TECHNICAL REPORT
R-113
May, 2024

Counterfactual Realizability and Decision-Making

Arvind Raghavan and **Elias Bareinboim**
Causal Artificial Intelligence Lab
Columbia University
ar4284@columbia.edu, eb@cs.columbia.edu

Abstract

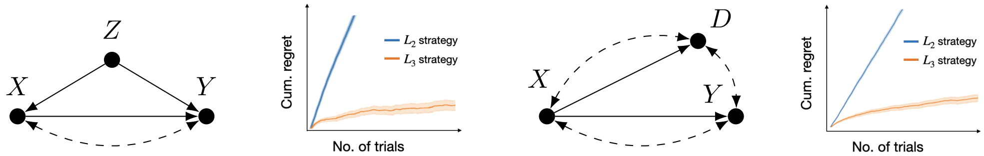
A commonly accepted belief suggests that, in a real-world environment, one can only draw samples from observational and interventional distributions, corresponding to Layers 1 and 2 of the *Pearl Causal Hierarchy* (or PCH). According to this understanding, Layer 3, representing counterfactual distributions (what would happen), remains inaccessible by definition. However, Bareinboim, Forney, and Pearl (2015) introduced a procedure that allows an agent to sample directly from a counterfactual distribution, opening up the possibility that other counterfactual quantities can be estimated directly via physical experimentation. In this paper, we investigate this proposition by introducing a formal definition of *realizability*, the ability to draw samples from a distribution, and then developing an algorithm to determine whether an arbitrary counterfactual distribution is realizable given fundamental physical constraints, such as the inability to go back in time and subject the same unit to a different experimental condition. Building on this new characterization, we further develop an algorithm for an agent to construct an optimal *realizable strategy* in multi-arm bandit settings. Contrary to general practice that assumes that an interventional strategy is the best that an agent can achieve, we show that a counterfactual strategy dominates interventional and observational ones (i.e. it is as good as or better than), and demonstrate the practical performance of counterfactual agents in simulations.

1 Introduction

The *Pearl Causal Hierarchy*, or PCH, is an important recent milestone in our understanding of causality [27, 5]. The three layers of the PCH represent the distinct regimes of *seeing*, *doing*, and *imagining*, with regard to an environment. Consider an environment involving a decision variable X (say, whether a person follows intermittent fasting) and an outcome Y (BMI). Layer 1, or $\mathcal{L}_1$, of the PCH represents distributions from the *observational* regime, such as $P(Y \mid x)$: the distribution of BMI among people who happen to adopt diet x. Layer 2 ($\mathcal{L}_2$) represents *interventional* distributions , such as the distribution of Y if a person is made to follow diet x, written symbolically as $P(Y; do(x))$. Layer 3 ($\mathcal{L}_3$) represents *counterfactual* distributions dealing with conflicting realities, such as $P(Y_x \mid x', y')$: the distribution of Y had X been fixed as x, given that X, Y were in fact observed to be x', y'. Each layer subsumes the one before it, but is under-determined by it [16, 5].

Suppose a scientist were interested in the $\mathcal{L}_3$-quantity $P(y_x \mid x')$, often called the *effect of the treatment on the treated*, or ETT [13, 14]. One approach to computing such quantities is through *identification* [24, §3.2.4]: leveraging causal knowledge about the environment, typically a causal graph or parametric assumptions, to infer the higher-layer quantity using lower-layer data. This approach fails when the quantity is non-identifiable, e.g. ETT in the general case [33, 8].

There is, however, another approach that uses physical experimentation to directly draw samples from the relevant distribution, $P(Y_x, X)$ in the ETT case, and then uses statistical methods to estimate $P(Y_x = y, X = x')$. This is only possible if there is some sequence of physical actions by which an



(a) Algo. sentencing (b) Cum. regret in a. (c) Tech surveillance (d) Cum. regret in c.

Figure 1: (a,c) Problems with decision variable X, outcome Y, and covariates Z, D; (b,d) Cumulative regret grows sub-linearly for $\mathcal{L}_3$-strategy but linearly for $\mathcal{L}_2$ strategy. For details, see Sec. 4.1.

agent can measure these random variables simultaneously for a single unit. It is generally believed that sampling from a distribution is only feasible when considering $\mathcal{L}_1$- and $\mathcal{L}_2$-distributions, the latter by interventions like randomized controlled trials (RCT), à la Fisher [11], and the former by simply observing the natural behaviour of the system. $\mathcal{L}_3$-distributions like $P(Y_x, X)$ are deemed non-realizable in general, as a single unit can only adopt either decision x or x'. However, Bareinboim, Forney & Pearl have shown it is feasible to draw samples from the ETT distribution through a novel *counterfactual randomization* procedure [4, 12]. This leaves open the possibility that other $\mathcal{L}_3$-distributions, say perhaps $P(Y_x, X, Y)$, are also realizable through clever experimental setups, allowing one to estimate important quantities like the *probability of sufficiency*, $P(y_x \mid y', x')$ [23].

This brings us to the central question motivating this work: *from which $\mathcal{L}_3$-distributions is it physically possible to draw samples given the fundamental constraints of nature?* More broadly, given that there are fundamental limitations like the inability to travel back in time and subject the original unit to a different experimental condition, how far up the PCH can one go via experimental procedures? Answering these questions has implications for decision-making. For concreteness, consider Fig. 1, showing simulations for two multi-arm bandit (MAB) settings. The standard approach in the literature uses allocation procedures (e.g., UCB, Thompson Sampling) to discover which arm x optimizes the expected outcome $\mathbb{E}[Y \mid c; do(x)]$ in Fig. 1(a) and $\mathbb{E}[Y; do(x)]$ in (c), which are both interventional ($\mathcal{L}_2$) strategies [34, 17]. It turns out there are superior strategies based on optimizing the counterfactual ($\mathcal{L}_3$) quantities $\mathbb{E}[Y_x \mid c, x']$ in (a) and $\mathbb{E}[Y_x \mid x', d_{x''}]$ in (c) (to be defined). Dismissing such possibilities leads to regret that grows linearly, since the $\mathcal{L}_2$-strategies do not come asymptotically closer to discovering the optimal arm in each round, as highlighted in Fig. 1(b,d).

The conceptual roadmap of this work is shown in Fig. 2 and its contributions are as follows:

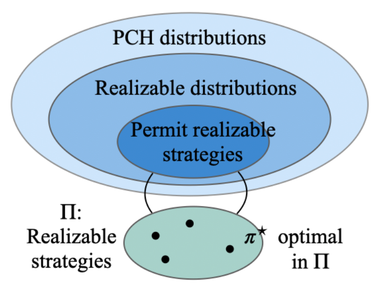


Figure 2: Paper's roadmap.

1. We introduce a formal definition for the *realizability of an $\mathcal{L}_3$-distribution* (Def. 3.4), and develop an algorithm to decide whether an arbitrary distribution is realizable (Algo. 1). We prove that the algorithm is complete (Thm. 3.5), and derive important corollaries characterizing realizable distributions (Cor. 3.7, 3.8).
2. We leverage this result to study *realizable decision strategies*. We develop a general algorithm by which an agent can adapt common MAB solvers to enact an optimal counterfactual ($\mathcal{L}_3$) strategy in decision problems with a known causal structure. We prove that this strategy dominates standard interventional ($\mathcal{L}_2$) ones (Thm. 4.1, Cor. 4.2), and empirically validate these results.

Preliminaries. We denote variables by capital letters, X, and values by small letters, x. Bold letters, $\mathbf{X}$, are sets of variables and $\mathbf{x}$ sets of values. $P(\mathbf{x})$ is shorthand for $P(\mathbf{X} = \mathbf{x})$. $\mathbb{1}[.]$ is the indicator function. We use *Structural Causal Models (SCM)* to describe the generative process for a system of interest [5, Def. 1][24]. An SCM $\mathcal{M}$ is a tuple $\langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$. $\mathbf{V}$ is the set of observable variables. $\mathbf{U}$ is the set of unobservable variables exogenous to the system, distributed according to $P^{\mathcal{M}}(\mathbf{U})$. $\mathcal{F} = \{f_V\}$ is a set of functions s.t. each f_V causally generates the value of $V \in \mathbf{V}$ as $V \leftarrow f_V(\mathbf{U}_V, \mathbf{Pa}_V)$, where $\mathbf{U}_V \subseteq \mathbf{U}$ and $\mathbf{Pa}_V \in \mathbf{V} \setminus V$. Each $\mathcal{M}$ induces a *causal diagram* $\mathcal{G}$ [5, Def. 13], which is a graph containing a vertex for each $V \in \mathbf{V}$, a directed edge from each node in $\mathbf{Pa}_V$ to V, and a bidirected edge between V, V' if $\mathbf{U}_V, \mathbf{U}_{V'}$ are not independent. Given a graph $\mathcal{G}$, $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{W}}}$ is the result of removing edges coming into variables in $\mathbf{X}$, and edges coming out of $\mathbf{W}$. We use standard terminology like parents, descendants of a node (see App. A). Our treatment is limited to *recursive* SCMs, which implies acyclic diagrams, with finite discrete domains over $\mathbf{V}$. The $do(\mathbf{x})$ operator indexes a sub-model $\mathcal{M}_{\mathbf{x}}$ where the functions generating variables $\mathbf{X}$ are replaced with constant values $\mathbf{x}$. A variable $Y \notin \mathbf{X}$ evaluated in this regime is called a *potential response*, denoted $Y_{\mathbf{x}}$. ($\mathbf{W}_\star = \mathbf{w}$) denotes an arbitrary counterfactual event, e.g. ($Y_x = y \wedge Y_{x'} = y' \wedge X = x''$). The probability of such an event is given by the $\mathcal{L}_3$-valuation [5, Def. 7]: $P^{\mathcal{M}}(\mathbf{W}_\star = \mathbf{w}) = \sum_{\mathbf{u}} \left(\prod_{W_{\mathbf{t}} \in \mathbf{W}_\star} \mathbb{1}[W_\star(\mathbf{u}) = w] \right) P^{\mathcal{M}}(\mathbf{u})$, with w taken from $\mathbf{w}$.

2 Data-collection procedures

In this section, we define a procedure, *counterfactual randomization*, that extends the scope of traditional *Fisherian* experimentation (discussed below). Consider a system of interest modeled by unknown SCM $\mathcal{M}$. Interventions and counterfactual events are typically defined in terms of *symbolic* operations on $\mathcal{M}$. To conceptually separate this from the *physical* constraints experienced by an agent (natural or artificial), we define the following physical actions that an agent can perform in the system. These are simply the physical counterparts to symbolic procedures.

We call each discrete episode of the system's behaviour a *unit*. Examples of units are patients in a clinical trial, neighbourhoods in a social science experiment, rounds played on a slot machine etc. We index units WLOG by $i = 1, 2, 3...$, which constitute a *target population* in the system.

Definition 2.1 (Physical actions)**.** (1) $\textsc{Select}^{(i)}$: randomly choosing, without replacement, a unit i from the target population, to observe in the system; (2) $\textsc{Read}(V)^{(i)}$: measuring the way in which a causal mechanism $f_V \in \mathcal{F}$ has physically affected unit i, by observing its realized feature $V^{(i)}$; (3) $\textsc{Rand}(X)^{(i)}$: erasing and replacing i's natural mechanism f_X for a decision variable X with an enforced value drawn from a randomizing device having support over Domain(X). ■

$\textsc{Read}(V)^{(i)} = v$ and $\textsc{Rand}(X)^{(i)} = x$ are also overloaded to refer to the *values* read and enforced, respectively. $\textsc{Rand}(X)^{(i)}$ is the standard Fisherian randomization of a decision variable X, corresponding to the symbolic procedure of a *stochastic* intervention on X [7].[1] As $\textsc{Rand}(X)^{(i)}$ erases the unit i's natural decision, $\textsc{Read}(X)^{(i)}$ will yield the value randomly assigned to unit i. The discovery of this procedure marked an important achievement in the history of science and experiment-design [10, 11]. Since the use of a randomizing device eliminates by design any confounding between the assigned decision and the unit's latent attributes $\mathbf{U}^{(i)}$, it allows researchers to estimate causal effects.

The actions in Def. 2.1 are sufficient for an agent to physically draw samples from any $\mathcal{L}_1$- or $\mathcal{L}_2$-distribution, as discussed in App. C.4. Until recently, it was generally presumed these were the only physical actions possible on units in a system. However, we discuss some important extensions of experimental capabilities next.

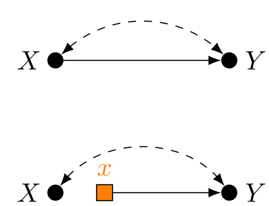


Figure 3: (Top) Causal diagram with decision variable X; (Bottom) Schematic illustration of the procedure of randomizing the actual decision without erasing the unit's natural decision.

Counterfactual data-collection procedures. In an important work from the causal reinforcement learning literature, Bareinboim, Forney & Pearl describe an experimental setting in which it is possible to both randomize a unit's actual decision, and also record the natural decision the unit *would have normally* taken [4, 12]. This procedure has subsequently been used to set benchmarks in counterfactual decision-making [36]. These settings involve an agent introspecting to gauge their natural choice, or otherwise revealing their natural choice by some indication, e.g. physical gestures prior to decision-time. Importantly, this form of randomization does not erase the unit's natural choice of decision variable X, as schematically illustrated in Fig. 3.

Building on this idea, we note a natural extension to the agent's capabilities: the ability to intervene on a decision variable X's *perceived* value along a specific path. For concreteness, consider the $\mathcal{L}_3$-quantity known as *natural direct effect*, or NDE, which tracks the effect of X on Y via a "direct" path, as opposed to an "indirect" path via a mediator Z [25]. The NDE is generally regarded as identifiable from experimental data only under certain conditions [26, 8]. The next example illustrates a procedure by which it is possible to compute the NDE even when these identification conditions are not met, by randomizing the *perception* of X.

[1]Note: if the device used for enforcing the value of X is a constant function, this action simply becomes $\textsc{Write}(X : x)^{(i)}$, corresponding to an *atomic* intervention $do(x)$.

Example 1 (Traffic camera) A computer vision company's tool is being considered for an automated speeding ticket system that uses footage from traffic cameras. But the government's audit team has a fairness concern: it is possible the model is trained on footage with a strong correlation between the color of the car and speeding (perhaps due to color preference of different socioeconomic neighbourhoods), and unfairly penalizes certain car colors. This amounts to a hypothesis that X (car's color) affects Y (AI decision to issue a ticket) via a direct path as opposed to the indirect path via Z (speeding), as illustrated in Fig 4(Top). The NDE is defined as the following expression: $\text{NDE}_{x,x'}(y) = P(y_{x' Z_x}) - P(y_x)$ [25].

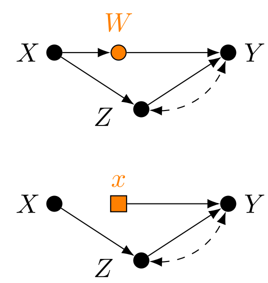


Figure 4: (Top) "Expanded" diagram for *traffic camera* example, where W is *counterfactual mediator* for X; (b) Schematic illustration of randomizing the value of X as perceived by Y.

The second term, $P(y_x)$, can be estimated from a Fisherian randomization of X (say, an experiment recruiting drivers and assigning them random cars). But it is not immediate how to estimate the first term, $P(y_{x' Z_x})$, even with RCT data. However, the audit team recognizes there exists a special mediator, viz. the features W in the video which reveal the car's color to the model (say, RGB values of pixels in the video frames). They use standard video-editing tools to randomly swap the color of the car in the footage. By randomly assigning a particular car $W \leftarrow$ red, they are able to affect the mechanism f_Y's *perception* of X:

$$\begin{aligned}
&P(Y_{W=\text{red}} \mid X = \text{blue}) && \text{estimated by Lemma C.14} && (1)\\
=&P(Y_{W=\text{red},Z} \mid X = \text{blue}) && Z: \text{natural value} && (2)\\
=&P(Y_{W=\text{red},Z_{X=\text{blue}}} \mid X = \text{blue}) && \text{consistency property} && (3)\\
=&P(Y_{X=\text{red},Z_{X=\text{blue}}} \mid X = \text{blue}) && \text{Def. } \textit{D.2}, X \equiv W && (4)\\
=&P(Y_{X=\text{red},Z_{X=\text{blue}}}) && \text{d-separation} && (5)
\end{aligned}$$

Eq. 4 is justified because W controls Y's perception of X given a fixed z (formalized in Lemma D.4). Thus, they are able to directly sample from the $\mathcal{L}_3$-distribution $P(Y_{x' Z_x}, X)$ via a physical procedure, and use identification rules to obtain $P(y_{x' Z_x})$. Using the formula for NDE, they can evaluate whether a car's color has a direct effect on the odds of getting a speeding ticket. A detailed version of this example, including a discussion of assumptions involved, is in App. D.4. ■

One is able to randomize X as *perceived* by only one of its children in Example 1, by leveraging the special variable W (RGB values) that fully encodes information about X (color) and mediates its effect on Y. Other examples of interventions on "perception" of attributes include changing details on a job application (name, pronouns, keywords) to simulate a perceived alternate demographic identity [6], or editing specific portions of text input to a language model [9]. This has also been discussed in [28, §4.4.4]. We provide a formal treatment in App. D.1 of the structural assumptions involved, including Def. D.2 of a *counterfactual mediator*. We also describe another example in App. D.5.

