 2020.

Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In *International conference on machine learning*, pages 26548–26558. PMLR, 2022.

On Transportability for Structural Causal Bandits (Supplementary Material)

Min Woo Park[1] **Sanghack Lee***[1]

[1]Graduate School of Data Science, Seoul National University, Seoul, Republic of Korea

CONTENTS

1 Introduction **1**

2 Preliminaries **2**

3 Structural Causal Bandits with Transportability **2**

4 Partial Transportability and Causal Bounds **4**

4.1 Partial Transportability 4

5 Dominance Bounds **6**

6 UCB with Transferred Bounds **8**

7 Limitations **9**

8 Conclusion **9**

A Related Works **15**

B Background **16**

C Proofs **17**

C.1 Proof of Proposition 1 17

C.2 Proof of Theorem 1 18

C.3 Proof of Corollary 1 19

C.4 Proof of Proposition 2 19

C.5 Proof of Theorem 2 20

D Details on Experimental Settings **20**

E Discussions **23**

A RELATED WORKS

Identification. Causal effect identification, also known as the *identification problem* [Pearl, 1995], concerns whether the causal effect of an intervention on a set of variables can be uniquely computed from the observational distribution over observed variables and a given causal diagram. Foundational works [Tian and Pearl, 2002, Shpitser and Pearl, 2006, Huang and Valtorta, 2006, 2008] culminated in a complete graphical and algorithmic characterization of the problem. Beyond purely observational sources, there has been growing interest in generalizing the identification problem to settings where both observational and experimental data are available. Bareinboim and Pearl [2012a] studied the conditions under which the causal effect is uniquely computable, given a causal diagram and a collection of observational and experimental distributions over *all* subsets of a given set. Lee et al. [2019] and Kivva et al. [2022] investigated the identification problem over arbitrary collections of distributions and established necessary and sufficient graphical conditions for generalized identification. Lee and Bareinboim [2020b] extended this problem to settings with partially observable source variables.

Partial identification. Given a causal diagram, one can express a causal effect in terms of the observational distribution using standard identification algorithms. However, challenges related to non-identifiability may arise, and the target effect may not be uniquely computable from observational data. The framework of *partial identification* [Balke and Pearl, 1995, 1997] addresses this issue by constraining the parameter space of causal effects within an interval. Zhang et al. [2022] proposed a polynomial programming approach to solve partial identification problems for arbitrary causal diagrams. Partial identification has also been applied in the causal decision-making literature to estimate dynamic treatment regimes [Zhang and Bareinboim, 2019, 2020, 2022], reinforcement learning [Zhang and Bareinboim, 2025, Bareinboim et al., 2024], bandit algorithms [Zhang and Bareinboim, 2017, 2021, Joshi et al., 2024], and other domains [Jalaldoust et al., 2024, Bellot and Chiappa, 2024, Ruan et al., 2024].

Transportability. In the causal inference literature, the problem of identifying causal effects under potential environment discrepancies has been extensively studied through the theory of *transportability* [Pearl and Bareinboim, 2011, Bareinboim and Pearl, 2012b, Bareinboim et al., 2013, Bareinboim and Pearl, 2016]. These works focus on determining whether a causal effect can be identified across environments, and which aspects of causal knowledge can be transferred. Lee et al. [2020] investigated transportability under arbitrary combinations of experiments conducted in both the source and target environments. Correa and Bareinboim [2020] extended this framework to handle soft interventions, while Correa et al. [2022] further generalized it to the setting of counterfactual effects. More recently, Jalaldoust et al. [2024] proposed a parameterization model approach [Zhang et al., 2022, Xia et al., 2021] to bound non-transportable causal effects. Zhang and Bareinboim [2017] explored the integration of prior knowledge into the multi-armed bandit (MAB) framework under restrictive graph structures such as the Bow and IV settings. Building upon this, Bellot et al. [2023] extended this idea to arbitrary causal diagrams, and Deng et al. [2025] further generalized it to settings where contextual information is available, although their focus was limited to single-node interventions.

Structural causal bandits. Lee and Bareinboim [2018] formalized the *structural causal bandit* framework, in which a bandit instance is structured by an SCM, and each action corresponds to an intervention on a subset of variables. The authors proposed a sound and complete graphical characterization for identifying actions that could be part of an optimal strategy, enabling an agent to avoid unnecessary exploration *a priori*, without any actual interaction. Lee and Bareinboim [2019] extended the framework to settings involving non-manipulable variables. Lee and Bareinboim [2020a] established the framework under stochastic policies and demonstrated the informativeness of such policies. Everitt et al. [2021] and Carey et al. [2024] further investigated the completeness of the graphical characterization of optimal policy spaces, although the general completeness remains an open problem. Wei et al. [2023] proposed a parameterization-based approach to incorporate shared information among possibly-optimal actions. Elahi et al. [2024] extended the SCB framework to settings where no causal graph is assumed to be accessible, requiring online discovery. Park et al. [2025] investigated SCB in settings where the available information is not a full partial ancestral graph representing the Markov equivalence class of the true causal diagram. Bhatija et al. [2025] extended the framework to a multi-outcome variant incorporating Pareto optimality. Recently, Park and Lee [2026] generalized it from interventional actions to realizable counterfactual actions.

B BACKGROUND

D-separation. In a causal diagram $\mathcal{G}$, a path p between vertices X and Y is a *d-connecting* (*active*) path relative to a set $\mathbf{Z}$ if:

1. every non-collider[1] on p is not a member of $\mathbf{Z}$; and
2. every collider on p is an ancestor of some member of $\mathbf{Z}$.

Two variables X and Y are said to be *d-separated* by $\mathbf{Z}$ if there is no d-connecting path between X and Y relative to $\mathbf{Z}$. Furthermore, two disjoint sets $\mathbf{X}$ and $\mathbf{Y}$ are said to be d-separated by $\mathbf{Z}$ if every variable in $\mathbf{X}$ is d-separated from every variable in $\mathbf{Y}$ by $\mathbf{Z}$; this is denoted as $(\mathbf{X} \perp\!\!\!\perp_d \mathbf{Y} \mid \mathbf{Z})_{\mathcal{G}}$.

Do-calculus. Pearl [1995] devised *do-calculus*, which acts as a bridge between observational and interventional distributions given a causal diagram without relying on any parametric assumptions.

Theorem 3 (Do-calculus [Pearl, 1995])**.** *Let $\mathcal{G}$ be a causal diagram compatible with a structural causal model $\mathcal{M}$, with endogenous variables $\mathbf{V}$. For any disjoint $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{W}$, $\mathbf{Z} \subseteq \mathbf{V}$, the following rules are valid.*

Rule 1. $P(\mathbf{y} \mid do(\mathbf{w}), \mathbf{x}, \mathbf{z}) = P(\mathbf{y} \mid do(\mathbf{w}), \mathbf{z})$ *if $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{W} \cup \mathbf{Z}$ in $\mathcal{G}_{\overline{\mathbf{W}}}$*

Rule 2. $P(\mathbf{y} \mid do(\mathbf{w}), do(\mathbf{x}), \mathbf{z}) = P(\mathbf{y} \mid do(\mathbf{w}), \mathbf{x}, \mathbf{z})$ *if $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{W} \cup \mathbf{Z}$ in $\mathcal{G}_{\overline{\mathbf{W}},\underline{\mathbf{X}}}$*

Rule 3. $P(\mathbf{y} \mid do(\mathbf{w}), do(\mathbf{x}), \mathbf{z}) = P(\mathbf{y} \mid do(\mathbf{w}), \mathbf{z})$ *if $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{W} \cup \mathbf{Z}$ in $\mathcal{G}_{\overline{\mathbf{W}},\overline{\mathbf{X}(\mathbf{Z})}}$*

where $\mathbf{X}(\mathbf{Z}) \triangleq \mathbf{X} \setminus \mathtt{An}(\mathbf{Z})_{\mathcal{G}[\mathbf{V}\setminus\mathbf{W}]}$.

Latent projection. The *latent projection* of a causal diagram $\mathcal{G}$ over $\mathbf{V}$ onto $\mathbf{X}$ is a subgraph over $\mathbf{X}$ such that, in addition to including edges in $\mathcal{G}[\mathbf{X}]$, for every pair of distinct vertices $V_i, V_j \in \mathbf{X}$:

1. Add a directed edge $V_i \to V_j$ if there exists a directed path from V_i to V_j in $\mathcal{G}$ such that every non-endpoint vertex on the path is not in $\mathbf{X}$.
2. Add a bidirected edge $V_i \leftrightarrow V_j$ if there exists a divergent path $V_i \leftarrow \cdots \to V_j$ in $\mathcal{G}$ such that every non-endpoint vertex on the path is not in $\mathbf{X}$.

We denote by $\mathcal{G}\langle\mathbf{X}\rangle$ the latent projection of $\mathcal{G}$ onto $\mathbf{X}$.

Counterfactual variables. Given a set of variables $\mathbf{Y} \subseteq \mathbf{V}$, the solution for $\mathbf{Y}$ in $\mathcal{M}_{\mathbf{x}}$ defines a *potential response* for a unit $\mathbf{u}$, denoted as $\mathbf{Y}_{\mathbf{x}}(\mathbf{u})$. Averaging over the space of $\mathbf{U}$, a potential response $\mathbf{Y}_{\mathbf{x}}(\mathbf{u})$ induces a *counterfactual variable* $\mathbf{Y}_{\mathbf{x}}$ [Pearl, 2000]. By definition, $P_{\mathbf{x}}(\mathbf{y}) = P(\mathbf{y}_{\mathbf{x}})$.

Recently, Correa and Bareinboim [2025] introduced a novel calculus over probability quantities defined at the counterfactual level, called the *ctf-calculus*. We use only CTF-Rule 1 in this paper:

$$\textbf{CTF-Rule 1. (Consistency rule)} \qquad P(\mathbf{y}_{\mathbf{T}_*\mathbf{x}}, \mathbf{x}_{\mathbf{T}_*}, \mathbf{w}_*) = P(\mathbf{y}_{\mathbf{T}_*}, \mathbf{x}_{\mathbf{T}_*}, \mathbf{w}_*).$$

where $\mathbf{T}_* = \{T_{1[\mathbf{w}_1]}, T_{2[\mathbf{w}_2]}, \cdots\}$ represents an arbitrary counterfactual event (a set of counterfactual variables).

Graphical characterizations of MIS and POMIS. We present graphical criteria for MIS and POMIS.

Proposition 3 (Proposition 1 in Lee and Bareinboim [2018])**.** *Let $\mathcal{G}$ be a causal diagram over the set of variables $\mathbf{V}$. A set $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}$ is an MIS relative to $\langle\mathcal{G}, Y\rangle$ if and only if $\mathbf{X} \subseteq \mathtt{An}(Y)_{\mathcal{G}_{\overline{\mathbf{X}}}}$.*

Given a causal diagram $\mathcal{G}$, *minimal unobserved confounders' territory* (MUCT) and *interventional border* (IB) provide a graphical characterization of POMIS.

[1] A node is called a *collider* on a path if the path contains two edges having arrowheads toward the node, e.g., Z is a collider on $X \to Z \leftarrow Y$.

Definition 5 (Unobserved-confounders' territory [Lee and Bareinboim, 2018])**.** Let $\mathcal{H} = \mathcal{G}[\mathtt{An}(Y)_{\mathcal{G}}]$. A set of variables $\mathbf{T} \subseteq \mathbf{V}(\mathcal{H})$ containing Y is called a *UC-territory* on $\mathcal{G}$ with respect to Y if $\mathtt{De}(\mathbf{T})_{\mathcal{H}} = \mathbf{T}$ and $\mathtt{cc}(\mathbf{T})_{\mathcal{H}} = \mathbf{T}$. If there is no $\mathbf{T}' \subsetneq \mathbf{T}$, we refer to it as a *minimal UC-territory* (MUCT) denoted as $\mathsf{MUCT}(\mathcal{G}, Y)$.