These extensions to experimental capabilities are captured in the following definition of a new physical action that an agent may be able to perform in an environment.

Definition 2.2 (Counterfactual (ctf-) randomization). CTF-RAND$(X \to \mathbf{C})^{(i)}$: fixing the value of X *as an input to* the mechanisms generating $\mathbf{C} \subseteq Ch(X)_{\mathcal{G}}$ using a randomizing device having support over Domain(X), for unit i, given causal diagram $\mathcal{G}$. ■

The essential differences between the Fisherian RAND$(X)^{(i)}$ and CTF-RAND$(X \to \mathbf{C})^{(i)}$ are (1) CTF-RAND does not erase the unit i's natural decision $X^{(i)}$[2]; and (2) while RAND affects all children of X, CTF-RAND does not affect $Ch(X) \setminus \mathbf{C}$. CTF-RAND can only be enacted under certain structural conditions, viz., either in environments which permit the measurement of a unit's natural decision while simultaneously randomizing the actual decision [4], or where counterfactual mediators can be used to alter the value of X as perceived by a subset of children. We provide a systematic way to translate such structural conditions into a list of the CTF-RAND procedures possible in the given environment, in Algo. 4. Ctf-randomization makes it possible to physically perform multiple randomizations involving the same variable X on a single unit i (explained with example in App. D.2). Further, CTF-RAND may only be performed w.r.t a graphical child variable; it is not possible to bypass a child and directly affect a descendant's perception of X (justified in App. D.3).

[2] Another way of understanding this difference is that the unit's natural inclination is taken into account.

3 Counterfactual realizability

Given the ability to perform ctf-randomization, we are interested in knowing which $\mathcal{L}_3$-distributions can be accessed directly by experimentation. In this section, we discuss the physical constraints imposed by nature on an agent. We formally define *realizability* and then develop a complete algorithm to decide whether an $\mathcal{L}_3$-distribution is realizable given a set of feasible physical actions.

The most basic constraints experienced by the agent are physical. Each mechanism $f_V \in \mathcal{F}$ represents some physical process that transforms a unit i according to the laws of nature. For instance, taking a drug, X, produces a side effect in the patient, Y, by a biochemical reaction $f_Y(X, U_Y)$, which depends on the drug and the patient's latent health condition, U_Y. Once patient i has been subjected to mechanism f_Y under $X = x$, there appears to be no way to go back in time and subject the same patient to mechanism f_Y under $X = x'$. Even if technologically feasible to reverse the process (e.g., by taking an antidote to the drug), the latent factors $\mathbf{U} = \mathbf{u}$ might have changed after the experiment (e.g., the patient could have developed tolerance to the drug). Repeating the experiment on this patient is tantamount to testing a *new* unit with unknown latent features $\mathbf{U} = \mathbf{u}'$. [3] This observation can be made more formal through the following assumption.

Assumption 3.1 (Fundamental constraint of experimentation (FCE))**.** A unit i in the target population can physically undergo a causal mechanism $f_V \in \mathcal{F}$ at most once. ■

Remark 3.2. The FCE assumption entails that a unit i can only be submitted to a particular mechanism $f_V(\mathbf{Pa}_V, \mathbf{U}_V)$ under a single set of experimental conditions, received as input to f_V. By implication, the physical actions in Defs. 2.1, 2.2 can only be performed at most once per unit i. ■

Once unit i has been subjected to f_V, it is not possible to re-run f_V with differently fixed inputs. $\textsc{Read}(V)^{(i)}$ thus only yields one value for i. Although ctf-randomization permits multiple interventions involving the same variable X (App. D.2), each such intervention can only be performed once, since it impacts different child mechanisms that can each only occur once for unit i. We also assume that the agent can only perform the physical actions in Defs. 2.1, 2.2, up to isomorphism.

Definition 3.3 (I.i.d sample)**.** Given an $\mathcal{L}_3$-distribution $Q = P(\mathbf{W}_\star)$ and a sequence of physical actions $\mathcal{A}^{(i)}$ performed on unit i in an environment modeled by SCM $\mathcal{M}$, producing a vector of realized values $\mathbf{W}_\star^{(i)} = \mathbf{w}$ for the variables in $\mathbf{W}_\star$, the vector is said to be an *i.i.d sample* from Q if $P^{\mathbb{C}}(\mathbf{W}_\star^{(i)} = \mathbf{w} \mid \mathcal{A}^{(i)}) = P^{\mathcal{M}}(\mathbf{W}_\star = \mathbf{w}), \forall \mathbf{w}$, where $P^{\mathbb{C}}$ is the probability measure over the beliefs of exogenous agent $\mathbb{C}$, and the l.h.s is the probability of physical actions $\mathcal{A}^{(i)}$ producing the vector $\mathbf{w}$ when performed on some unit i. ■

Definition 3.4 (Realizability)**.** Given a causal diagram $\mathcal{G}$ and the set of physical actions $\mathbb{A}$, an $\mathcal{L}_3$-distribution $P(\mathbf{W}_\star)$ is *realizable given* $\mathbb{A}$ *and* $\mathcal{G}$ iff there exists a sequence of actions $\mathcal{A}$ from $\mathbb{A}$ by which an agent can draw an i.i.d sample (Def. 3.3) from $P^{\mathcal{M}}(\mathbf{W}_\star)$, for any $\mathcal{M} \in M(\mathcal{G})$, the class of SCMs compatible with $\mathcal{G}$. ■

We emphasize the *distinction between realizability and identifiability*. Identifiability [24, Def. 3.2.3] from $\mathcal{G}$ states that a distribution (say, $P(\mathbf{v}; do(x))$) can be uniquely computed from the available data (say, $P(\mathbf{v})$) for any SCM compatible with the assumptions in $\mathcal{G}$. Realizability of a distribution states that it is physically possible for an agent to actually gather data samples according to this distribution.

We next develop an algorithm to decide whether a distribution is realizable. For intuition, suppose that an agent is able to perform $\textsc{ctf-Rand}(V \to C), \forall V, C \in Ch(V)$, w.r.t an input causal diagram, and wants to obtain samples from $P(Z_x, W_t)$. Consider the diagram $\mathcal{G}_2$ in Fig. 5. By performing $\textsc{ctf-Rand}(T \to W)$ and $\textsc{ctf-Rand}(X \to Z)$, the distribution is realizable. However, suppose the input diagram is $\mathcal{G}_1$. A necessary condition to measure Z_x for a unit is for mechanism f_A to receive the natural value of T, illustrated in green. While a necessary condition to simultaneously measure W_t is for f_W to receive A_t, which in turn requires f_A to receive a fixed t, shown in red. This conflict in necessary conditions renders the query non-realizable.[4]

[3] In the philosophy of science literature, this has been discussed under the temporal asymmetry of causation [30, §III-IV].

[4] To be clear, the input to the algorithm is a graph and an accurate set of actions the agent can perform in the environment. If the graph is per $\mathcal{G}_1$ in Fig. 5, then $\textsc{ctf-Rand}(T \to Z)$ is not possible in this environment. Marginalizing out A and providing graph $\mathcal{G}_2$ as input does not help. See Remark C.3 and App. D.3.

Algorithm 1 CTF-REALIZE

1: **Input:** $\mathcal{L}_3$-distribution $Q = P(\mathbf{W}_\star)$; causal diagram $\mathcal{G}$; action set $\mathbb{A}$
2: **Output:** I.i.d sample $\mathbf{W}_\star^{(i)}$ from Q; FAIL if Q is not realizable given $\mathcal{G}, \mathbb{A}$
3: Fix a topological ordering Top($\mathcal{G}$)
4: SELECT$^{(i)}$ for a new unit i
5: **for** V in order Top($\mathcal{G}$) **do**
6: INT$_V \leftarrow \emptyset$ {Interventions for V}
7: OUTPUT$_V \leftarrow \emptyset$ {Index in output vector}
8: **for** each term $W_\mathbf{t}$ in expression $\mathbf{W}_\star$ **do**
9: **if** $V \in An(W)_{\mathcal{G}_{\overline{T}}}$ **and** $V \neq W$ **then**
10: Call **COMPATIBLE**($V, W_\mathbf{t}$) 3
11: **end if**
12: **if** $V = W$ **then**
13: Add $\{W_\mathbf{t}\}$ to OUTPUT$_V$
14: **end if**
15: **end for**
16: **for** each {action : tag} $\in$ INT$_V$ **do**
17: Perform the randomization on unit i
18: If the random-generated value $\neq$ tag, discard the unit and return to Line 4
19: **end for**
20: **for** each $W_t \in$ OUTPUT$_V$ **do**
21: **if** $\{$RAND$(V) : .\} \in$ INT$_V$ **then**
22: Return FAIL
23: **else**
24: Perform READ$(V)^{(i)} = v'$
25: Assign v' to each index $W_t^{(i)}$ in output vector $\mathbf{W}_\star^{(i)} = \mathbf{w}$
26: **end if**
27: **end for**
28: **end for**
29: Return i.i.d sample $\mathbf{W}_\star^{(i)} = \mathbf{w}$

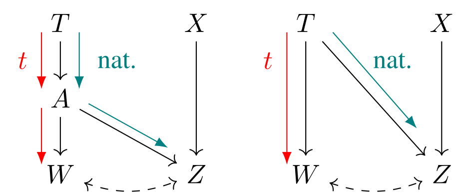


Figure 5: Evaluating realizability of $P(Z_x, W_t)$ for graphs $\mathcal{G}_1$ (left) and $\mathcal{G}_2$ (right). $\mathcal{G}_1$ yields conflicting requirements.

This "edge-coloring" intuition is formalized in Algo. 1. The algorithm **CTF-REALIZE** takes as input an $\mathcal{L}_3$-distribution $P(\mathbf{W}_\star)$, a graph $\mathcal{G}$, and a set of physical actions $\mathbb{A}$ the agent is able to perform in the environment (viz., the RAND and CTF-RAND actions which are possible in the environment). It returns an i.i.d sample if the distribution is realizable, and FAIL otherwise.

The algorithm works as follows. The necessary and sufficient conditions to measure each potential response $W_t \in \mathbf{W}_\star$ are [i] T is fixed as t (by intervention) as an input to all children $C \in Ch(T) \cap An(W)$; [ii] each $A \in An(W)_{\mathcal{G}_{\overline{T}}}$, $A \notin \{T, W\}$ is received "naturally" (i.e., without intervention) by its children $C \in Ch(A) \cap An(W)$; and [iii] f_W is not erased and overwritten (by Fisherian intervention). If these conditions can be met for all the terms in $\mathbf{W}_\star$, the distribution is realizable. If there is a conflict in the necessary conditions for evaluating two terms (as we saw for $P(Z_x, W_t)$ in Fig. 5, $\mathcal{G}_1$), the query is non-realizable.

The algorithm is general and does not make assumptions about the ability to perform any particular interventions. If the action set $\mathbb{A}$ does not contain counterfactual randomization capabilities, the algorithm returns FAIL for non-$\mathcal{L}_2$ queries. If the agent cannot perform any interventions at all, the algorithm returns FAIL for non-$\mathcal{L}_1$ queries (we assume the ability to READ all variables).

Theorem 3.5 (Correctness and Completeness)**.** *An $\mathcal{L}_3$-distribution $Q = P(\mathbf{W}_\star)$ is realizable given action set $\mathbb{A}$ and causal diagram $\mathcal{G}$ iff the algorithm **CTF-REALIZE**$(Q, \mathcal{G}, \mathbb{A})$ returns a sample.* ■

A further question we may ask is which $\mathcal{L}_3$-distributions are realizable if we assume maximum experimental capabilities, notably, the ability to perform separate ctf-randomization for each child of each variable. Given a causal diagram $\mathcal{G}$, we define the *maximal feasible action set* $\mathbb{A}^\dagger(\mathcal{G})$ as the set containing all of the following actions: SELECT$^{(i)}$, READ$(V)^{(i)}$,$\forall V$, and CTF-RAND$(X \to C)^{(i)}$,$\forall X$ and $C \in Ch(X)$. $\mathbb{A}^\dagger(\mathcal{G})$ thus gives the agent the most granular interventional capabilities.

Definition 3.6 (Ancestors of a counterfactual [8])**.** Given a causal diagram $\mathcal{G}$ and a potential response $Y_\mathbf{x}$, the set of (counterfactual) ancestors of $Y_\mathbf{x}$, denoted $An(Y_\mathbf{x})$, consists of each $W_\mathbf{z}$ s.t. $W \in An(Y)_{\mathcal{G}_{\underline{\mathbf{X}}}}$, and $\mathbf{z} = \mathbf{x} \cap An(W)_{\mathcal{G}_{\overline{\mathbf{X}}}}$. For a set $\mathbf{W}_\star$, $An(\mathbf{W}_\star)$ is defined to be the union of the ancestors of each potential response in the set. ■

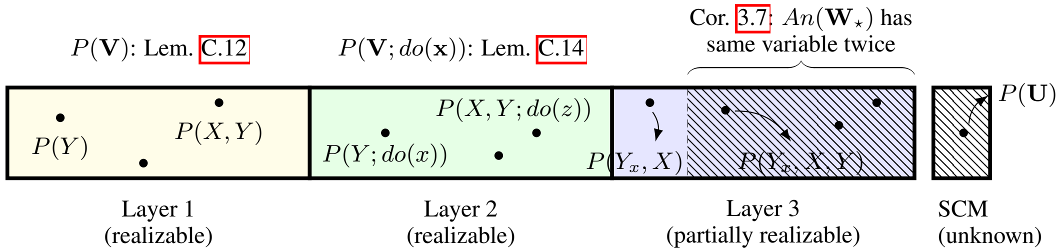


Figure 6: Pearl Causal Hierarchy (PCH) induced by an unknown SCM $\mathcal{M}$. An $\mathcal{L}_3$-distribution is realizable given a graph $\mathcal{G}$ and the maximal feasible action set $\mathbb{A}^\dagger(\mathcal{G})$ iff the ancestor set $An(\mathbf{W}_\star)$ does not contain the same variable under different regimes.

Corollary 3.7. *An $\mathcal{L}_3$-distribution $Q = P(\mathbf{W}_\star)$ is realizable given causal diagram $\mathcal{G}$ and action set $\mathbb{A}^\dagger(\mathcal{G})$ iff the ancestor set $An(\mathbf{W}_\star)$ does not contain a pair of potential responses $W_\mathbf{t}, W_\mathbf{s}$ of the same variable W under different regimes.* ■

For instance, if $\mathbf{W}_\star = \{Z_x, W_t\}$ w.r.t graph $\mathcal{G}_1$ in Fig. 5, then $An(\mathbf{W}_\star) = \{A, T, A_t\}$, which contains both A, A_t. Thus, $P(\mathbf{W}_\star)$ is not realizable even with maximal experimentation capabilities. Cor. 3.7 thus provides a graphical criterion to delineate how far up the PCH an agent can go via experimental methods, as illustrated in Fig. 6. In App. B.2, we provide further examples of using the **CTF-REALIZE** algorithm, and the graphical criterion, to demonstrate the realizability of the ETT distribution $P(Y_x, X)$, the non-realizability of the PS distribution $P(Y_x, X, Y)$.