Definition 6 (Interventional border [Lee and Bareinboim, 2018])**.** Let $\mathbf{T}$ be a minimal UC-territory on a causal diagram $\mathcal{G}$ with respect to Y. Then $\mathtt{Pa}(\mathbf{T})_{\mathcal{G}} \setminus \mathbf{T}$ is called an *interventional border* (IB) for $\mathcal{G}$ with respect to Y denoted as $\mathsf{IB}(\mathcal{G}, Y)$.

MUCT is the minimal set of variables that is closed under both descendants and bidirected connections; and IB consists of the parents of MUCT, excluding MUCT itself. Intuitively, MUCT is the minimal closed mechanism that conveys all hidden information from unobserved confounders to the downstream reward, while IB consists of the nodes that directly affect this closed mechanism.

Theorem 4 (Theorem 6 in Lee and Bareinboim [2018])**.** *Given information $\langle \mathcal{G}, Y \rangle$, $\mathbf{X} \subseteq \mathbf{V} \setminus \{Y\}$ is a POMIS with respect to $\langle \mathcal{G}, Y \rangle$ if and only if $\mathsf{IB}(\mathcal{G}_{\overline{\mathbf{X}}}, Y) = \mathbf{X}$.*

Proposition 4 (Theorem 4 in Lee and Bareinboim [2019])**.** *Given $\langle \mathcal{G}, Y, \mathbf{N} \rangle$, $\mathbb{P}^{\mathbf{N}}_{\mathcal{G},Y} = \mathbb{P}_{\mathcal{H},Y}$ where $\mathcal{H} = \mathcal{G}\langle \mathbf{V} \setminus \mathbf{N} \rangle$.*

C PROOFS

In this section, we provide detailed proofs of the propositions and theorems presented in the main body of the paper.

C.1 PROOF OF PROPOSITION 1

Lemma 1 (Lemma 3 in Tian and Pearl [2003])**.** *Let $\mathbf{C} \subseteq \mathbf{C}' \subseteq \mathbf{V}$, and suppose that $\mathbf{C}$ is an ancestral set in $\mathcal{G}[\mathbf{C}']$, then:*

$$\sum_{\mathbf{c}' \setminus \mathbf{C}} Q[\mathbf{C}'] = Q[\mathbf{C}]. \tag{26}$$

Proposition 1. *Let $Q^i[\mathbf{D}]$ be an identifiable c-factor from $\pi^i \in \Pi$ such that $\mathbf{C} \cap \Delta^i = \emptyset$ and $\mathbf{C} \subsetneq \mathbf{D}$. Let $\mathbf{C}' = \mathtt{An}(\mathbf{C})_{\mathcal{G}[\mathbf{D}]}$. Suppose that $\mathbf{C} = \mathtt{An}(\mathbf{C})_{\mathcal{G}[\mathbf{C}']}$. Then $Q^*[\mathbf{C}]$ is bounded as follows. If $Q^i[\mathbf{C}']$ is not identifiable in π^i, then:*

$$\sum_{\mathbf{c}' \setminus \mathbf{C}} Q^i[\mathbf{D}] \leq Q^*[\mathbf{C}] \tag{10}$$

$$\leq \sum_{\mathbf{c}' \setminus \mathbf{C}} \{Q^i[\mathbf{D}] + 1 - \sum_{\mathbf{c}'} Q^i[\mathbf{D}]\}. \tag{11}$$

Otherwise, $Q^*[\mathbf{C}] = \sum_{\mathbf{c}' \setminus \mathbf{C}} Q^i[\mathbf{C}']$ (i.e., transportable).

Proof of Prop. 1. We have:

$$\begin{aligned} Q^*[\mathbf{C}] &= P^{\mathbf{\Delta}}_{\mathbf{v} \setminus \mathbf{C}}(\mathbf{c} \mid \mathbf{S}^i = \mathbf{0}, \mathbf{S}^{-i} = \mathbf{0}) = P^{\mathbf{\Delta}}_{\mathbf{v} \setminus \mathbf{C}}(\mathbf{c} \mid \mathbf{S}^i = \mathbf{i}, \mathbf{S}^{-i} = \mathbf{0}) \\ &= Q^i[\mathbf{C}] \end{aligned}$$

holds by $\mathbf{C} \cap \Delta^i = \emptyset$.

According to Lem. 1 and $\mathbf{C} = \mathtt{An}(\mathbf{C})_{\mathcal{G}[\mathbf{C}']}$, we have:

$$Q^*[\mathbf{C}] = Q^i[\mathbf{C}] \tag{27}$$

$$= \sum_{\mathbf{c}' \setminus \mathbf{C}} Q^i[\mathbf{C}']. \tag{28}$$

First, suppose that $Q^i[\mathbf{C}']$ is identifiable. Then it is straightforward that $Q^*[\mathbf{C}]$ is transportable from π^i as shown in Eq. (28).

Otherwise, since $Q^i[\mathbf{C}']$ is not identifiable from π^i, we bound the c-factor using the identifiable c-factor $Q^i[\mathbf{D}]$ as follows. Let $\mathbf{T} \triangleq \mathbf{D} \setminus \mathbf{C}'$, i.e., $\mathbf{D} = \mathbf{T} \cup \mathbf{C}'$. By basic probabilistic algebra,

$$Q^i[\mathbf{C}'](\mathbf{v}) = P^i(\mathbf{c}'_{\mathbf{v} \setminus \mathbf{C}'}) \qquad \text{def.} \tag{29}$$

$$= \sum_{\mathbf{t}'} P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{C}'}, \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \text{marg. over } \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}} \tag{30}$$

$$= \sum_{\mathbf{t}'} P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D},\mathbf{v}\cap\mathbf{T}}, \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \mathbf{v}\setminus\mathbf{C}' = (\mathbf{v}\setminus\mathbf{D})\cup(\mathbf{v}\cap\mathbf{T}) \tag{31}$$

$$\geq P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D},\mathbf{t}'}, \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \text{algebra w/ } \mathbf{t}' = \mathbf{v}\cap\mathbf{T} \tag{32}$$

$$= P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D}}, \mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) \quad \text{CTF-Rule 1} \tag{33}$$

$$= P^i(\mathbf{d}_{\mathbf{v}\setminus\mathbf{D}}) \quad \mathbf{D} = \mathbf{T}\cup\mathbf{C}' \tag{34}$$

$$= Q^i[\mathbf{D}](\mathbf{v}). \quad \text{def.} \tag{35}$$

We thus have $Q^i[\mathbf{C}'] \geq Q^i[\mathbf{D}]$.

Similarly, we derive an upper causal bound for $Q^i[\mathbf{C}']$.

Using basic probabilistic algebra,

$$Q^i[\mathbf{C}'](\mathbf{v}) = P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{C}'}) \quad \text{def.} \tag{36}$$

$$= \sum_{\mathbf{t}'} P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{C}'}, \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \text{marg. over } \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}} \tag{37}$$

$$= \sum_{\mathbf{t}'} P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D},\mathbf{v}\cap\mathbf{T}}, \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \mathbf{v}\setminus\mathbf{C}' = (\mathbf{v}\setminus\mathbf{D})\cup(\mathbf{v}\cap\mathbf{T}) \tag{38}$$

$$= P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D},\mathbf{v}\cap\mathbf{T}}, \mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) + \sum_{\mathbf{t}'\neq\mathbf{t}} P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D},\mathbf{v}\cap\mathbf{T}}, \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \text{decomposition by isolating } \mathbf{t} = \mathbf{v}\cap\mathbf{T} \tag{39}$$

$$= P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D},\mathbf{t}}, \mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) + \sum_{\mathbf{t}'\neq\mathbf{t}} P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D},\mathbf{v}\cap\mathbf{T}}, \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \mathbf{t} = \mathbf{v}\cap\mathbf{T} \tag{40}$$

$$= P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D}}, \mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) + \sum_{\mathbf{t}'\neq\mathbf{t}} P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D},\mathbf{v}\cap\mathbf{T}}, \mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \text{CTF-Rule 1} \tag{41}$$

$$\leq P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D}}, \mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) + \sum_{\mathbf{t}'\neq\mathbf{t}} P^i(\mathbf{t}'_{\mathbf{v}\setminus\mathbf{D}}) \quad \text{algebra} \tag{42}$$

$$= P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D}}, \mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) + 1 - P^i(\mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) \tag{43}$$

$$= P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D}}, \mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) + 1 - \sum_{\mathbf{c}'} P^i(\mathbf{c}'_{\mathbf{v}\setminus\mathbf{D}}, \mathbf{t}_{\mathbf{v}\setminus\mathbf{D}}) \quad \text{marg. over } \mathbf{c}'_{\mathbf{v}\setminus\mathbf{D}} \tag{44}$$

$$= P^i(\mathbf{d}_{\mathbf{v}\setminus\mathbf{D}}) + 1 - \sum_{\mathbf{c}'} P^i(\mathbf{d}_{\mathbf{v}\setminus\mathbf{D}}) \quad \mathbf{D} = \mathbf{T}\cup\mathbf{C}' \tag{45}$$

$$= Q^i[\mathbf{D}](\mathbf{v}) + 1 - \sum_{\mathbf{c}'} Q^i[\mathbf{D}](\mathbf{v}). \quad \text{def.} \tag{46}$$

We thus have:

$$Q^i[\mathbf{C}'] \leq Q^i[\mathbf{D}] + 1 - \sum_{\mathbf{c}'} Q^i[\mathbf{D}] \tag{47}$$

Therefore, combining this result with R.H.S in Eq. (28), we have:

$$\sum_{\mathbf{c}'\setminus\mathbf{C}} Q^i[\mathbf{D}] \leq Q^*[\mathbf{C}] = \sum_{\mathbf{c}'\setminus\mathbf{C}} Q^i[\mathbf{C}'] \leq \sum_{\mathbf{c}'\setminus\mathbf{C}} \{Q^i[\mathbf{D}] + 1 - \sum_{\mathbf{c}'} Q^i[\mathbf{D}]\} \tag{48}$$

which concludes the proof. □

C.2 PROOF OF THEOREM 1

Lemma 2 (Proposition 1 in Lee et al. [2020]). *$(\mathbf{S}^i \perp\!\!\!\perp_d \mathbf{Y})_{\mathcal{G}^{\Delta}\setminus\mathbf{X}}$ at Line 9 in* GTRU *is equivalent to $\Delta^i \cap (\mathbf{V}\setminus\mathbf{X}) = \emptyset$.*

Theorem 1. PATR *returns valid closed-form expressions for lower/upper bounds.*

Proof of Thm. 1. Let a superscript l denote variables and values local to the function. The soundness of the algorithm was partially proved by Lee et al. [2020] for the case in which the given query is transportable.

It suffices to show that:

$$Q^i[\mathbf{V}_l] \leq P^*_{\mathbf{x}_l}(\mathbf{v}_l \setminus \mathbf{X}_l) \leq Q^i[\mathbf{V}_l] + 1 - \sum_{\mathbf{v}_l \setminus \mathbf{X}_l} Q^i[\mathbf{V}_l] \tag{49}$$

at Line 12. Note that $\mathbf{Y}_l = \mathtt{An}(\mathbf{Y}_l)_{\mathcal{G} \setminus \mathbf{X}_l}$ by Line 4 in PATR and $P^*_{\mathbf{x}_l}(\mathbf{v}_l \setminus \mathbf{X}_l) = P^i_{\mathbf{x}_l}(\mathbf{v}_l \setminus \mathbf{X}_l)$ by Lem. 2. Here, Line 9 in GTRU corresponds to Line 6 in PATR, which has already been applied.