Corollary 3.8 (Fundamental problem of causal inference (FPCI) [15])**.** *The distribution $Q = P(Y_x, Y_{x'})$ is not realizable given maximal feasible action set $\mathbb{A}^\dagger(\mathcal{G})$, for any causal diagram $\mathcal{G}$, and any variables $X, Y \in Desc(X)$.* ■

The FPCI is an influential notion in the literature, and is often taken as a primitive, or in an axiomatic fashion. We show that it is rather a specific consequence of the more general FCE assumption 3.1, and follows from Thm. 3.5 and Cor. 3.7. By itself, the FPCI does not translate to an operational criterion for determining which $\mathcal{L}_3$-distributions are realizable (Def. 3.4). For instance, it does not clarify that a query with potential responses under different regimes like $P(Y_x, Z_{x'})$ may indeed be realizable via counterfactual randomization, as we show in App. D.5. It also does not tell us that $P(Z_x, W_t)$ may be realizable given causal diagram $\mathcal{G}_2$ in Fig. 5, but not realizable given $\mathcal{G}_1$.

4 Counterfactual decision-making

In this section, we study the practical implications of counterfactual realizability for decision-making. To focus the discussion, we provide a generic *MAB template* (Fig. 7) that is representative of a broad class of multi-arm bandit (MAB) problems in the literature (the discussion can also be extended to other settings such as sequential or Markov decision processes). X is the decision variable, Z is a context variable, Y is the reward, and D is a descendant of X confounded with Y.

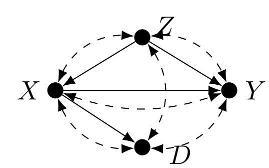


Figure 7: MAB template.

Given a decision problem following the MAB template (Fig. 7), a *decision strategy* π is a mapping from a set of variables $\mathbf{W}_\star$ (possibly counterfactual) to a set of actions $\mathcal{A}$ involving X. The expected reward of following this strategy is notated $\mu_\pi := \mathbb{E}[Y_\mathcal{A} \mid \mathbf{W}_\star]$, where $Y_\mathcal{A}$ is the potential response of Y under the actions $\mathcal{A}$. For example, the $\mathcal{L}_1$-strategy of merely observing the natural behaviour of some behavioral agent is $\pi^{\text{obs}} : \{\} \mapsto \{\}$, which incurs the observational reward of $\mu_{\pi^{\text{obs}}} = \mathbb{E}[Y]$. Whereas, the typical approach in the literature is the $\mathcal{L}_2$-strategy $\pi^{\text{int}} : \{z\} \mapsto \{\text{WRITE}(X : x)\}$, where $x := \arg\max_{x'} \mathbb{E}[Y_{x'} \mid z]$. In words, this strategy involves observing context $Z = z$ for a each round, and then performing the intervention $do(x)$ that maximizes $\mathbb{E}[Y_x \mid z]$ [34, 17].

It was shown in [4, 12, 36] that there exists a superior counterfactual strategy π^{ett} based on based on maximizing $\mathbb{E}[Y_x \mid x', z]$, related to ETT (discussed in Sec. 1). We improve upon this benchmark by proving that the following $\mathcal{L}_3$-strategy is optimal in an MAB problem, $\pi^{\text{opt}} : \{X, Z, D_{x''}\} \mapsto \{\text{CTF-WRITE}(x \to Y), \text{CTF-WRITE}(x'' \to D)\}$, where $x, x'' := \arg\max_{x,x''} \mathbb{E}[Y_x \mid Z, X, D_{x''}]$.[5]

[5] CTF-WRITE is simply the deterministic equivalent of CTF-RAND (Def. 2.2).

With minor abuse of notation, this is the strategy that (1) observes $Z = z, X = x'$ for a round; (2) maps from $\{z, x'\} \mapsto x''$, to perform the counterfactual intervention CTF-WRITE$(x'' \to D)$ to observe $D_{x''} = d$; and (3) maps from $\{z, x', d_{x''}\} \mapsto x$, to perform the counterfactual intervention CTF-WRITE$(x \to Y)$ that maximizes $\mathbb{E}[Y_x \mid z, x', d_{x''}]$. For each mapping in (2), (3) yields a local optimum, in expectation over $X, Z, D_{x''}$. Optimizing over all mappings in (2) yields a global optimum. Translating this to practice, we provide a general algorithm (Algo. 2) that adapts any standard MAB solver to implement the optimal $\mathcal{L}_3$-strategy π^{opt}. We provide examples using Thompson Sampling in the following section (Algos. 5,6).

Theorem 4.1 (Optimality)**.** *Given a decision problem following the MAB template (Fig. 7), π^{opt} is an optimal realizable strategy. I.e., $\mu_{\pi^{opt}} \geq \mu_\pi, \forall \pi \in \Pi$, the space of realizable strategies.* ■

Corollary 4.2 ($\mathcal{L}_3$-dominance)**.** *Given an MAB decision problem with causal diagram described by the MAB template (Fig. 7), the optimal $\mathcal{L}_3$-strategy π^{opt} dominates the $\mathcal{L}_1$-strategy π^{obs} and the optimal $\mathcal{L}_2$-strategy π^{int}. I.e., $\mu_{\pi^{opt}} \geq \mu_{\pi^{obs}}$ and $\mu_{\pi^{opt}} \geq \mu_{\pi^{int}}$.* ■

Algorithm 2 MAB-OPT

```
Input: MAB problem following Fig. 7; MAB solver (e.g. UCB, EXP3, TS); No. of rounds T; Obs. data P(v)
for each z, x' do
    Initialize D-arms x''
end for
for each z, x', x'', d do
    Initialize Y-arms x
    If x = x' = x'', hot-start using P(v)
end for
for t ∈ [T] do
    Observe x', z
    Draw D-arm x'' using MAB solver
    Perform CTF-WRITE(x'' → D) and get D_{x''} = d
    Draw Y-arm x using MAB solver
    Perform CTF-WRITE(x → Y) and get Y_x = y
    Update D-arms and Y-arms according to MAB solver rules using y
end for
```

For decades, the Fisherian RCT methodology used to enact π^{int} was deemed to be the "gold standard" for decision-making. We show that the $\mathcal{L}_3$-strategy π^{opt} is at least as good (and often better) than $\mathcal{L}_2$-strategies. This means that if MAB solvers UCB, EXP3 etc. were deployed to enact an $\mathcal{L}_2$-strategy in an environment where π^{opt} is better, the agent would incur linear cumulative regret, since the learning approach comes no closer to discovering the optimal strategy as the number of trials increases.

4.1 Experiments

We illustrate the benefits of $\mathcal{L}_3$-approaches over purely $\mathcal{L}_2$-ones, through two scenarios. We evaluate four algorithms in simulations: (i) TS: Thompson Sampling optimizing $\mathbb{E}[Y_x \mid \text{context}]$; (ii) TS^{ett}: optimizing the $\mathcal{L}_3$-quantity $\mathbb{E}[Y_x \mid x', \text{context}]$; (iii) TS^{opt}: optimizing the $\mathcal{L}_3$-quantity $\mathbb{E}[Y_x \mid x', d_{x''}]$; and (iv) TS^{aug}: TS that follows an $\mathcal{L}_2$-strategy and treats the natural decision X and $D_{x''}$ as merely additional context variables. (ii) and (iii) are detailed in Alg. 5 and 6.

Experiment 1 (Algorithmic sentencing). Consider a drug court [20], where offenders are assigned to undertake either counseling ($X = 1$) or peer-support ($X = 0$), as shown in Fig. 8(a - Top). The outcome Y is a binary indicator that the individual does not re-offend within 3 months. An experienced judge represents the natural regime ($\mathcal{L}_1$), and an external agent (perhaps a policy-maker) wants to optimize Y. Z is a binary score summarizing the individual's risk (1: high risk), based on a standardized score card. From past observational data, the judge appears to perform well on high-risk individuals ($\mathbb{E}[Y \mid z_1] = 0.75$), but poorly on low-risk ones ($\mathbb{E}[Y \mid z_0] = 0.25$). Data from a recent social science RCT ($\mathcal{L}_2$) shows $\mathbb{E}[Y_{x_1} \mid z] = 0.6, \forall z$. This suggests a naive "mixed" strategy of using the judge for high-risk individuals and the best treatment recommended by the RCT, for low-risk ones. However, following [4, 36], the agent can both sample the judge's natural decision and randomize the treatment per defendant, leading to a strategy based on ETT-like quantity $\mathbb{E}[Y_x \mid x', z]$ that yields best results as shown in Table 1. The optimal strategy turns out to be to follow the judge's natural decision for high-risk individuals, and flip the natural decision for low-risk ones. Details of the underlying SCM and discussions are in App. E.1.

Strategy	**Avg.** Y
$\mathcal{L}_1$	0.35
$\mathcal{L}_2$	0.6
Naive mixed ($\mathcal{L}_1 + \mathcal{L}_2$)	0.63
ETT [4, 36] ($\mathcal{L}_3$)	0.75

Table 1: Performance of different strategies in Experiment 1.

Simulations in the online, adaptive setting corroborate this conclusion. Fig. 8(c - Top) shows the cumulative regret (CR) for 1000 iterations averaged over 200 epochs for TS, TS^{aug}, TS^{ett} (CI=95%), while Fig. 8(d - Top) shows the optimal arm probability (OAP) over time. The $\mathcal{L}_3$-strategy in red

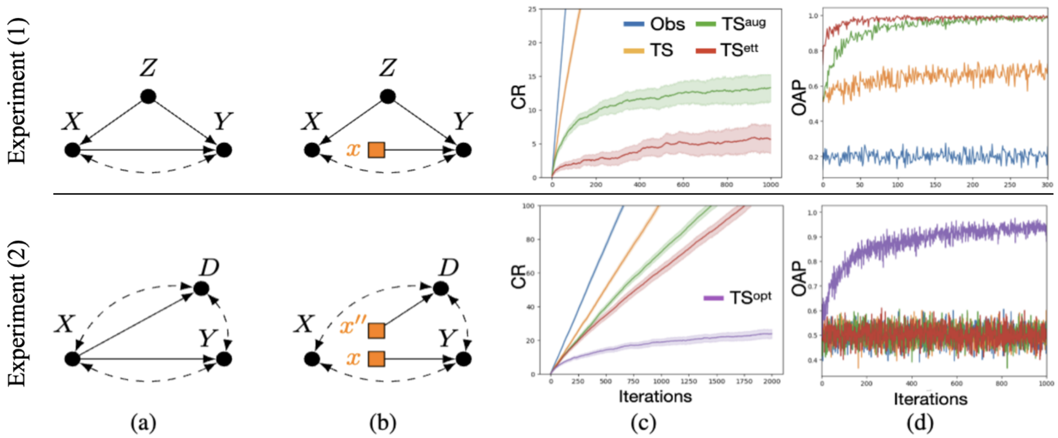


Figure 8: (Top) Experiment 1; (Bottom) Experiment 2; (a) MAB causal diagram; (b) Illustration of the optimal $\mathcal{L}_3$-decision strategy; (c) Cumulative Regret for algorithms TS, TS^{aug}, TS^{ett}, TS^{opt} and Obs. strategy as baseline; (d) Optimal Arm Probability for all algorithms.

shows the quickest convergence and lowest CR. The $\mathcal{L}_2$-strategy in green of using X as merely a context variable, takes around 200 more iterations for OAP to converge.

Experiment 2 (Big Tech surveillance). Alice is a user of a social networking platform run by Omega, a Big Tech firm that uses surveillance and predictions to increase user engagement through addictive notifications and recommendations [38]. Alice chooses every evening whether to use Omega via desktop ($X = 0$) or mobile ($X = 1$). Y is a binary indicator of whether Alice stays within her self-determined social media usage limit per day. Alice also notices that she receives ads when she logs in each evening as D (0: streaming service, 1: food delivery ads). The usage type X affects D, Y, as shown in Fig. 8(a - Bottom). In reality, the decisions, covariates and confounders can be high-dimensional.

Strategy	Avg. Y
$\mathcal{L}_1$	0.65
$\mathcal{L}_2$	0.7
ETT [4, 36] ($\mathcal{L}_3$)	0.75
Optimal $\mathcal{L}_3$ (this work)	0.80

Table 2: Performance of different strategies in Experiment 2.

Omega's PR department claims an RCT shows that the avg. user sticks to their self-determined usage limit 70% of the time. Alice verifies that if she randomizes her daily choice ($\mathcal{L}_2$) she indeed incurs $\mathbb{E}[Y_x] = 0.7, \forall x$. But she suspects that they are tracking and exploiting her latent preferences since she normally experiences $\mathbb{E}[Y] = 0.65$ from following her natural choices ($\mathcal{L}_1$) each day. She then decides to follow an ETT-based strategy ($\mathcal{L}_3$) by recording what she naturally feels like doing each day, and optimizing $\mathbb{E}[Y_x \mid x']$, getting avg. performance of 0.75. At this point, she realizes can perform *another* counterfactual randomization, by sampling her natural choice X, randomly logging in to just see what ads she gets $D_{x''}$, and again randomizing how she actually uses Omega that day to get Y_x. This $\mathcal{L}_3$-strategy optimizes $\mathbb{E}[Y_x \mid x', d_{x''}]$, which performs best as shown in Table 2. Detailed discussion of the underlying SCM is in App. E.2.

Simulations in the online setting corroborate this finding. Fig. 8(c,d - Bottom) shows the cumulative regret (CR) and optimal arm probability (OAP) over 2000 iterations averaged over 200 epochs for TS, TS^{aug}, TS^{ett}, and TS^{opt} (CI=95%). The optimal strategy (purple) performs best, improving on the baseline performance of the ETT-based strategy (red) in [4, 36]. Indeed, all other algorithms fail to improve in OAP, and incur constant average regret in the limit.

5 Conclusions

In this paper, we formalize the ability to draw samples from a distribution by direct experimentation, which we call *realizability*. We develop a complete algorithm to determine whether a counterfactual distribution is realizable given certain fundamental constraints of experimentation (FCE). We build on this to identify an optimal counterfactual decision strategy in a decision problem with a generic template, and demonstrate how it improves upon a previous baseline in counterfactual decision-making. We hope these results can help researchers identify novel experiment-design ideas that permit counterfactual randomization, and thus more powerful, personalized, decision-making strategies.

Acknowledgements

This research is supported in part by the NSF, ONR, AFOSR, DoE, Amazon, JP Morgan, and The Alfred P. Sloan Foundation.

References

[1] Tara V. Anand, Adele H. Ribeiro, Jin Tian, and Elias Bareinboim. Causal effect identification in cluster dags. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i10.26435. URL https://doi.org/10.1609/aaai.v37i10.26435.

[2] Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996. doi: 10.1080/01621459.1996.10476902. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1996.10476902.

[3] Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1172–1176, 9 1997.

[4] Elias Bareinboim, Andrew Forney, and Judea Pearl. Bandits with unobserved confounders: A causal approach. In *Advances in Neural Information Processing Systems*, pages 1342–1350, 2015.

[5] Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. On Pearl's Hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 507–556. Association for Computing Machinery, New York, NY, USA, 1st edition, 2022.

[6] Marianne Bertrand and Sendhil Mullainathan. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. Working Paper 9873, National Bureau of Economic Research, July 2003. URL http://www.nber.org/papers/w9873.

[7] J. Correa and E. Bareinboim. A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY, 2020. AAAI Press.

[8] Juan Correa, Sanghack Lee, and Elias Bareinboim. Nested counterfactual identification from arbitrary surrogate experiments. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 6856–6867. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/36bedb6eb7152f39b16328448942822b-Paper.pdf.

[9] Amir Feder, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar, Zach Wood-Doughty, Jacob Eisenstein, Justin Grimmer, Roi Reichart, Margaret E. Roberts, Brandon M. Stewart, Victor Veitch, and Diyi Yang. Causal inference in natural language processing: Estimation, prediction, interpretation and beyond. *Transactions of the Association for Computational Linguistics*, 10:1138–1158, 2022. doi: 10.1162/tacl_a_00511. URL https://aclanthology.org/2022.tacl-1.66.

[10] Ronald A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925.

[11] Ronald A. Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.

[12] Andrew Forney, Judea Pearl, and Elias Bareinboim. Counterfactual Data-Fusion for Online Reinforcement Learners. In *Proceedings of the 34th International Conference on Machine Learning*, 2017. ISBN 9781510855144. doi: http://dx.doi.org/10.1037/a0022750.

[13] James J. Heckman and Richard Robb Jr. Alternative Methods for Evaluating the Impact of Interventions. In J J Heckman and B Singer, editors, *Longitudinal Analysis of Labor Market Data*. Cambridge University Press, New York, NY, 1985.

[14] James J. Heckman and Richard Robb Jr. Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes. In H. Wainer, editor, *Drawing Inference From Self Selected Samples*, pages 63–107. Springer-Verlag, New York, NY, 1986.

[15] P W Holland. Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396): 945–960, 12 1986.