Moreover, it holds that $\mathbf{V}_l = \mathtt{An}(\mathbf{Y}_l)_{\mathcal{G}_l}$ and $\mathbf{V}_l$ is a c-component in $\mathcal{G}_l$, as ensured by earlier steps in PAID. Therefore, we have:

$$Q^i[\mathbf{V}_l] = P^i(\mathbf{v}_l) \leq P^*_{\mathbf{x}_l}(\mathbf{v}_l \setminus \mathbf{X}_l) \tag{50}$$

$$\leq P^i(\mathbf{v}_l) + 1 - \sum_{\mathbf{v}_l \setminus \mathbf{X}_l} P^i(\mathbf{v}_l) = Q^i[\mathbf{V}_l] + 1 - \sum_{\mathbf{v}_l \setminus \mathbf{X}_l} Q^i[\mathbf{V}_l] \tag{51}$$

due to Prop. 1.

Hence, we have:

$$P^i(\mathbf{v}_l) \leq P^*_{\mathbf{x}_l}(\mathbf{v}_l \setminus \mathbf{X}_l) \leq P^i(\mathbf{v}_l) + 1 - \sum_{\mathbf{v}_l \setminus \mathbf{X}_l} P^i(\mathbf{v}_l). \tag{52}$$

Finally, taking $P^*_{\mathbf{x}_l}(\mathbf{y}_l) = \sum_{\mathbf{v}_l \setminus (\mathbf{X}_l \cup \mathbf{Y}_l)} P^*_{\mathbf{x}_l}(\mathbf{v}_l \setminus \mathbf{X}_l)$ leads to the following:

$$\sum_{\mathbf{v}_l \setminus (\mathbf{X}_l \cup \mathbf{Y}_l)} P^i(\mathbf{v}_l) \leq P^*_{\mathbf{x}_l}(\mathbf{y}_l) \leq \sum_{\mathbf{v}_l \setminus (\mathbf{X}_l \cup \mathbf{Y}_l)} \left\{ P^i(\mathbf{v}_l) + 1 - \sum_{\mathbf{v}_l \setminus \mathbf{X}_l} P^i(\mathbf{v}_l) \right\} \tag{53}$$

which concludes the proof. □

C.3 PROOF OF COROLLARY 1

Corollary 1. *Let $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$ be a bound from* CAUSALBOUND. *Then $\mu^*_{\mathbf{x}}$ is partially transportable with respect to $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$ with the bound given by $[\ell_{\mathbf{x}}, u_{\mathbf{x}}]$.*

Proof of Cor. 1. In Thm. 1, it has been established that the expressions returned by PATR$(\mathsf{type} = \ell)$ and PATR$(\mathsf{type} = u)$ are sound. Since $0 \leq$ PATR$(\mathsf{type} = \ell) \leq P^*_{\mathbf{x}}(y)$ for every $y \in \Omega_Y$, the lower causal bound $0 \leq \ell_{\mathbf{x}} \leq \mu^*_{\mathbf{x}}$ holds.

Moreover, the min operator ensures that $P^*_{\mathbf{x}}(y) \leq \min\{1, \text{PATR}(\mathsf{type} = u)\} \leq 1$ for any $y \in \Omega_Y$, so the upper bound $u_{\mathbf{x}}$ is valid. Furthermore, the resulting bounds are strictly contained in $[0, 1]$ as long as the source distributions $\mathcal{P}^{\Pi}_{\mathbb{Z}}$ are informative. Hence, the resulting interval constitutes a valid partially transportable bound. □

C.4 PROOF OF PROPOSITION 2

Proposition 2. *A POMIS with respect to $\langle \mathcal{G}, Y, \mathbf{N} \rangle$ is minimal with respect to $\langle \mathcal{G}, Y \rangle$.*

Proof of Prop. 2. Let $\mathbf{X}$ be a POMIS with respect to $\langle \mathcal{G}, Y, \mathbf{N} \rangle$. Then, $\mathbf{X}$ is an MIS under the same information, by definition. We will show that $\mathbf{X}$ is also an MIS with respect to $\langle \mathcal{G}, Y \rangle$ by proving the contrapositive statement.

Suppose that $\mathbf{X}$ is *not* an MIS with respect to $\langle \mathcal{G}, Y \rangle$. Then, there exists a node $X \in \mathbf{X}$ such that there is no *proper directed path*[2] from X to Y with respect to $\mathbf{X}$ in $\mathcal{G}$ (Thm. 4). That is, every directed path from X to Y in $\mathcal{G}$ forms:

$$X \to V_1 \to \cdots \to V_n \to Z \to W_1 \to \cdots \to W_m \to Y \tag{54}$$

[2]We refer to a directed path from $X \in \mathbf{X}$ to Y as a *proper* directed path with respect to $\mathbf{X}$ if only the first node X belongs to $\mathbf{X}$.

with $n, m \geq 0$ for an arbitrary $Z \in \mathbf{X} \setminus \{X\}$. Since the latent projection does not introduce any directed edges from V_i to W_j in $\mathcal{G}\langle \mathbf{V} \setminus \mathbf{N} \rangle$, all such paths correspond to non-proper directed paths from X to Y with respect to $\mathbf{X}$ in $\mathcal{H} \triangleq \mathcal{G}\langle \mathbf{V} \setminus \mathbf{N} \rangle$. Therefore, $\mathbf{X}$ is *not* an MIS with respect to $\langle \mathcal{H}, Y \rangle$. Combining this with Prop. 4 establishes the contrapositive, which concludes the proof. □

C.5 PROOF OF THEOREM 2

Theorem 2 (Upper dominance). *Let $\mathbf{r}^\star$ be the optimal action with respect to $\langle \mathcal{G}, Y, \mathbf{N} \rangle$ where $\mathbf{N}$ is a proper subset of $\mathbf{N}^*$. Then $\mu^*_{\mathbf{x}^\star}$ is upper bounded by $\mu^*_{\mathbf{r}^\star}$.*

Proof of Thm. 2. Without loss of generality, we assume $\mathbf{N} = \emptyset$ since we can equivalently reformulate the graph as $\mathcal{G} = \mathcal{G}\langle \mathbf{V} \setminus \mathbf{N} \rangle$ and the constraint set as $\mathbf{N}^* = \mathbf{N}^* \setminus \mathbf{N}$.

According to Prop. 2, any POMIS with respect to $\langle \mathcal{G}, Y, \mathbf{N}^* \rangle$ is also an unconstrained MIS. Therefore, the following holds:

$$\underbrace{\mu^*_{\mathbf{x}^\star}}_{\text{constraint } \mathbf{N}^*} \leq \overbrace{\mu^*_{\mathbf{r}^\star}}^{\text{unconstrained}} \tag{55}$$

since the optimal action $\mathbf{x}^\star$ in $\mathcal{A}^{\mathbf{N}^*}_{\mathcal{G},Y}$ belongs to the collection of MISs $\mathbb{M}_{\mathcal{G},Y}$. □

D DETAILS ON EXPERIMENTAL SETTINGS

The experiments were conducted on a Linux server equipped with an Intel® Xeon® Gold 5317 processor running at 3.0 GHz and 64 GB of RAM. No GPUs were used during the simulations.

Detailed explanation of worked examples. We provide details on how the TRUCB (Alg. 4) works and the specific SCMs used in all bandit instances. We denote the exclusive-or operation by $\oplus$, and use $\texttt{Bern}$ to represent a Bernoulli distribution. We randomly generate structural functions $\mathcal{F}$ using binary logical operations ($\wedge$, $\vee$, $\oplus$, $\neg$), and randomly select the parameters of the exogenous variable distributions. In the example below, we truncate all numerical values to three decimal places for readability.

Task 1. The bandit instance is associated with an SCM $\mathcal{M}$ where

$$\mathcal{M} = \begin{cases} \mathbf{U} & = \{U_A, U_B, U_Y, U_{AY}\} \\ \mathbf{V} & = \{A, B, Y\} \\ \mathcal{F} & = \begin{cases} f_A = u_A \vee (b \oplus u_{AY}), f_B = u_B, \\ f_Y = (1 - b) \vee ((u_{AY} \vee a) \oplus u_Y) \end{cases} \\ P(\mathbf{u}) & = \begin{cases} U_A \sim \texttt{Bern}(0.33), U_B \sim \texttt{Bern}(0.38), \\ U_Y \sim \texttt{Bern}(0.41), U_{AY} \sim \texttt{Bern}(0.75). \end{cases} \end{cases} \tag{56}$$

The learner has an experimental prior $\mathbb{Z}^1 = \{\{B\}\}$ from $\Delta^1 = \{A\}$. The action space without prior information corresponds to $\mathbb{P}^{\mathbf{N}^*=\emptyset}_{\mathcal{G},Y} = \{\{B\}, \{A, B\}\}$, which corresponds to the action space of the baseline algorithm POUCB (i.e., the initialized target action space).

In this setting, we observe that μ^*_b cannot be bounded from the source π^1, since $\{A\} \cap \Delta^1 \neq \emptyset$. In contrast, $\mu^*_{a,b}$ can be bounded as follows:

$$\textstyle\sum_y y P^1_b(a, y) \leq \mu^*_{a,b} \leq \sum_y y \big\{P^1_b(a, y) + 1 - P^1_b(a)\big\}. \tag{57}$$

Using these expressions, the learner can estimate causal bounds for four actions corresponding to the POMIS:

Then, TRUCB computes a dominance bound as $[\ell^\star, u^\star] = [\mathbf{0.8325}, 1]$. Here, $u^\star = 1$ arises since the causal bounds for $do(b)$ are $[0, 1]$, i.e., no information can be transported for that intervention. Among the four actions, $do(A = 0, B = 1)$ and $do(A = 1, B = 1)$ have upper causal bounds lower than the lower dominance bound, i.e., $u_{\mathbf{x}} < \ell^\star$ (red), leading them to be eliminated by TRUCB. Accordingly, the algorithm begins the online interaction, excluding these two actions.

A	B	$\ell_{\mathbf{x}}$	$\mu^*_{\mathbf{x}}$	$u_{\mathbf{x}}$
0	0	0.1675	1.0000	1.0000
0	1	0.2965	0.5450	0.7940
1	0	**0.8325**	1.0000	1.0000
1	1	0.2935	0.5900	0.7960

Task 2. The bandit instance is associated with an SCM $\mathcal{M}$ where

$$\mathcal{M} = \begin{cases} \mathbf{U} & = \{U_A, U_B, U_C, U_Y, U_{BC}, U_{BY}\} \\ \mathbf{V} & = \{A, B, C, Y\} \\ \mathcal{F} & = \begin{cases} f_A = u_A, f_B = (u_{BY} \wedge u_{BC}) \oplus u_B, \\ f_C = b \oplus (u_{BC} \wedge ((1-u_C) \oplus a)), \\ f_Y = c \oplus (u_{BY} \wedge ((1-u_Y) \oplus a)) \end{cases} \\ P(\mathbf{u}) & = \begin{cases} U_A \sim \texttt{Bern}(0.28), U_B \sim \texttt{Bern}(0.42), U_C \sim \texttt{Bern}(0.52) \\ U_Y \sim \texttt{Bern}(0.47), U_{BC} \sim \texttt{Bern}(0.62), U_{BY} \sim \texttt{Bern}(0.49). \end{cases} \end{cases} \tag{58}$$

A learner has access to prior data from two environments with potential discrepancies $\Delta^1 = \{A\}$ and $\Delta^2 = \{B\}$. The first source π^1 provides observational prior information, with $\mathbb{Z}^1 = \{\emptyset\}$, and the second source π^2 provides experimental prior information, with $\mathbb{Z}^2 = \{\{C\}\}$. The non-manipulable variables constraint is $\mathbf{N}^* = \{A, C\}$, and the corresponding initial action space is $\mathbb{P}^{\mathbf{N}^*}_{\mathcal{G},Y} = \{\emptyset, \{B\}\}$.