[16] Duligur Ibeling and Thomas Icard. Probabilistic reasoning across the causal hierarchy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10170–10177, 2020.

[17] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[18] Ang Li and Judea Pearl. Unit selection based on counterfactual logic. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 1793–1799. International Joint Conferences on Artificial Intelligence Organization, 7 2019. doi: 10.24963/ijcai.2019/248. URL https://doi.org/10.24963/ijcai.2019/248.

[19] Ang Li and Judea Pearl. Unit selection with causal diagram. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(5):5765–5772, Jun. 2022. doi: 10.1609/aaai.v36i5.20519. URL https://ojs.aaai.org/index.php/AAAI/article/view/20519.

[20] Douglas B. Marlowe, David S. Festinger, Patricia L. Arabia, Karen L. Dugosh, Kathleen M. Benasutti, Jason R. Croft, and James R. McKay. Adaptive interventions in drug court: A pilot experiment. *Criminal Justice Review*, 33(3):343–360, 2008. doi: 10.1177/0734016808320325.

[21] Robert Nozick. Newcomb's problem and two principles of choice. In Nicholas Rescher, editor, *Essays in Honor of Carl G. Hempel*, pages 114–146. Reidel, 1969.

[22] Yushu Pan and Elias Bareinboim. Counterfactual image editing. Technical Report R-103, Causal Artificial Intelligence Lab, Columbia University, December 2023.

[23] Judea Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*, 121(1–2):93–149, 11 1999.

[24] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition, 2000. ISBN 978-0-521-89560-6.

[25] Judea Pearl. Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pages 411–420. Morgan Kaufmann, San Francisco, CA, 2001.

[26] Judea Pearl. Direct and indirect effects. In *Proceedings of the American Statistical Association, Joint Statistical Meetings*, pages 1572–1581. {MIRA} Digital Publishing, Minn., MN, 2005.

[27] Judea Pearl and Dana Mackenzie. *The Book of Why*. Basic Books, New York, 2018. ISBN 978-0-465-09760-9.

[28] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A Primer*. John Wiley & Sons, 2016. ISBN 9781119186847.

[29] Drago Plecko and Elias Bareinboim. Causal fairness analysis: A causal toolkit for fair machine learning. *Foundations and Trends in Machine Learning*, 17(3):304–589, Jan 2024. ISSN 1935-8237. doi: 10.1561/2200000106. URL https://doi.org/10.1561/2200000106.

[30] Hans Reichenbach. *The Direction of Time*. University of California Press, Berkeley, 1956.

[31] Thomas S. Richardson and James M. Robins. Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Working Paper 128, Center for the Statistics and the Social Sciences, 2013.

[32] Ilya Shpitser. *Complete Identification Methods for Causal Inference*. PhD thesis, Computer Science Department, University of California, Los Angeles, CA, 4 2008.

[33] Ilya Shpitser and Judea Pearl. Effects of Treatment on the Treated: Identification and Generalization. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, Montreal, Quebec, 2009. AUAI Press.

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

[35] Kevin Xia and Elias Bareinboim. Neural causal abstractions. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(18):20585–20595, Mar. 2024. doi: 10.1609/aaai.v38i18.30044. URL https://ojs.aaai.org/index.php/AAAI/article/view/30044.

[36] Junzhe Zhang and Elias Bareinboim. Can humans be out of the loop? In Bernhard Schölkopf, Caroline Uhler, and Kun Zhang, editors, *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 177 of *Proceedings of Machine Learning Research*, pages 1010–1025. PMLR, 11–13 Apr 2022. URL https://proceedings.mlr.press/v177/zhang22a.html.

[37] Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 26548–26558. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/zhang22ab.html.

[38] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. 1st edition, 2018. ISBN 1610395697.

Appendices

A Graphical terminology

Structural Causal Models (SCM) and causal diagrams are described in the preliminaries in Sec. 1. See [5] for full treatment. We use the following graphical kinship nomenclature w.r.t causal diagram $\mathcal{G}$:

- Parent(s) of V, denoted $\mathbf{Pa}_V$: the set of variables $\{V'\}$ s.t. there is a direct edge $V' \to V$ in $\mathcal{G}$. $\mathbf{Pa}_V$ does not include V.
- Children of V, denoted $Ch(V)$: the set of variables $\{V'\}$ s.t. there is a direct edge $V \to V'$ in $\mathcal{G}$. $Ch(V)$ does not include V.
- Ancestors of V, denoted $An(V)$: the set of variables $\{V'\}$ s.t. there is a path (possibly length 0) from V' to V consisting only of edges pointing toward V, $V' \to ... \to V$. $An(V)$ is defined to include V.
- Descendants of V, denoted $Desc(V)$: the set of variables $\{V'\}$ s.t. there is a path (possibly length 0) from V to V' consisting only of edges pointing toward V', $V \to ... \to V'$. $Desc(V)$ is defined to include V.
- Non-descendants of V, denoted $NDesc(V)$: the set $\mathbf{V} \setminus Desc(V)$. $NDesc(V)$ does not include V.

Given a graph $\mathcal{G}$, $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{W}}}$ is the result of removing edges coming into variables in $\mathbf{X}$, and edges coming out of $\mathbf{W}$.

B Details on the CTF-REALIZE algorithm

B.1 Sub-routine of CTF-REALIZE algorithm (Algo. 1)

Algorithm 3 COMPATIBLE (sub-routine)

Input: $V \in \mathbf{V}$ of $\mathcal{G}$; $W_{\mathbf{t}} \in \mathbf{W}_\star$ of Q
for each $C \in Ch(V)$ **do**
  if $C \in An(W)$ **then**
    if $V \in \mathbf{T}$ **then**
      Let $v :=$ value of V in subscript $\mathbf{t}$
      Find smallest $\mathbf{C} \ni C$ s.t. CTF-RAND$(V \to \mathbf{C}) \in \mathbb{A}$
      if $\{$CTF-RAND$(V \to \mathbf{C}) : .\} \in$ INT$_V$ and its label is not "v" **then**
        Return FAIL
      else
        Add $\{$CTF-RAND$(V \to \mathbf{C}) : v\}$ to INT$_V$, with the label "v"
      end if
      if no such $\mathbf{C} \ni C$ s.t. CTF-RAND$(V \to \mathbf{C}) \in \mathbb{A}$ **then**
        if $\{$RAND$(V) : .\} \in$ INT$_V$ and its label is not "v" **then**
          Return FAIL
        else if RAND$(V) \notin \mathbb{A}$ **then**
          Return FAIL
        else
          Add $\{$RAND$(V) : v\}$ to INT$_V$, with the label "v"
        end if
      end if
    end if
    if $V \notin \mathbf{T}$ **then**
      for each $\mathbf{C} \ni C$ s.t. CTF-RAND$(V \to \mathbf{C}) \in \mathbb{A}$ **do**
        if $\{$CTF-RAND$(V \to \mathbf{C}) : .\} \in$ INT$_V$ and its label is not "Natural" **then**
          Return FAIL
        else
          Add $\{$CTF-RAND$(V \to \mathbf{C})$: Natural$\}$ to INT$_V$, with the label "Natural"
        end if
      end for
      if $\{$RAND$(V) : .\} \in$ INT$_V$ and its label is not "Natural" **then**
        Return FAIL
      else if RAND$(V) \in \mathbb{A}$ **then**
        Add $\{$RAND(V) : Natural$\}$ to INT$_V$, with the label "Natural"
      end if
    end if
  end if
end for

B.2 Examples using the CTF-REALIZE algorithm

Example B.1. (ETT realizability)

Query, $Q = P(Y_x, X)$

Graph, $\mathcal{G}$: Fig. 9

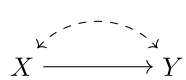


Figure 9: Graph for Example B.1

Suppose action set $\mathbb{A} = \mathbb{A}^\dagger(\mathcal{G}) := \{$CTF-RAND$(X \to Y)\}$

CTF-REALIZE$(Q, \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ trace:

- Start with X (first in topological order)
- For the first term in $\mathbf{W}_\star$: Y_x
  - Since $X \in An(Y)$, call Algo. 3 **COMPATIBLE**(X, Y_x)
    - $Y \in Ch(X)$ and $Y \in An(Y)$
    - $X \in$ subscript of Y_x
    - CTF-RAND$(X \to Y) \in \mathbb{A}^\dagger(\mathcal{G})$
    - INT$_X \leftarrow \{$CTF-RAND$(X \to Y) : x\}$

- For the second term in $\mathbf{W}_\star$: X
  - OUTPUT$_X \leftarrow \{X\}$
- Moving to Y (next in topological order)
- For the first term in $\mathbf{W}_\star$: Y_x
  - OUTPUT$_Y \leftarrow \{Y_x\}$
- Perform interventions in INT$_X$, followed by READ, and assign output vector based on OUTPUT$_X$, OUTPUT$_Y$
- Return i.i.d sample

For simplicity, we don't show the steps $\textsc{Select}^{(i)}$ and the rejection sampling involving in the randomization procedure (steps 17-18 of Algo. 1).

Thus, Q is realizable given $\mathcal{G}, \mathbb{A}^\dagger$. This is validated by the ancestor set $An(Y_x, X)_\mathcal{G} = \{Y_x, X\}$, which doesn't repeat any variables. This is also illustrated in Fig. 10.

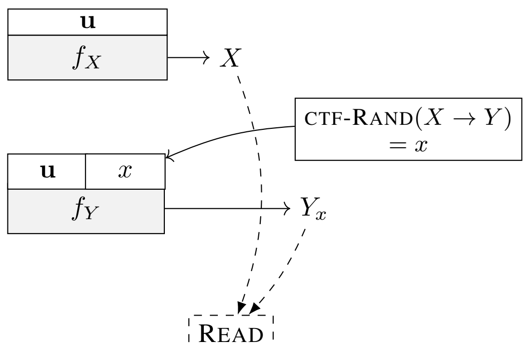


Figure 10: $P(Y_x, X)$ is realizable given the graph in Fig. 9 and $\mathbb{A}^\dagger(\mathcal{G})$.

However, suppose the agent's action set is

$\mathbb{A} = \{\textsc{Rand}(X)\}$, i.e., does not permit any counterfactual randomization procedures.

In this case,

CTF-REALIZE$(Q, \mathcal{G}, \mathbb{A})$ trace:

- Start with X (first in topological order)
- For the first term in $\mathbf{W}_\star$: Y_x
  - Since $X \in An(Y)$, call Algo. 3 **COMPATIBLE**(X, Y_x)
    * $Y \in Ch(X)$ and $Y \in An(Y)$
    * $X \in$ subscript of Y_x
    * $\textsc{Rand}(X) \in \mathbb{A}$, and no other ctf-randomization procedure
    * INT$_X \leftarrow \{\textsc{Rand}(X) : x\}$
- For the second term in $\mathbf{W}_\star$: X
  - OUTPUT$_X \leftarrow \{X\}$
- Moving to Y (next in topological order)
- For the first term in $\mathbf{W}_\star$: Y_x
  - OUTPUT$_Y \leftarrow \{Y_x\}$
- OUTPUT$_X$ contains X, but the intervention set INT$_X$ contains $\textsc{Rand}(X)$

- FAIL (Line 22 of Algo. 1)

■

Example B.2. (Probability of sufficiency (PS) realizability)

Query, $Q = P(Y_x, X, Y)$

Graph, $\mathcal{G}$: Fig. 11

$$X \longrightarrow Y$$

Figure 11: Graph for Example B.2

Suppose action set $\mathbb{A} = \mathbb{A}^\dagger(\mathcal{G}) := \{\textsc{ctf-Rand}(X \to Y)\}$

CTF-REALIZE$(Q, \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ trace:

- Start with X (first in topological order)
- For the first term in $\mathbf{W}_\star$: Y_x
  - Since $X \in An(Y)$, call Algo. 3 **COMPATIBLE**(X, Y_x)
    - $*$ $Y \in Ch(X)$ and $Y \in An(Y)$
    - $*$ $X \in$ subscript of Y_x
    - $*$ $\textsc{ctf-Rand}(X \to Y) \in \mathbb{A}^\dagger(\mathcal{G})$
    - $*$ $\text{INT}_X \leftarrow \{\textsc{ctf-Rand}(X \to Y) : x\}$
- For the second term in $\mathbf{W}_\star$: X
  - $\text{OUTPUT}_X \leftarrow \{X\}$
- For the third term in $\mathbf{W}_\star$: Y
  - Since $X \in An(Y)$, call Algo. 3 **COMPATIBLE**(X, Y)
    - $*$ $Y \in Ch(X)$ and $Y \in An(Y)$
    - $*$ $X \notin$ subscript of Y ; X needs to be received naturally
    - $*$ But INT_X already contains $\{\textsc{ctf-Rand}(X \to Y) : x\}$ with label $x \neq$ "Natural"
    - $*$ FAIL (Line 25 of Algo. 3)

Thus, Q is not realizable given $\mathcal{G}, \mathbb{A}^\dagger$. This is validated by the ancestor set $An(Y_x, X, Y)_\mathcal{G} = \{Y_x, X, Y\}$, which contains both Y_x, Y. This is also illustrated in Fig. 12.

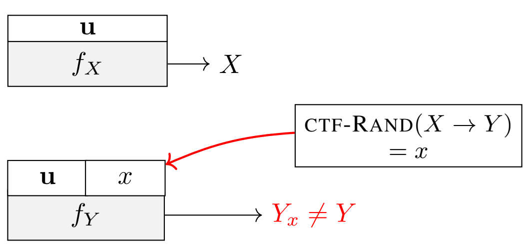


Figure 12: $P(Y_x, X, Y)$ is not realizable given the graph in Fig. 11 and $\mathbb{A}^\dagger(\mathcal{G})$.

■

Example B.3. Query, $Q = P(W_{xt}, Z_{x'})$

Graph, $\mathcal{G}$: Fig. 13

Suppose action set $\mathbb{A} = \mathbb{A}^\dagger(\mathcal{G}) := \{\textsc{ctf-Rand}(T \to A), \textsc{ctf-Rand}(X \to A), \textsc{ctf-Rand}(A \to W), \textsc{ctf-Rand}(A \to Z)\}$

CTF-REALIZE$(Q, \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ trace:

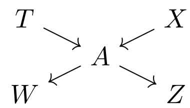


Figure 13: Graph for Example B.3

- Start with X (first in topological order)
- For the first term in $\mathbf{W}_\star$: W_{xt}
  - Since $X \in An(W)$, call Algo. 3 **COMPATIBLE**(X, W_{xt})
    - $*$ $A \in Ch(X)$ and $A \in An(W)$
    - $*$ $X \in$ subscript of W_{xt}
    - $*$ CTF-RAND$(X \to A) \in \mathbb{A}^\dagger(\mathcal{G})$
    - $*$ INT$_X \leftarrow \{$CTF-RAND$(X \to A) : x\}$
- For the second term in $\mathbf{W}_\star$: $Z_{x'}$
  - Since $X \in An(Z)$, call Algo. 3 **COMPATIBLE**$(X, Z_{x'})$
    - $*$ $A \in Ch(X)$ and $A \in An(Z)$
    - $*$ $X \in$ subscript of $Z_{x'}$; X needs to be fixed as x'
    - $*$ But INT$_X$ already contains $\{$CTF-RAND$(X \to A) : x\}$ with label $x \neq x'$
    - $*$ FAIL (Line 8 of Algo. 3)

Thus, Q is not realizable given $\mathcal{G}, \mathbb{A}^\dagger$. This is validated by the ancestor set $An(W_{xt}, Z_{x'})_\mathcal{G} = \{W_{xt}, A_{xt}, Z_{x'}, A_{x'}\}$, which contains both $A_{xt}, A_{x'}$. This is also illustrated in Fig. 14.

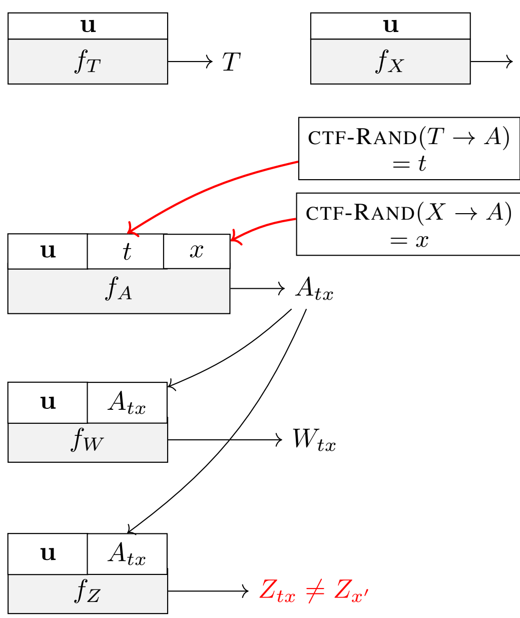


Figure 14: $P(W_{xt}, Z_{x'})$ is not realizable given the graph in Fig. 13 and $\mathbb{A}^\dagger(\mathcal{G})$.