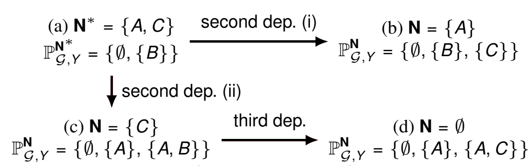
First, the action $do(\emptyset)$ is transportable as follows:

$$\mu^*_\emptyset = \textstyle\sum_{a,b,c,y} yQ^2[A]Q^1[B, C, Y] = \sum_{a,b,c,y} yP^2_c(a)P^1(b, c, y \mid a) = \mathbf{0.4844}. \tag{59}$$

On the other hand, $do(b)$ is *not* transportable, since $Q^*[C]$ is *non*-transportable from the sources Π. We have the following causal bounds (left):

B	$\ell_{\mathbf{x}}$	$\mu^*_{\mathbf{x}}$	$u_{\mathbf{x}}$
0	0.2097	0.4031	0.6783
1	0.2752	0.5969	0.8066

(a) Causal bounds



(b) UDB traversal

We now proceed to compute the dominance bounds. We have $\mathbb{M}^{\mathbf{N}^*}_{\mathcal{G},Y} = \{\emptyset, \{B\}\}$. Since $\mu^*_\emptyset = 0.4844$ and $\ell_{b^*} = 0.2752$, the lower dominance bound is given by $\ell^\star = \mathbf{0.4844}$. Thus, no actions are eliminated.

To compute the upper dominance bound, UDB initializes $u^\star$ to 0.8066.

First depth. Since it is initialized with the upper causal bound rather than the transferred value, it traverses weaker constraints.

Second depth (i). For $\mathbf{N} = \{A\}$, we have $\mathbb{P}^{\mathbf{N}}_{\mathcal{G},Y} = \{\emptyset, \{B\}, \{C\}\}$ and $\mu^*_{c^*} = 0.7485$. Therefore, this call returns $u^\star = 0.8066$.

Second depth (ii). For $\mathbf{N} = \{C\}$, we have $\mathbb{P}^{\mathbf{N}}_{\mathcal{G},Y} = \{\{A\}, \{A, B\}\}$. This call sets $u^\star = 0.8070$, which corresponds to the upper causal bound for $do(A{=}0, B{=}1)$. Again, since it is updated with the upper causal bound rather than the transferred value, this call traverses weaker constraints.

Third depth. For $\mathbf{N} = \emptyset$, we have $\mathbb{P}^{\mathbf{N}}_{\mathcal{G},Y} = \{\{A\}, \{A, C\}\}$. This call updates $u^\star = 0.7697$, which corresponds to the transportable action $do(A = 1, C = 1)$—$\mu^*_{do(A=1,C=1)} = \sum_y yP^2_{do(C=1)}(y \mid A = 1) = 0.7697$.

Consequently, the final upper dominance bound is set to $u^\star = 0.7697 = \min\{0.8066, \underbrace{\min\{0.8070, \overbrace{0.7697}^{\text{third dep.}}\}}_{\text{second dep.}}\}$.

Since $u_{do(B=1)} = 0.8066 > 0.7697 = u^\star$, the bound for $do(B = 1)$ is strengthened to $u^\star = 0.7697$ in the online learning phase, i.e., $\min(u_{do(B=1)}, u^\star) = u^\star$. Although the size of the action space remains unchanged (i.e., no action is removed from $\mathcal{A}^*$, implying that the action spaces of all three algorithms—UCB, POUCB and TRUCB—are identical), we observe that accounting for the transferred bounds improves performance.

Task 3. The bandit instance is associated with an SCM $\mathcal{M}$ where

$$\mathcal{M} = \begin{cases} \mathbf{U} & = \{U_R, U_T, U_W, U_X, U_Z, U_Y, U_{RW}, U_{RY}, U_{XY}, U_{WX}, U_{WZ}\} \\ \mathbf{V} & = \{R, T, W, X, Z, Y\} \\ \mathcal{F} & = \begin{cases} f_R = u_R \vee ((1 - u_{RW}) \wedge (t \wedge u_{RY})), f_T = u_T \\ f_W = u_W \vee ((1 - u_{WX}) \wedge (u_{RW} \wedge u_{WZ})), \\ f_X = w \vee ((1 - u_{WX}) \wedge (u_{XY} \wedge u_X)), \\ f_Z = x \vee ((1 - u_{WZ}) \wedge (r \wedge u_Z)), \\ f_Y = z \vee ((1 - u_{XY}) \wedge (u_{RY} \wedge u_Y)) \end{cases} \\ P(\mathbf{u}) & = \begin{cases} U_R \sim \texttt{Bern}(0.53), U_T \sim \texttt{Bern}(0.63), U_W \sim \texttt{Bern}(0.38), \\ U_X \sim \texttt{Bern}(0.27), U_Z \sim \texttt{Bern}(0.4), U_Y \sim \texttt{Bern}(0.26), \\ U_{RW} \sim \texttt{Bern}(0.52), U_{RY} \sim \texttt{Bern}(0.63), U_{XY} \sim \texttt{Bern}(0.79), \\ U_{WX} \sim \texttt{Bern}(0.74), U_{WZ} \sim \texttt{Bern}(0.31). \end{cases} \end{cases} \tag{60}$$

A learner has access to priors from two environments with potential discrepancies $\Delta^1 = \{T\}$ and $\Delta^2 = \{R\}$, with priors $\mathbb{Z}^1 = \{\emptyset, \{Z\}\}$ and $\mathbb{Z}^2 = \{\{Z\}\}$, and constraint $\mathbf{N}^* = \{T, W\}$. The initial action space is $\mathbb{P}^{\mathbf{N}^*}_{\mathcal{G},Y} = \{\emptyset, \{R\}, \{X\}, \{Z\}\}$. We have:

$$\mu^*_\emptyset = \textstyle\sum_{r,t,w,x,z,y} yP^1(r, x, y, z \mid t, w)P^2_z(t) \tag{61}$$
$$\mu^*_z = \textstyle\sum_y yP^1_z(y) = \sum_y yP^2_z(y) \tag{62}$$

which are transportable, while the following quantities are non-transportable:

$$\mu^*_x = \textstyle\sum_{r,t,z,y} yQ^1[R, Y]Q^2[T]\, \overbrace{Q^*[Z]}^{\text{non-transportable}} \tag{63}$$
$$\mu^*_r = \textstyle\sum_{w,x,z,y} y\, \underbrace{Q^*[W, X, Z, Y]}_{\text{non-transportable}}. \tag{64}$$

The factor $Q^*[Z]$ can be bounded using $Q^1[W, R, X, Z, Y]$, leading to the following result:

$$\sum_y yP^1(w^\dagger) \sum_{r,t,z} P^1(r, x, z|t, w^\dagger)P^1_z(r, y|t)P^2_{z^\dagger}(t) \leq \mu^*_x \tag{65}$$
$$\leq \sum_y y \min\Big\{ \sum_{r,t,z} \big(P^1(r, x, z|t, w^\dagger)P^1(w^\dagger) + 1 - P^1(r, x|t, w^\dagger)P^1(w^\dagger)\big)P^1_z(r, y|t)P^2_{z^\dagger}(t)\Big\} \tag{66}$$

where † denotes externally assigned values.

Similarly, we can derive the causal bound for μ^*_r using $Q^1[W, R, X, Z, Y]$ as follows:

$$\sum_y yP^1(r, y|t^\dagger) \leq \mu^*_r \leq \sum_y y \min\Big\{1, \sum_{w,x,z} P^1(r, w, x, y, z \mid t^\dagger) + 1 - P^1(r, w, x, z \mid t^\dagger)\Big\}. \tag{67}$$

In this setting, a transportable action yields $\mu^*_{do(Z=1)} = 1$, resulting in $\ell^\star = u^\star = 1$.

Furthermore, we obtain the upper causal bounds $u_{do(\emptyset)} = 0.5514$, $u_{do(X=0)} = 0.7901$ and $u_{do(Z=0)} = 0.034$. Since these upper bounds are all smaller than $\ell^\star$, the corresponding actions are eliminated from $\mathcal{A}^*$ by TRUCB.

E DISCUSSIONS

Parametric approach. One might be concerned that the algorithm (Alg. 4) eliminates non-optimal actions before learning begins but then switches to a traditional approach that ignores additional observations available during each round.

Indeed, there exists a rich body of research that incorporates prior knowledge to iteratively update parameters of SCMs under graphical constraints [Zhang and Bareinboim, 2022, Bellot et al., 2023, Wei et al., 2023, Jalaldoust et al., 2024] in online learning. However, such methods often rely on optimization-based approaches—such as *canonical SCM* [Zhang et al., 2022] or *neural causal models* (NCMs) [Xia et al., 2021] —that assume full parameterization, which results in high computational overhead. These approaches are infeasible for larger or denser causal diagrams, and exploiting them effectively remains an open problem [Elahi et al., 2024]. Moreover, these methods are constrained to categorical settings, which makes them inapplicable to continuous domains, such as those encountered in causal Bayesian optimization [Aglietti et al., 2020] and the multi-outcome variant [Bhatija et al., 2025], where POMISs are also leveraged for structural pruning in continuous action spaces.

In contrast, our approach focuses on leveraging structural knowledge offline, before any online interaction, and without requiring parameterization or any strong assumptions beyond a given graphical structure and sources. This distinction allows us to scale to settings where parameter learning is infeasible or computationally prohibitive (dense graphs or continuous domains), while still retaining provable regret guarantees through structure-informed pruning and closed-form bounding.

Computational complexity. Regarding the scalability of our framework, particularly the hierarchical traversal over constrained POMISs (i.e., UDB), an exhaustive search would incur a worst-case complexity exponential in the number of non-manipulable variables. However, our approach takes advantage of a strictly monotonic hierarchy: *bounds computed under weaker constraints are naturally looser*. This structural property allows our algorithm to traverse the space hierarchically and terminate early as soon as a valid dominance bound is found, bypassing unnecessary computation and significantly reducing practical runtime. Furthermore, all steps involving structural querying and transportability bounding (e.g., Algs. 1 and 3) are executed entirely *offline* as a pre-processing step. During the online interaction phase, our algorithm introduces negligible overhead over standard UCB, retaining high sample efficiency without trading off real-time performance.

Regret analysis. Our theoretical regret analysis bridges canonical frameworks for UCB-based online learning with transportability theory. The derived structural bounds—both the transportable causal bounds and the dominance bounds—are theoretically guaranteed to be sound and deterministic given a valid super-model of the environment (as discussed above). Because the true expectations lie within these bounds, they seamlessly satisfy the admissibility conditions required by existing bounded-reward analyses [Zhang and Bareinboim, 2017, Gong et al., 2023]. Rather than requiring a bespoke regret analysis from scratch, our framework demonstrates how offline structural information can deterministically shrink the effective action space and suppress the constant factors in the sub-linear regret bounds of any optimistic exploration algorithm.

Future work. Recently, research on transportability theory in practical settings has grown [Jung and Bellot, 2024, Jalaldoust et al., 2024, Jalaldoust and Bareinboim, 2024]. Beyond structural causal bandits, we believe that transportability-based decision making will offer substantial practical value when integrated with causal reinforcement learning [Zhang and Bareinboim, 2022, Hwang et al., 2024, Bareinboim et al., 2024], rehearsal learning [Qin et al., 2023, 2025, Du et al., 2024, 2025, Tao et al., 2025, 2026], and sequential planning [Pearl and Robins, 1995, Jung et al., 2024].