■

Example B.4. Query, $Q = P(Y_x, Z_{x'}, W_{x''})$

Graph, $\mathcal{G}$: Fig. 15

Suppose action set $\mathbb{A} = \{$RAND(X), CTF-RAND$(X \to \{Z, W\})\}$

CTF-REALIZE$(Q, \mathcal{G}, \mathbb{A})$ trace:

$$\begin{tikzcd} & X \arrow[dl] \arrow[d] \arrow[dr] & \\ Y & Z & W \end{tikzcd}$$

Figure 15: Graph for Example B.4

- Start with X (first in topological order)
- For the first term in $\mathbf{W}_\star$: Y_x
  - Since $X \in An(Y)$, call Algo. 3 **COMPATIBLE**(X, Y_x)
    * $Y \in Ch(X)$ and $Y \in An(Y)$
    * $X \in$ subscript of Y_x
    * $\textsc{Rand}(X) \in \mathbb{A}$; no other ctf-randomization procedures affecting Y
    * $\text{INT}_X \leftarrow \{\textsc{Rand}(X) : x\}$
- For the second term in $\mathbf{W}_\star$: $Z_{x'}$
  - Since $X \in An(Z)$, call Algo. 3 **COMPATIBLE**$(X, Z_{x'})$
    * $Z \in Ch(X)$ and $Z \in An(Z)$
    * $X \in$ subscript of $Z_{x'}$
    * $\textsc{ctf-Rand}(X \to \{Z, W\}) \in \mathbb{A}$; no smaller ctf-randomization procedures affecting Z
    * $\text{INT}_X \leftarrow \text{INT}_X \cup \{\textsc{ctf-Rand}(X \to \{Z, W\}) : x'\}$
- For the third term in $\mathbf{W}_\star$: $W_{x''}$
  - Since $X \in An(W)$, call Algo. 3 **COMPATIBLE**$(X, W_{x''})$
    * $W \in Ch(X)$ and $W \in An(W)$
    * $X \in$ subscript of $W_{x''}$; X needs to be fixed as x''
    * But INT_X already contains $\{\textsc{ctf-Rand}(X \to \{Z, W\}) : x'\}$ with label $x' \neq x''$
    * No smaller ctf-randomization procedures affecting W
    * FAIL (Line 8 of Algo. 3)

Thus, Q is not realizable given $\mathcal{G}, \mathbb{A}$.

However, suppose instead that action set $\mathbb{A}' = \{\textsc{Rand}(X), \textsc{ctf-Rand}(X \to \{Z, W\}), \textsc{ctf-Rand}(X \to Z)\}$

CTF-REALIZE$(Q, \mathcal{G}, \mathbb{A}')$ trace:

- Start with X (first in topological order)
- For the first term in $\mathbf{W}_\star$: Y_x
  - Since $X \in An(Y)$, call Algo. 3 **COMPATIBLE**(X, Y_x)
    * $Y \in Ch(X)$ and $Y \in An(Y)$
    * $X \in$ subscript of Y_x
    * $\textsc{Rand}(X) \in \mathbb{A}$; no other ctf-randomization procedures affecting Y
    * $\text{INT}_X \leftarrow \{\textsc{Rand}(X) : x\}$
- For the second term in $\mathbf{W}_\star$: $Z_{x'}$
  - Since $X \in An(Z)$, call Algo. 3 **COMPATIBLE**$(X, Z_{x'})$
    * $Z \in Ch(X)$ and $Z \in An(Z)$
    * $X \in$ subscript of $Z_{x'}$
    * $\textsc{ctf-Rand}(X \to Z) \in \mathbb{A}$
    * $\text{INT}_X \leftarrow \text{INT}_X \cup \{\textsc{ctf-Rand}(X \to Z) : x'\}$
- For the third term in $\mathbf{W}_\star$: $W_{x''}$

- Since $X \in An(W)$, call Algo. 3 **COMPATIBLE**$(X, W_{x''})$
  - $W \in Ch(X)$ and $W \in An(W)$
  - $X \in$ subscript of $W_{x''}$
  - $\textsc{ctf-Rand}(X \to \{Z, W\}) \in \mathbb{A}$
  - $\text{INT}_X \leftarrow \text{INT}_X \cup \{\textsc{ctf-Rand}(X \to \{Z, W\}) : x''\}$

- Moving to Y (next in topological order)
  - $\text{OUTPUT}_Y \leftarrow \{Y_x\}$
- Moving to Z (next in topological order)
  - $\text{OUTPUT}_Z \leftarrow \{Z_{x'}\}$
- Moving to W (next in topological order)
  - $\text{OUTPUT}_W \leftarrow \{W_{x''}\}$
- Perform interventions in INT_X, followed by READ, and assign output vector based on OUTPUT_Y, OUTPUT_Z, OUTPUT_W
- Return i.i.d sample

Thus, Q is realizable given $\mathcal{G}, \mathbb{A}'$.

Lastly, it is evident that Q is realizable given $\mathcal{G}, \mathbb{A}^\dagger$. This is validated by the ancestor set $An(Y_x, Z_{x'}, W_{x''})_\mathcal{G} = \{Y_x, Z_{x'}, W_{x''}\}$, which does not repeat any variables. This is also illustrated in Fig. 16.

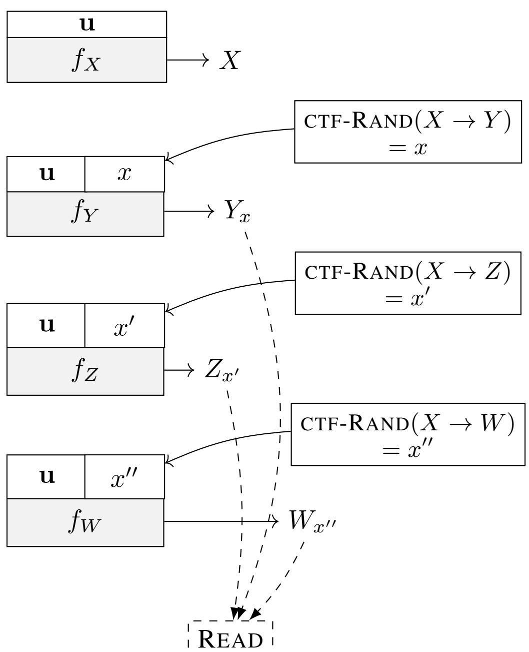


Figure 16: $P(Y_x, Z_{x'}, W_{x''})$ is realizable given the graph in Fig. 15 and $\mathbb{A}^\dagger(\mathcal{G})$.

■

C Structural assumptions and proofs

C.1 Structural assumptions

In this section, we gather together all the structural assumptions we make, for ease of reference. We also include related remarks.

Assumption C.1 (Unobservability)**.** An agent deployed in the environment does not know the underlying SCM $\mathcal{M}$ of the environment, and does not know the latent features $\mathbf{U}^{(i)}$ of any unit i in the target population. ■

Assumption C.2 (Feasible actions)**.** Given causal diagram $\mathcal{G}$, the physical actions that an agent can perform on any unit i in the target population are limited to: $\text{SELECT}^{(i)}$, $\text{READ}(V)^{(i)}$, $\text{RAND}(X)^{(i)}$, and $\text{CTF-RAND}(X \to \mathbf{C})^{(i)}$, for some $V, X \in \mathbf{V}$ and $\mathbf{C} \subseteq Ch(X)_{\mathcal{G}}$, per Defs. 2.1,2.2. ■

Assumption 3.1 (Fundamental constraint of experimentation (FCE))**.** A unit i in the target population can physically undergo a causal mechanism $f_V \in \mathcal{F}$ at most once.app ■

Regarding the structural conditions involving counterfactual randomization (Def. 2.2), we make the following assumption.

Assumption D.3 (Tree structure)**.** Given a variable X, causal diagram $\mathcal{G}$, and an "expanded" diagram $\mathcal{G}^+$ (Def. D.1) including the set of all the counterfactual mediators $\mathbf{W}$ (Def. D.2) of X in the environment, each $W \in \mathbf{W}$ has only one parent in $\mathcal{G}^+$, and each $C \in Ch(X)_{\mathcal{G}}$ has at most one $W \in \mathbf{W}$ as a parent in $\mathcal{G}^+$. ■

From this assumption, and from the definition of a counterfactual mediator (Def. D.2), we can derive the following observations:

Remark C.3 (No bypassing children). Given causal diagram $\mathcal{G}$, the procedure $\text{CTF-RAND}(X \to \mathbf{C})$, either by eliciting a unit's natural decision or via a counterfactual mediator, can only be performed w.r.t $\mathbf{C} \subseteq Ch(X)_{\mathcal{G}}$. It cannot by-pass child mechanisms and directly affect a descendant. This is elaborated in App. D.3, and specifically in Lemma D.7. ■

Remark D.6 (Procedure containment). Assumption D.3 implies that if an agent is capable of performing both $\text{CTF-RAND}(X \to \mathbf{C})^{(i)}$ and $\text{CTF-RAND}(X \to \mathbf{C}')^{(i)}$ s.t. $\mathbf{C} \neq \mathbf{C}'$ and $\mathbf{C} \cap \mathbf{C}' \neq \emptyset$, then either $\mathbf{C} \subseteq \mathbf{C}'$ or $\mathbf{C}' \subseteq \mathbf{C}$. ■

Remark D.5 (Superseding action). Given a decision variable X, the action $\text{CTF-RAND}(X \to \mathbf{C}')^{(i)}$ can *supersede* the action $\text{CTF-RAND}(X \to \mathbf{C})^{(i)}$ if $\mathbf{C}' \subsetneq \mathbf{C}$, where *supersede* means that the former action $\text{CTF-RAND}(X \to \mathbf{C}')^{(i)}$ blocks any effect that the latter action has on the variables $\mathbf{C}'$. Additionally, the action $\text{CTF-RAND}(X \to \mathbf{C})^{(i)}$ *supersedes* the action $\text{RAND}(X)^{(i)}$. ■

Counterfactual randomization permits multiple randomizations for the same variable X for a single unit i. But some randomizations block the effects of others. See App. D.2.

C.2 Proofs for Section 3

Recall, $P^{\mathbb{C}}(.)$ is the probability measure from the perspective of an exogenous agent $\mathbb{C}$'s beliefs about the environment, distinguished by superscript from $P^{\mathcal{M}}(.)$, the true unknown distribution.

Since unit selection is randomized, $\text{SELECT}^{(i)}$ yields an unbiased sample of a unit with latent features distributed according to the target population frequency $P(\mathbf{u})$. I.e., $P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \text{SELECT}^{(i)}) = P^{\mathcal{M}}(\mathbf{u})$.

Lemma C.4 (I.i.d requirement)**.** *Consider a sequence of actions $\mathcal{A}^{(i)}$ performed on unit i in the target population, that yields a vector of realized values $\mathbf{W}_\star^{(i)}$. $\mathbf{W}_\star^{(i)}$ is an i.i.d sample from $P^{\mathcal{M}}(\mathbf{W}_\star)$, for arbitrary $\mathcal{M}$ iff* *i.* $P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \mathcal{A}^{(i)}) = P^{\mathcal{M}}(\mathbf{U} = \mathbf{u})$*; and*

ii. $\mathbb{1}[\mathbf{W}_\star^{(i)} = \mathbf{w} \mid \mathcal{A}^{(i)}, \mathbf{U}^{(i)} = \mathbf{u}] = \mathbb{1}[\mathbf{W}_\star(\mathbf{u}) = \mathbf{w}]$.

Proof. Recall from Def. 3.3 that $\mathbf{W}_\star^{(i)}$ being an i.i.d sample from $P^{\mathcal{M}}(\mathbf{W}_\star)$ means that

$$P^{\mathbb{C}}(\mathbf{W}_\star^{(i)} = \mathbf{w} \mid \mathcal{A}^{(i)}) = P^{\mathcal{M}}(\mathbf{W}_\star = \mathbf{w}), \forall \mathbf{w} \tag{6}$$

Reverse direction:

We simply multiply respective l.h.s and r.h.s of conditions [i] and [ii] and sum over all $\mathbf{u}$ to get

$$\sum_{\mathbf{u}} P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \mathcal{A}^{(i)}).\mathbb{1}[\mathbf{W}_\star^{(i)} = \mathbf{w} \mid \mathcal{A}^{(i)}, \mathbf{U}^{(i)} = \mathbf{u}] = \sum_{\mathbf{u}} P^{\mathcal{M}}(\mathbf{U} = \mathbf{u}).\mathbb{1}[\mathbf{W}_\star(\mathbf{u}) = \mathbf{w}] \tag{7}$$

$$= P^{\mathcal{M}}(\mathbf{W}_\star = \mathbf{w}), \tag{8}$$

which we get from the Layer 3 valuation formula (see preliminaries in Sec. 1). On the l.h.s, we apply the chain rule to get the result we need.

$$P^{\mathbb{C}}(\mathbf{W}_\star^{(i)} = \mathbf{w} \mid \mathcal{A}^{(i)}) = P^{\mathcal{M}}(\mathbf{W}_\star = \mathbf{w}) \tag{9}$$

Forward direction:

Assume Eq. 6.

From Remark C.11, we conclude that condition [i] automatically holds. Since the agent acts exogenously to the system, $P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \mathcal{A}^{(i)}) = P^{\mathcal{M}}(\mathbf{U} = \mathbf{u}), \forall \mathbf{u}$.

Applying the chain rule on both sides of Eq. 6,

$$\sum_{\mathbf{u}} P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \mathcal{A}^{(i)}).\mathbb{1}[\mathbf{W}_\star^{(i)} = \mathbf{w} \mid \mathcal{A}^{(i)}, \mathbf{U}^{(i)} = \mathbf{u}] = \sum_{\mathbf{u}} P^{\mathcal{M}}(\mathbf{U} = \mathbf{u}).\mathbb{1}[\mathbf{W}_\star(\mathbf{u}) = \mathbf{w}] \tag{10}$$

The probability terms are equal, for each $\mathbf{u}$.

Since the probability terms are free parameters, and we need this equation to hold for any arbitrary probability simplex, it must be the case that the indicator terms are also equal. Thus begetting condition [ii].

■

Lemma C.5. *Given a causal diagram $\mathcal{G}$, for any SCM $\mathcal{M}$ compatible with $\mathcal{G}$, the jointly necessary and sufficient conditions to measure a potential response $W_{\mathbf{t}}(\mathbf{u})$ are [i] $\mathbf{T}$ is fixed as $\mathbf{t}$ (by intervention) as an input to all children $C \in Ch(\mathbf{T}) \cap An(W)$; [ii] each $A \in An(W)_{\mathcal{G}_{\overline{\mathbf{T}}}}$, $A \notin \{\mathbf{T}, W\}$ is received "naturally" (i.e., without intervention) by its children $C \in Ch(A) \cap An(W)$; and [iii] the mechanism f_W is not erased and overwritten (by a Fisherian intervention).*

Proof. $W_{\mathbf{t}}$ is the variable W evaluated in the sub-model $\mathcal{M}_{\mathbf{t}}$, where the equations for $\mathbf{T}$ are replaced by constant values in $\mathbf{t}$.

For any changes to the function for $T \in \mathbf{T}$, the function f_W is only affected by any effect on the children of T which are also ancestors of W. Any effect of T on some $C' \in Ch(T)$ s.t. $C' \notin An(W)$ has no effect on W.

Further, in the submodel $\mathcal{M}_{\mathbf{t}}$ there are no interventions on any other ancestors of W in $\mathcal{G}_{\overline{\mathbf{T}}}$, besides $\mathbf{T}$. Even if there were tnterventions involving some $X \notin An(W)_{\mathcal{G}_{\overline{\mathbf{T}}}}$, this would have no effect on W in the sub-model $\mathcal{M}_{\mathbf{t}}$, by Rule 3 of do-calculus.

It is evident that f_W evaluated according to the sub-model $\mathcal{M}_{\mathbf{t}}$, and evaluated according to a sub-model satisfying conditions [i] and [ii] are identical for each $\mathbf{u}$, since the sequence of structural equations that eventually generate W are the same.

Finally, in order to measure $W_{\mathbf{t}}$, we need to measure the output of the mechanism f_W in the real world. The mechanism cannot not be erased and overwritten, as per condition [iii]. ■

Lemma C.6. *Given a set $\mathbf{W}_\star$ and graph $\mathcal{G}$, where each member $W_\mathbf{t} \in \mathbf{W}_\star$ has its respective conditions [i-iii] (per Lemma C.5), suppose these conditions introduce conflicts when combined across $\mathbf{W}_\star$. Removing X from $\mathbf{W}_\star$ and from all subscripts in $\mathbf{W}_\star$ to get a new set $\mathbf{W}'_\star$ does not introduce new conflicts between the terms, if X is first in a topological ordering of $\mathcal{G}$.*

Proof. Let us consider the conditions in the necessary-and-sufficient set given in Lemma C.5 for each $W_\mathbf{t} \in \mathbf{W}_\star$.

Condition [i] add requirements for each $t \in \mathbf{t}$ of some $W_\mathbf{t} \in \mathbf{W}_\star$. Since X is removed from every subscript, this no longer applies to any term in $\mathbf{W}'_\star$.

Condition [ii] requires that if X is an ancestor of some $W_\mathbf{t} \in \mathbf{W}_\star$, and it doesn't appear in $\mathbf{T}$, then X must be received without intervention by mediating children. Since X is removed from every subscript, X not being intervened upon at all meets this condition [ii] for every term $\mathbf{W}'_\star$, without conflicting with condition [i], which no longer applies.

Importantly, even though X is no longer being intervened upon, for each $W_\mathbf{t} \in \mathbf{W}_\star$, removing X from $\mathbf{T}$ (if it appears) does not add any additional ancestors in $\mathcal{G}_{\overline{T}}$ that need to be tracked for condition [ii], since X is first in topological order.

Condition [iii] would only apply to X itself, which is not present as a potential response in $\mathbf{W}'_\star$. ■

Theorem 3.5 :

Let $\mathcal{A}^{(i)}$ be a sequence of actions conducted by an exogenous agent to beget a vector of values $\mathbf{W}_\star^{(i)}$ for a unit i.

By Lemma C.4, if an agent wants $\mathbf{W}_\star^{(i)}$ to be an i.i.d sample from $P(\mathbf{W}_\star)$, then for each possible $\mathbf{U}^{(i)} = \mathbf{u}$, the vector $\mathbf{W}_\star^{(i)}$ needs to be identical to the $\mathbf{W}_\star(\mathbf{u})$ as evaluated according to the SCM. Essentially, this says that the agent's actions need to output the same vector $\mathbf{W}_\star(\mathbf{u})$ as if it has been evaluated according to the SCM.

By Lemma C.5, $\mathbf{W}_\star(\mathbf{u})$ can be evaluated if and only if the following three conditions are met for each $W_\mathbf{t} \in \mathbf{W}_\star$ simultaneously:

i $\mathbf{T}$ is fixed as $\mathbf{t}$ (by intervention) as an input to all children $C \in Ch(\mathbf{T}) \cap An(W)$;

ii Each $A \in An(W)_{\mathcal{G}_{\overline{\mathbf{T}}}}$, $A \notin \{\mathbf{T}, W\}$ is received "naturally" (i.e., without intervention) by its children $C \in Ch(A) \cap An(W)$; and

iii The mechanism f_W is not erased and overwritten (by a Fisherian intervention).

Inductive hypothesis (IH):

CTF-REALIZE$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A})$ returns FAIL if and only if conditions [i-iii] are not met simultaneously, when combined across all $W_\mathbf{t} \in \mathbf{W}_\star$ w.r.t a causal diagram $\mathcal{G}$ having $\leq n$ nodes

Base case:

Consider an SCM with only one variable $V \in \mathbf{V}$. IH is trivially true, since the conditions are always met, and since **CTF-REALIZE** will just return the value $\text{READ}(V)$.

Assume IH is true for any SCM with causal diagram having $\leq n$ nodes.

n+1 case:

Consider an SCM whose causal diagram $\mathcal{G}$ has $n+1$ nodes. Let X be the first in some topological ordering of $\mathcal{G}$. Consider an action set $\mathbb{A}$ that the agent can perform in the environment, and an arbitrary distribution $P(\mathbf{W}_\star)$.

WLOG, we can begin the outer loop of **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}))$ with X (first in topological order).

- The inner loop calls **COMPATIBLE**$(X, W_\mathbf{t})$ for each $W \in Desc(X), X \neq W$.

- It maintains a tracker INT_X of the smallest counterfactual interventions needed to satisfy condition [i] for each $W_{\mathbf{t}}$, resorting to Fisherian intervention if needed. Note (per Remark D.6), interventions follow a tree-like structure, so conflicts can be tracked by tagging the smallest available intervention that is needed for each $W_{\mathbf{t}}$ w.r.t each child of X.
- Note also (per Remark D.5) that if there are two simultaneous interventions added to INT_X, $\text{CTF-RAND}(X \to \mathbf{C})$, $\text{CTF-RAND}(X \to \mathbf{C}')$, where $\mathbf{C}' \subseteq \mathbf{C}$, then the set $\mathbf{C}'$ is unaffected by the first procedure.
- This inner loop exactly checks if there are any conflicts in conditions [i-ii] among $\mathbf{W}_\star$ w.r.t X, by "tagging" each procedure with the fixed value x needed for that intervention (including the requirement of no intervention).
- Finally the outer loop in Line 20 of Algo. 1 checks if X appears as a potential response anywhere in $\mathbf{W}_\star$. If so, INT_X cannot contain the requirement of Fisherian $\text{RAND}(X)$, since this violates condition [iii] w.r.t X.
- X does not appear anywhere else in subsequent algorithm iterations.

Thus, we conclude that **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}))$ does not return FAIL on the outer loops evaluated for X, if and only if there are no conflicts in the conditions [i-iii] for $\mathbf{W}_\star$ w.r.t X. *In other words, all conflicts w.r.t X, in the conditions [i-iii] combined across the terms in $\mathbf{W}_\star$, are identified in the algorithm steps that involve X.*

Next, define the new set $\mathbf{W}'_\star$ by dropping x from the subscript (if it appears) for each $W_{\mathbf{t}} \in \mathbf{W}_\star$, and dropping X from $\mathbf{W}_\star$ (if it appears). Since X is first in topological order of $\mathcal{G}$, this does not add any *new* conflicts across conditions [i-iii] induced by each term in $\mathbf{W}'_\star$ (by Lemma C.6).

It is also clear that if there are conflicts *not* involving X, that are induced by conditions [i-iii] across the terms in $\mathbf{W}_\star$, then these conflicts are also induced by $\mathbf{W}'_\star$. Suppose there are two terms $W_{\mathbf{t}}, Y_{\mathbf{h}} \in \mathbf{W}_\star$ s.t. $\mathbf{T} \setminus X$ needs to be received as $\mathbf{t} \setminus X$ by mediating children (condition [i]) for $W_{\mathbf{t}}$, and this conflicts with the requirement that $\mathbf{T} \setminus X$ needs to be received as $\mathbf{t}' \setminus X$ (or naturally) by the same mediating children, for $Y_{\mathbf{h}}$. Removing X does not affect this conflict, since X is topologically prior.

Next, define the graph $\mathcal{G}'$ as the projection of $\mathcal{G}$ that marginalizes out X (and adds bidirected edges if needed). $\mathcal{G}'$ has $\leq n$ nodes. Therefore, from the IH, we conclude that **CTF-REALIZE**$(P(\mathbf{W}'_\star), \mathcal{G}', \mathbb{A})$ does not return FAIL if and only if there are no conflicts induced by conditions [i-iii], combined across terms in $\mathbf{W}'_\star$.

Now, we note that **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A})$ is merely **CTF-REALIZE**$(P(\mathbf{W}'_\star), \mathcal{G}', \mathbb{A})$, plus all the steps involving X that we discussed earlier (can be verified from inspecting the algorithm - the former has an outer loop involving X and then contains the same steps as the latter). *Therefore, all conflicts induced by conditions [i-iii] that involve X and do not involve X are identified in the algorithm steps when run on $\mathcal{G}$ and $\mathbf{W}_\star$.*

Thus, we show that **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A})$ returns FAIL if and only if conditions [i-iii] are not met simultaneously, when combined across all $W_{\mathbf{t}} \in \mathbf{W}_\star$ w.r.t a causal diagram having $\leq n+1$ nodes. The IH stands proved.

By Lemma C.5, we know that conditions [i-iii] are necessary and sufficient to evaluate each term in $\mathbf{W}_\star(\mathbf{u})$ simultaneously, for any SCM compatible with $\mathcal{G}$. By Lemma C.4, we know that this is equivalent to drawing an i.i.d sample from $P(\mathbf{W}_\star)$. This gives us the proof of the theorem.

Note: we don't discuss the rejection sampling steps involved steps 17-18 of Algo. 1 as this is trivially equivalent to intervening using a fixed value. ■

Corollary 3.7 :

The proof intuition is as follows: given a graph $\mathcal{G}$ and a potential response $Y_{\mathbf{x}}$, the set of (counterfactual) ancestors of $Y_{\mathbf{x}}$ [8] lists each ancestor of Y and *what regime it must be realized in*, in order for $Y_{\mathbf{x}}$ to be evaluated. In other words $An(Y_{\mathbf{x}})$ tracks the regimes necessary and sufficient for its ancestors to be evaluated under to beget $Y_{\mathbf{x}}$.

For instance, in graph $\mathcal{G}_1$ in Fig. 5, in order to evaluate W_t, we need A_t to be evaluated in the regime $\mathcal{M}_t$. In order to evaluate Z_x, we need A, T to both be evaluated naturally. This reveals a conflict at the bottleneck f_A, which renders the distribution non-realizable.

Thus, Corollary 3.7 provides a *sufficient* condition to conclude that a distribution is non-realizable, if $An(\mathbf{W}_\star)$ contains two potential responses of the same variable under different regimes. It also becomes a *necessary* condition for non-realizability, if the agent can perform CTF-RAND$(X \to C)$, separately for each $C \in Ch(X)$, for all X. I.e., if the action set is $\mathbb{A}^\dagger(\mathcal{G})$.

The proof steps are similar to Theorem 3.5.

Inductive Hypothesis (IH):

Given a graph $\mathcal{G}$ with $\leq \ n$ nodes, and an arbitrary distribution $\mathbf{W}_\star$, **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ if and only if $An(\mathbf{W}_\star)$ does not contain a pair of potential responses $W_\mathbf{t}, W_\mathbf{s}$ of the same variable W under different regimes.

Base case:

For a graph containing only one variable Y, this is trivially true. $An(Y) = Y$, and the distribution $P(Y)$ is realizable.

Assume IH is true for a graph of $\leq n$ nodes.

n+1 case:

Consider an SCM whose causal diagram $\mathcal{G}$ has $n+1$ nodes. Let X be the first in some topological ordering of $\mathcal{G}$. The agent can perform $\mathbb{A}^\dagger$ in the environment, and the distribution is some arbitrary $P(\mathbf{W}_\star)$. WLOG, we can begin the outer loop of **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}))$ with X (first in topological order).

From Lemmas C.5 and C.4, we know that conditions [i-iii] for each $W_t \in \mathbf{W}_\star$, combined across $\mathbf{W}_\star$ form a necessary and sufficient set to realize $P(\mathbf{W}_\star)$.

Note that condition [iii] is always satisfied because the agent need never perform a Fisherian RAND(V) for any V. It can get the same effect by performing CTF-RAND$(V \to C)$ for each $C \in Ch(V)$. Step 12 of the sub-routine, Algo. 3 would never be invoked.

From Theorem 3.5, we know that **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ returns FAIL if and only if there are conflicts in conditions [i-ii] when combined across all the terms $\mathbf{W}_\star$.

Define the new set $\mathbf{W}'_\star$ by dropping x from the subscript (if it appears) for each $W_\mathbf{t} \in \mathbf{W}_\star$, and dropping X from $\mathbf{W}_\star$ (if it appears). Since X is first in topological order of $\mathcal{G}$, this does not add any *new* conflicts across conditions [i-ii] induced by each term in $\mathbf{W}'_\star$ (by Lemma C.6). It also doesn't *remove* any conflicts that are not related to X, as argued in the proof of Theorem 3.5, since X comes topologically first.

Define the graph $\mathcal{G}'$ as the projection of $\mathcal{G}$ that marginalizes out X (and adds bidirected edges if needed). $\mathcal{G}'$ has $\leq n$ nodes. From the IH, we conclude that **CTF-REALIZE**$(P(\mathbf{W}'_\star), \mathcal{G}', \mathbb{A}^\dagger(\mathcal{G}'))$ does not return FAIL if and only if $An(\mathbf{W}'_\star)$ does not contain two potential responses $W_\mathbf{t}, W_\mathbf{s}$ of the same variable under different regimes.

However, note that (as discussed in the proof of Theorem 3.5, and from inspecting the algorithm), the only difference between **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ and **CTF-REALIZE**$(P(\mathbf{W}'_\star), \mathcal{G}', \mathbb{A}^\dagger(\mathcal{G}'))$ is that in the former, the outer loop of **CTF-REALIZE** first checks for conflicts in the conditions [i-ii] across $\mathbf{W}_\star$ w.r.t X. After that, the steps for both algorithms are the identical.

Therefore, any conflicts detected by **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ that are *not* detected by **CTF-REALIZE**$(P(\mathbf{W}'_\star), \mathcal{G}', \mathbb{A}^\dagger(\mathcal{G}'))$ must be conflicts w.r.t X. By the IH, these additional conflicts (unrelated to X) cannot be because of a pair of conflicting potential responses in $An(\mathbf{W}'_\star)$.

We have already established that removing X to make $\mathbf{W}'_\star$ does not remove or add any conflicting potential response pairs that don't involve X. Therefore, our task is to now show that each of these additional conflicts (involving X) must correspond to at least one conflicting pair of potential responses in $An(\mathbf{W}_\star)$, that are not present in $An(\mathbf{W}'_\star)$. And conversely, we need to show that each pair of conflicting potential responses in $An(\mathbf{W}_\star)$ involving X (i.e., that is not present in $An(\mathbf{W}'_\star)$)

corresponds to at least one conflict detected by **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ in the outer loop involving X.

Forward direction:

As discussed in the proof of Theorem 3.5, **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ returns FAIL in the outer loop involving X if and only if the input of X to some $C \in Ch(X)$ is required to be some x to satisfy condition [i] w.r.t some $W_\mathbf{t} \in \mathbf{W}_\star$, but also required to be x' or "natural" to satisfy condition [i/ii] w.r.t some $Y_\mathbf{h} \in \mathbf{W}_\star$.

Note that the action set is $\mathbb{A}^\dagger(\mathcal{G})$. Therefore step 6 of sub-routine Algo. 3 would always pick only the procedure CTF-RAND$(X \to C)$ whenever C needs to receive a fixed value. The interventions affecting other $C' \in Ch(X)$ would not affect C.

In this case, it is easy to see that the set $An(W_\mathbf{t})$ must contain $C_{x\ldots}$ per Def. 3.6, and $An(Y_\mathbf{h})$ must contain $C_{x'\ldots}$ or a potential response of C without X in the subscript. Thus, if **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ returns FAIL in the outer loop involving X, there must be a pair of conflicting potential responses in $An(\mathbf{W}_\star)$.

Reverse direction:

Assume there exists a conflicting pair of potential responses $A_\mathbf{t}, A_\mathbf{s} \in An(\mathbf{W}_\star)$ where $x \in \mathbf{t}$ and $\mathbf{s}$ contains some x' or does not contain X at all.

This means there is some $W_\mathbf{h} \in \mathbf{W}_\star$ s.t. $A \in An(W)_{\mathcal{G}_{\underline{X}}}$ and some $Y_\mathbf{j} \in \mathbf{W}_\star$ s.t. $A \in An(Y)_{\mathcal{G}_{\underline{X}}}$. I.e., A mediates the effect of X on W, Y. Further, from Def. 3.6, it means that A needs to be realized in conflicting regimes w.r.t X.

From Lemma C.5, such conflict happens because for $A_\mathbf{t}$, condition [i] requires that X is fixed by intervention to be x for all children $C \in Ch(X) \cap An(A)$. Whereas, for $A_\mathbf{s}$, condition [i] or [ii] requires that each child $C \in Ch(X) \cap An(A)$ receives X either fixed as x', or naturally (as the case may be, for $\mathbf{s}$). For any such $C \in Ch(X) \cap An(A)$, it is clear from the proof of Theorem 3.5 that this conflict will trigger a FAIL from **CTF-REALIZE**$(P(\mathbf{W}_\star), \mathcal{G}, \mathbb{A}^\dagger(\mathcal{G}))$ in the first outer loop involving X.

Thus, we have shown that the IH holds for any $\mathbf{W}_\star$ involving a graph with $n + 1$ nodes.

Since Theorem 3.5 shows **CTF-REALIZE** is complete, we have thus proved that $\mathbf{W}_\star$ is realizable given $\mathcal{G}$ and $\mathbb{A}^\dagger(\mathcal{G})$ if and only if $An(\mathbf{W}_\star)$ does not contain a pair of conflicting potential responses for the same variable under different regimes.

■

Corollary 3.8 :

This follows from Corollary 3.7. For any causal diagram, the ancestral set of $\{Y_x, Y_{x'}\}$ would include both these potential responses.

Thus, the query is not realizable.

■

C.3 Proofs for Section 4

Remark C.7. In order for an agent to enact a non-trivial decision strategy $\pi : \{\mathbf{W}_\star = \mathbf{w}\} \mapsto \mathcal{A}$, we observe that (1) the distribution $P(Y_\mathcal{A}, \mathbf{W}_\star)$ must be realizable (Def. 3.4); and (2) the agent must be able to observe $\mathbf{W}_\star$ before performing actions $\mathcal{A}$. We call this a *realizable decision strategy*, and notate the space of all realizable strategies in a MAB problem as Π (Fig. 2). ■

Theorem 4.1 :

From Lemma C.8, all strategies involve mappings, where each mapping maps to one of the following 5 possible action sets: (1) $\{\}$ (no action); (2) $\textsc{Write}(X : x)$, for some x; (3) only $\textsc{ctf-Write}(x \to Y)$ for some x; (4) only $\textsc{ctf-Write}(x'' \to D)$ for some x''; or (5) both $\textsc{ctf-Write}(x \to Y)$, $\textsc{ctf-Write}(x'' \to D)$ for some x, x''.

Define Π_5 to be the space of strategies where every mapping of each strategy in Π_5 is mapping to a pair of actions $\textsc{ctf-Write}(x \to Y)$, $\textsc{ctf-Write}(x'' \to D)$ for some x, x''. I.e., all mappings only involve possibility (5) under these strategies, for all any unit encountered in the decision problem.

Let π_5 be an optimal strategy in this space. I.e., $\pi_5 \in \arg\max_{\pi \in \Pi_5} \mu_\pi$.

By Lemma C.10, π_5 is also an optimal strategy in the space of all possible strategies. This means we only need to consider strategies whose mappings are mappings to a pair of $\textsc{ctf-Write}$ procedures.

Let $\mathbf{W}_\star$ be the context used by π_5. If $\mathbf{W}_\star$ does not already contain the natural variables X, Z, we can always define π_5' that use context $\mathbf{W}_\star' = \mathbf{W}_\star \cup \{X, Z\}$ s.t. $\mu_{\pi_5'} = \mu_{\pi_5}$, where π_5' simply ignores the extra context variable in the mapping.

Such a move would not affect the realizability of π_5' because $\textsc{ctf-Write}$ does not override the natural value of X, and both X, Z can be observed before decision-making.

Combinatorially, there are only 3 possibilities for picking each mapping in π_5'.

1. Mapping from $\{x', z\}$ to a pair of actions $\{\textsc{ctf-Write}(x \to Y), \textsc{ctf-Write}(x'' \to D)\}$
2. Mapping from $\{x', z\}$ to some $\textsc{ctf-Write}(x \to Y)$, observing $Y_x = y$, and mapping from $\{x', z, y_x\}$ to $\textsc{ctf-Write}(x'' \to D)$; or
3. Mapping from $\{x', z\}$ to some $\textsc{ctf-Write}(x'' \to D)$, observing $D_{x''} = d$, and mapping from $\{x', z, d_{x''}\}$ to $\textsc{ctf-Write}(x \to Y)$

We can use similar arguments to Lemma C.10, where we restricted our attention to the space of strategies Π_5 which could mimic all other optimal strategies.

Possibility 2 can be mimicked by some mapping following possibility 1 which maps to a joint pair of actions. The two are equivalent in terms of outcome, because conditioning on y_x to choose x'' does not affect the outcome Y. So we can restrict our attention to possibilities 1 and 2.

Each mapping of possibility 1 can be mimicked by possibility 3, where the extra step of conditioning on $d_{x''}$ just ignores the extra information about $d_{x''}$. Thus, we can replace all mappings in the optimal strategy π_5' with mappings of possibility 3, to get a strategy π_5'' that also performs optimally.

Since there are two mappings in π_5'', they must be the mappings which maximize the outcome.

This is precisely the definition of the strategy π^{opt} given in Sec. 4 and in the description immediately following it.

■

Lemma C.8. *Any decision strategy π for a decision problem having causal structure same as the MAB template is s.t. each mapping of the strategy maps from domain of the context to one of the five following possible sets of actions: (1) $\{\}$ (no action); (2) $\textsc{Write}(X : x)$, for some x; (3) only $\textsc{ctf-Write}(x \to Y)$ for some x; (4) only $\textsc{ctf-Write}(x'' \to D)$ for some x''; or (5) both $\textsc{ctf-Write}(x \to Y)$, $\textsc{ctf-Write}(x'' \to D)$ for some x, x''.*

Proof. Since the physical action space only involves doing nothing, $\textsc{Write}$ or $\textsc{ctf-Write}$. Any other combination would be equivalent to one of the 5 above. E.g., $\textsc{Write}(X : x)$ and $\textsc{ctf-Write}(x'' \to D)$ is the equivalent to the pair $\textsc{ctf-Write}(x \to Y)$ and $\textsc{ctf-Write}(x'' \to D)$ (see Remark D.5).

We ignore randomized actions for simplicity. From standard results in learning theory, there is an optimum to be found at a simplex corner so we need only search over the space of hard interventions.

■

Lemma C.9. *The context $\mathbf{W}_\star$ used in the strategy $\pi : \{\mathbf{W}_\star = \mathbf{w}\} \mapsto \mathcal{A}$ can only possibly contain a subset of $X, Z, D, D_{x''}$ for some x'', and at most one potential response of D.*

Proof. There are only 4 variables to consider: X, Y, Z, D.

By the definition of a realizable strategy (Remark C.7), we need $P(Y_\mathcal{A}, \mathbf{W}_\star)$ to be realizable. By Cor. 3.7 there cannot be two potential responses of the same variable in a realizable distribution. This rules out any other potential response of Y, and ensures only one each of X, Z, D.

Since the only possible actions are interventions involving X, which do not affect Z and X (natural variable), these are the only potential responses that could appear involving these variables.

Likewise, with D, D (natural value) and $D_{x''}$ are the only possible potential responses that could appear, and at most one of them can. ■

Lemma C.10. *If π_5 is an optimal strategy in Π_5, the set of all strategies which map to a pair of* CTF-WRITE *procedures, then π_5 is also an optimal strategy in the set of all strategies possible in the MAB decision problem.*

Proof. Let Π_1 be the space of all strategies possible in the problem. Note that $\Pi_5 \subseteq \Pi_1$. Let $\pi_1 \notin \Pi_5$ be an optimal strategy. I.e. $\pi_1 \in \arg\max_{\pi \in \Pi_1} \mu_\pi$. If no such π_1 the Lemma stands proved.

Let $\mathbf{W}_\star$ be the context used by π_1. If $\mathbf{W}_\star$ does not already contain the natural variable X, we can always define π_1' that uses context $\mathbf{W}_\star' = \mathbf{W}_\star \cup \{X\}$ s.t. $\mu_{\pi_1'} = \mu_{\pi_1}$, where π_1' simply ignores the extra context variable in the mapping. For now, it doesn't matter whether such π_1' is realizable or not. Just that it is also an optimal strategy.

Each mapping in the strategy π_1' maps from the domain of $\mathbf{W}_\star'$ to one of the five possible action sets mentioned in Lemma C.8. E.g., for some $\mathbf{W}_\star' = \mathbf{w}$, the strategy π_1' maps this to $\mathbf{w} \mapsto \{\}$ or $\mathbf{w} \mapsto$ WRITE(X).

Consider a mapping in π_1' from the domain of $\mathbf{W}_\star' = \{x', ...\}$ to possibility (1), empty set of actions (recall, the context includes natural X). Such a mapping can be mimicked by an equivalent mapping $\mathbf{W}_\star' = \{x', ...\} \mapsto \{\text{CTF-WRITE}(x' \to Y), \text{CTF-WRITE}(x' \to D)\}$. By the consistency property if $X(\mathbf{u}) = x'$, then $Y_{x'}(\mathbf{u}) = Y(\mathbf{u})$ and $D_{x'}(\mathbf{u}) = D(\mathbf{u})$.

Thus, we can replace all the mappings in π_1' that involve a mapping to the empty set of actions, with an equivalent pair of CTF-WRITE using the natural value of X observed in the context. Call this new strategy π_2. π_2 is as good as π_1' because the mappings are all equivalent. Thus, π_2 is also optimal in Π_1. Again, it doesn't matter that π_2 may not be realizable, just that it is optimal.

Next, consider a mapping in π_2 from the domain of $\mathbf{W}_\star' = \{x', ...\}$ to possibility (2), some action WRITE$(X : x)$. Such a mapping can be mimicked by an equivalent mapping $\mathbf{W}_\star' = \{x', ...\} \mapsto \{\text{CTF-WRITE}(x \to Y), \text{CTF-WRITE}(x \to D)\}$. The evaluation of $f_Y(x, Z, \mathbf{u})$ in both scenarios is the same, with the only difference being that f_X is overwritten, which doesn't affect the outcome Y for each $\mathbf{u}$. I.e., the outcome Y would be the same for every unit under both strategies.

Thus, we can replace all the mappings in π_2 that involve a mapping to some action WRITE$(X : x)$, with an equivalent pair of CTF-WRITE. Call this new strategy π_3. π_3 is as good as π_2 because the mappings are all equivalent in terms of outcome. Thus, π_3 is also optimal in Π_1.

Next, consider a mapping in π_3 from the domain of $\mathbf{W}_\star' = \{x', ...\}$ to possibility (3), some action CTF-WRITE$(x \to Y)$. Such a mapping can be mimicked by an equivalent mapping $\mathbf{W}_\star' = \{x', ...\} \mapsto \{\text{CTF-WRITE}(x \to Y), \text{CTF-WRITE}(x' \to D)\}$ for natural value x'. By the consistency property, if $X(\mathbf{u}) = x'$ then $D_{x'}(\mathbf{u}) = D(\mathbf{u})$.

Thus, we can replace all the mappings in π_3 that involve a mapping to some action CTF-WRITE$(x \to Y)$, with an equivalent pair of CTF-WRITE. Call this new strategy π_4. π_4 is as good as π_3 because the mappings are all equivalent in terms of outcome. Thus, π_4 is also optimal in Π_1.

Next, consider a mapping in π_4 from the domain of $\mathbf{W}_\star' = \{x', ...\}$ to possibility (4), some action CTF-WRITE$(x \to D)$. Such a mapping can be mimicked by an equivalent mapping $\mathbf{W}_\star' = \{x', ...\} \mapsto \{\text{CTF-WRITE}(x' \to Y), \text{CTF-WRITE}(x \to D)\}$ for natural value x'. By the consistency property, if $X(\mathbf{u}) = x'$ then $Y_{x'}(\mathbf{u}) = Y(\mathbf{u})$.

Thus, we can replace all the mappings in π_4 that involve a mapping to some action CTF-WRITE$(x \to D)$, with an equivalent pair of CTF-WRITE. Call this new strategy π_5'. π_5' is as good as π_4 because the mappings are all equivalent in terms of outcome. Thus, π_5' is also optimal in Π_1.

However, note that the only possible mappings in π_5' are possibility (5) involving a pair of CTF-WRITE actions. Which means $\pi_5' \in \Pi$.

Thus, we show that all optimal strategies in Π_5 are also optimal in the overall space of strategies. ■

Corollary 4.2 : This result follows immediately, by simply recognizing that $\pi^{int}, \pi^{obs} \in \Pi$, the space of realizable strategies (we assume the agent can perform RAND(X), WRITE$(X : x)$).

Therefore, $\mu_{\pi^{ctf}}$ cannot be less than $\mu_{\pi^{int}}, \mu_{\pi^{obs}}$, by Theorem 4.1. ■

C.4 Realizability of $\mathcal{L}_1$- and $\mathcal{L}_2$-distributions

We define the probability measure $P^{\mathbb{C}}(.)$ from the perspective of an exogenous agent $\mathbb{C}$'s beliefs about the environment, distinguished by superscript from $P^{\mathcal{M}}(.)$, the true unknown distribution.

Since unit selection is randomized, $\text{SELECT}^{(i)}$ yields an unbiased sample of a unit with latent features distributed according to the target population frequency $P(\mathbf{u})$. I.e., $P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \text{SELECT}^{(i)}) = P^{\mathcal{M}}(\mathbf{u})$. We also assume that target population size is large enough that $\text{SELECT}^{(i)}$ does not significantly change the distribution of the remaining population.

Further, we assume that the actions $\text{READ}(V)^{(i)}$ and $\text{RAND}(V)^{(i)}$ do not disrupt any other mechanism $f_{V'}$ for unit i.

Remark C.11. Let $\mathcal{A}^{(i)}$ be a sequence of actions taken by agent $\mathbb{C}$ on unit i that is not conditional on any data gathered regarding i. The assumption of $\mathbb{C}$ behaving exogenously means that $P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \mathcal{A}^{(i)}) = P^{\mathcal{M}}(\mathbf{u})$. ■

Lemma C.12 (Observational sample)**.** *An agent $\mathbb{C}$ can draw an i.i.d sample distributed according to the $\mathcal{L}_1$ query $P(\mathbf{V})$ associated with an SCM $\mathcal{M}$, by the following actions:*

i. $\text{SELECT}^{(i)}$

ii. $\text{READ}(\mathbf{V})^{(i)} = \mathbf{v} \sim P(\mathbf{V})$

Given N i.i.d samples, the consistent unbiased estimate of $P(\mathbf{v})$ is

$$\hat{P}(\mathbf{v}) := \frac{1}{N} \sum_i \prod_{v \in \mathbf{v}} \mathbb{1}[\text{READ}(V)^{(i)} = v] \tag{11}$$

Proof. This follows directly from the definitions of the actions. $\text{SELECT}^{(i)}$ chooses a unit at random from the population. By Remark C.11, $P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \text{SELECT}^{(i)}) = P^{\mathcal{M}}(\mathbf{u})$. For randomly selected unit i,

$$
\begin{aligned}
& P^{\mathbb{C}}(\text{READ}(\mathbf{V})^{(i)} = \mathbf{v} \mid \text{SELECT}^{(i)}) && && (12) \\
&= \sum_{\mathbf{u}} P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \text{SELECT}^{(i)}). && && (13) \\
& P^{\mathbb{C}}(\text{READ}(\mathbf{V})^{(i)} = \mathbf{v} \mid \mathbf{U}^{(i)} = \mathbf{u}, \text{SELECT}^{(i)}) && \text{Chain rule} && \\
&= \sum_{\mathbf{u}} P^{\mathbb{C}}(\mathbf{U}^{(i)} = \mathbf{u} \mid \text{SELECT}^{(i)}).\mathbb{1}^{\mathcal{M}}[\mathbf{V}(\mathbf{u}) = \mathbf{v}] && \text{Def. 2.1(ii)} && (14) \\
&= \sum_{\mathbf{u}} P^{\mathcal{M}}(\mathbf{u}).\mathbb{1}^{\mathcal{M}}[\mathbf{V}(\mathbf{u}) = \mathbf{v}] && \text{Rem. C.11} && (15) \\
&= P^{\mathcal{M}}(\mathbf{v}) && \text{Definition} && (16)
\end{aligned}
$$

I.e., this record is an i.i.d. sample from $P^{\mathcal{M}}(\mathbf{V})$. Now consider the estimator below.

$$\hat{P}(\mathbf{v}) := \frac{1}{N}\sum_{n}\prod_{v\in\mathbf{v}}\mathbb{1}^{\mathbb{C}}[\textsc{Read}(V)^{(i)} = v] \tag{17}$$

$$= \frac{1}{N}\sum_{n}\sum_{\mathbf{u}}\prod_{v\in\mathbf{v}}\mathbb{1}^{\mathcal{M}}[\mathbf{U}^{(i)} = \mathbf{u}].\mathbb{1}^{\mathcal{M}}[V(\mathbf{u}) = v] \tag{18}$$

Un-biasedness is established by taking expectation on either side, w.r.t the agent $\mathbb{C}$'s actions (choice of units to observe):

$$\mathbb{E}_{\mathbb{C}}[\hat{P}(\mathbf{v})] = \mathbb{E}_{\mathbb{C}}\left[\frac{1}{N}\sum_{n}\sum_{\mathbf{u}}\prod_{v\in\mathbf{v}}\mathbb{1}^{\mathcal{M}}[\mathbf{U}^{(i)} = \mathbf{u}].\mathbb{1}^{\mathcal{M}}[V(\mathbf{u}) = v]\right] \tag{19}$$

$$= \sum_{\mathbf{u}}\frac{1}{N}\mathbb{E}_{\mathbb{C}}\left[\sum_{n}\mathbb{1}^{\mathcal{M}}[\mathbf{U}^{(i)} = \mathbf{u}]\prod_{v\in\mathbf{v}}.\mathbb{1}^{\mathcal{M}}[V(\mathbf{u}) = v]\right] \quad \text{Linearity of expectation} \tag{20}$$

$$= \sum_{\mathbf{u}}\frac{1}{N}\mathbb{E}_{\mathbb{C}}\left[\sum_{n}\mathbb{1}^{\mathcal{M}}[\mathbf{U}^{(i)} = \mathbf{u}]\right]\prod_{v\in\mathbf{v}}\mathbb{1}^{\mathcal{M}}[V(\mathbf{u}) = v] \quad V(\mathbf{u}) \text{ constant wrt } \mathbb{C} \tag{21}$$

$$= \sum_{\mathbf{u}}\frac{1}{N}\left[N.P^{\mathcal{M}}(\mathbf{u})\right]\prod_{v\in\mathbf{v}}\mathbb{I}^{\mathcal{M}}[V(\mathbf{u}) = v] \quad \text{Def. 2.1(i), Rem. C.11} \tag{22}$$

$$= P^{\mathcal{M}}(\mathbf{v}) \quad \text{Definition} \tag{23}$$

Consistency is established by the fact that as $\mathcal{N}$ (target population size) $\to \infty$, and N (sample size) $\to \infty$,

$$\frac{1}{N}\sum_{n}\mathbb{I}^{\mathcal{M}}[\mathbf{U}^{(i)} = \mathbf{u}] \to P^{\mathcal{M}}(\mathbf{u}) \tag{24}$$

■

Lemma C.13. *The $\mathcal{L}_2$ distribution of an atomic intervention is equivalent to the $\mathcal{L}_2$ distribution of the corresponding conditional stochastic intervention.*

$$P^{\mathcal{M}}(\mathbf{v}; do(\mathbf{x})) = P^{\mathcal{M}}(\mathbf{v}|\mathbf{x}; \sigma_{\mathbf{X}}) \tag{25}$$

$$= \sum_{\mathbf{u}}\underbrace{\mathbb{1}[\mathbf{V}_{\sigma_{\mathbf{X}}}(\mathbf{u}) = \mathbf{v} \mid X_{\sigma_{\mathbf{X}}} = \mathbf{x}]}_{①}.\underbrace{P(\mathbf{u})}_{②} \tag{26}$$

Proof. The step from the r.h.s of Eq. 25 to Eq. 26 is derived as follows: in the submodel $\mathcal{M}_{\sigma_{\mathbf{X}}}$, if we are given that $\mathbf{X}$ has been randomly assigned $\mathbf{x}$, then the remaining variables are deterministically generated as a function of $\mathbf{u}$ and $\mathbf{x}$ via their respective equations. The probability mass is collected over all the $\mathbf{u}$ which produce the output $\mathbf{v}$ over all these equations.

$$P^{\mathcal{M}}(\mathbf{v}|\mathbf{x}; \sigma_{\mathbf{X}}) = \sum_{\mathbf{u}}\mathbb{I}[\mathbf{V}_{\sigma_{\mathbf{X}}}(\mathbf{u}) = \mathbf{v} \mid X_{\sigma_{\mathbf{X}}} = \mathbf{x}].P^{\mathcal{M}}(\mathbf{u}) \tag{27}$$

Notice: if $\mathbf{v}$ is incompatible with $\mathbf{x}$, the indicator in the r.h.s evaluates to 0. Next, we prove. Eq. 25.

In $\mathcal{M}_{\sigma\mathbf{X}}$, as defined, $\mathbf{X}$ is assigned according to an independent random vector. Notate this vector as $\mathbf{X}_{\sigma_{\mathbf{X}}}$ and let the distribution of this vector be $P_{\sigma_{\mathbf{X}}}(\mathbf{X})$, defined by the assignment frequency over the target population.

$\mathcal{M}_{\sigma\mathbf{X}}$ is defined such that the target population is split into groups, each assigned $(X_{\sigma_{\mathbf{X}}} = \mathbf{x})$ for some $\mathbf{x}$. Note, the assignment vector $\mathbf{X}_{\sigma_{\mathbf{X}}}$ is independent of the latent features $\mathbf{U}$ across the target population iff each finite group assigned $(X_{\sigma_{\mathbf{X}}} = \mathbf{x})$ has the same distribution of latent features $P(\mathbf{U})$ as in the overall target population.

The above discussion handles the finite size of the target population. Starting with the r.h.s of Eq. 25,

$$P^{\mathcal{M}}(\mathbf{v}|\mathbf{x}; \sigma_{\mathbf{X}}) = \frac{P(\mathbf{v}, \mathbf{x}; \sigma_{\mathbf{X}})}{P(\mathbf{x}; \sigma_{\mathbf{X}})} = \begin{cases} P(\mathbf{v}; \sigma_{\mathbf{X}})/P(\mathbf{x}; \sigma_{\mathbf{X}}) & \text{if } \mathbf{v} \text{ compatible with } \mathbf{x} \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

Evaluating for when $\mathbf{v}$ is compatible with $\mathbf{x}$:

$$\frac{P(\mathbf{v};\sigma_{\mathbf{X}})}{P(\mathbf{x};\sigma_{\mathbf{X}})} = \frac{P(\mathbf{v};\sigma_{\mathbf{X}})}{P_{\sigma_{\mathbf{X}}}(\mathbf{x})} \tag{29}$$

$$= \frac{\sum_{\mathbf{u}} \left(P(\mathbf{u}) \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}} P(v_i \mid \mathbf{pa}_i, \mathbf{u}_i) . P_{\sigma_{\mathbf{X}}}(\mathbf{x}) \right)}{P_{\sigma_{\mathbf{X}}}(\mathbf{x})} \quad \text{Truncated factorization product} \tag{30}$$

$$= \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}} P(v_i \mid \mathbf{pa}_i, \mathbf{u}_i) \tag{31}$$

$$= P^{\mathcal{M}}(\mathbf{v}; do(\mathbf{x})) \quad \text{Truncated factorization product} \tag{32}$$

Eq. 30 uses the fact that each sub-group assigned $(X_{\sigma_{\mathbf{X}}} = \mathbf{x})$, by independence, has the same frequency of latent features $P(\mathbf{u})$. ■

Lemma C.14 (Interventional sample)**.** *An agent $\mathbb{C}$ can draw an i.i.d sample distributed according to the $\mathcal{L}_2$ query $P(\mathbf{V}; do(\mathbf{x}))$ associated with an SCM $\mathcal{M}$, by the following actions:*

i. $\textsc{Select}^{(i)}$

ii. $\textsc{Rand}(\mathbf{X})^{(i)}$

iii. If $\textsc{Rand}(\mathbf{X})^{(i)} = \mathbf{x}$*, then* $\textsc{Read}(\mathbf{V})^{(i)} = \mathbf{v} \sim P(\mathbf{V}; do(\mathbf{x}))$*, else repeat i-iii.*

Given $N_{\mathbf{x}}$ i.i.d samples, the consistent unbiased estimate of Eq. 26 is given by

$$\hat{P}(\mathbf{v}; do(\mathbf{x})) = \underbrace{\frac{1}{N_{\mathbf{x}}} \sum_i}_{②} \underbrace{\mathbb{1}[\textsc{Read}(\mathbf{V})^{(i)} = \mathbf{v}, \textsc{Rand}(\mathbf{X})^{(i)} = \mathbf{x}]}_{①}, \tag{33}$$

Proof. The proof steps are similar to the ones used for the Observational i.i.d sample case. Note that Remark C.11 still hold since even though the agent is conditioning on the value randomly assigned to a particular unit i, this value is independent of the unit's latent features $\mathbf{U}^{(i)}$. ■

D Details on counterfactual randomization

In this appendix, we provide a full formal account for the procedure we define in Sec. 2, called CTF-RAND$(X \to \mathbf{C})$ (Def. 2.2).

- In Sec. D.1, we lay out the structural conditions under which it is possible to perform this procedure, and we provide an algorithm (Algorithm 4) by which an agent can translate the structural conditions in the environment into a list of CTF-RAND procedures that it is able to perform in the given setting.
- In Sec. D.2, we emphasize that it is possible for an agent to enact multiple randomization procedures involving the same variable X for a single unit i, and illustrate this with an example.
- In Sec. D.3, we discuss the constraints implied by the assumptions we make. In particular, we discuss why CTF-RAND$(X \to \mathbf{C})$ can only be performed on some $\mathbf{C} \subseteq Ch(X)$, and not by-pass children to directly affect some distant descendants of X.
- In Secs. D.4 and D.5, we provide detailed examples of experiments were the structural conditions in the environment permit the agent to perform counterfactual randomization.

D.1 Structural conditions required for counterfactual randomization

Counterfactual randomization (Def. 2.2) can be performed under two circumstances:

i. CTF-RAND$(X \to Ch(X))$ can be performed by eliciting a unit's natural decision X, while simultaneously randomizing its actual enforced decision. Thus, the agent can affect the value of the decision X as received by all the children of X, whilst also recording the natural realization of X. As discussed in Sec. 2, this was established in [4, 12, 36].

ii. CTF-RAND$(X \to \mathbf{C})$ can also be performed for some $\mathbf{C} \subseteq Ch(X)$ if there is a special *counterfactual mediator* (defined below) by which the mechanisms generating $\mathbf{C}$ perceive the value of X. This counterfactual mediator then allows the agent to intervene on the value of X as *perceived* by $\mathbf{C}$, thus mimicking an actual intervention on X.

Definition D.1 (Expanded SCM)**.** Given an SCM $\mathcal{M}$ containing observable variables $\mathbf{V}$, we define an *expanded SCM* $\mathcal{M}^+$ of the same environment to be a model containing a bigger set of observable variables $\mathbf{V}^+ \supset \mathbf{V}$, and which relaxes the positivity requirement. I.e., it is possible that $P^{\mathcal{M}^+}(\mathbf{v}^+) = 0$, for some $\mathbf{v}^+$ in $\mathcal{L}_1$. We call the causal diagram of $\mathcal{M}^+$ an *expanded* causal diagram $\mathcal{G}^+$. ■

Definition D.2 (Counterfactual mediator)**.** Given a variable X in a causal diagram $\mathcal{G}$, we call any variable $W \notin \mathbf{V}$ a *counterfactual mediator* of X w.r.t $Y \in Ch(X)_{\mathcal{G}}$ if

i. In an "expanded" SCM of the environment $\mathcal{M}^+$ (Def. D.1), W is generated according to an invertible mechanism $W \leftarrow f_W(X, \mathbf{U}_W)$ with $\mathbf{U}_W$ possibly empty, s.t. $f_W^{-1}(W) = X$;

ii. It is physically possible to perform RAND$(W)^{(i)}$ (Def. 2.1); and

iii. In $\mathcal{M}^+$, Y is generated by the mechanism $Y \leftarrow f_Y(f_W^{-1}(W), \mathbf{A}, \mathbf{U}_Y)$, where $\mathbf{A}$ is the set $\mathbf{Pa}_Y \setminus X$ in $\mathcal{G}$. ■

The intuition behind Def. D.2 is that a *counterfactual mediator* is a real variable in the environment which fully encodes information about the variable X, and which mediates how Y perceives the value of X via the "direct" causal path. For instance, in Example 1 (*Traffic Camera*), the RGB values of the video frames W are a counterfactual mediator for the mechanism f_Y (decision to issue a speeding ticket) to perceive the car's color X via the "direct" path, not via the actual speeding of the car).

Condition [i] of Def. D.2 divides the domain of W into *equivalence classes* s.t. each value w belongs to an equivalence class $\{w' : f_W^{-1}(w) = x\}$ for some value x.

Condition [iii] of Def. D.2 essentially says that the mechanism f_Y only cares about which equivalence class W belongs to. I.e., Y only cares about what W reveals about X.

Note: these conditions does not require an agent to have full knowledge of the SCM. They are rather structural assumptions about the underlying mechanisms which can be verified in a given setting. In

Example 1, treating W as counterfactual mediator means the assumption that (1) the video features W uniquely map back to the actual color of the car in the footage; and (2) the computer vision system only cares W reveals about X, and is indifferent to any stochasticity *within* some equivalence class $\{w' : f_W^{-1}(w) = x\}$.

Assumption D.3. (Tree structure) Given a variable X, causal diagram $\mathcal{G}$, and an "expanded" diagram $\mathcal{G}^+$ (Def. D.1) including the set of all the counterfactual mediators $\mathbf{W}$ (Def. D.2) of X in the environment, each $W \in \mathbf{W}$ has only one parent in $\mathcal{G}^+$, and each $C \in Ch(X)_\mathcal{G}$ has at most one $W \in \mathbf{W}$ as a parent in $\mathcal{G}^+$. ■

Assumption D.3 enforces that each child of X perceives X through at most one proxy pathway. This assumption rules out possible structures like Fig. 17(a) where a child perceives X through multiple proxy pathways. If X is a construct like gender identity, then it is possible that a child perceives X via a cluster of personal attributes $\mathbf{W}$ which indicate X. In this case, no single attribute solely satisfies Def. D.2 of a counterfactual mediator. However, the cluster of attributes $\mathbf{W}$ could be collapsed into a single variable having domain equal to the cartesian product of the sub-domains [1, 35]. This single node $\mathbf{W}$ would indeed satisfy the definition of a counterfactual mediator and would comply with the tree structure in Assumption D.3. For a comprehensive discussion of the semantics of interventions on the perception of a compound attribute such as race or gender identity, see [29, App. D.1]. The following Lemma is the key property that enables path-specific randomization.

Lemma D.4. *Given a causal diagram $\mathcal{G}$ containing variables X and $Y \in Ch(X)_\mathcal{G}$. Let W be a counterfactual mediator of X w.r.t Y (Def. D.2). For any value x, we have*

$$Y_{w\mathbf{a}}(\mathbf{u}) = Y_{x\mathbf{a}}(\mathbf{u}), \qquad \forall \mathbf{u}, \forall w \in \{w' : f_W^{-1}(w) = x\}, \tag{34}$$

where $\mathbf{A} := \mathbf{Pa}_Y \setminus X$ in $\mathcal{G}$.

Proof. This follows from Def. D.2. Suppose we are given values (w, x) where $f_W^{-1}(w) = x$. Let $\mathbf{A} := \mathbf{Pa}_Y \setminus X$ in $\mathcal{G}$.

The variable $W_x(\mathbf{u}) = W_{x\mathbf{a}}(\mathbf{u}) = f_W(x, \mathbf{u})$, in the enhanced submodel $\mathcal{M}^+_{x\mathbf{a}}$. Adding $\mathbf{a}$ to the subscript does not matter - by Assumption D.3 and Lemma D.7, $\mathbf{A}$ cannot be an ancestor of W in $\mathcal{M}^+$.

Since f_W is invertible by condition [i] in $\mathcal{M}^+$, it is also invertible in submodel submodel $\mathcal{M}^+_{x\mathbf{a}}$. Therefore, we have $f_W^{-1}(W_{x\mathbf{a}}(\mathbf{u})) = x$.

$$\begin{aligned} Y_{w\mathbf{a}}(\mathbf{u}) &= f_Y(f_W^{-1}(w), a, \mathbf{u}) \\ &= f_Y(x, a, \mathbf{u}) \\ Y_{W_{x\mathbf{a}}\mathbf{a}}(\mathbf{u}) &= f_Y(f_W^{-1}(W_{x\mathbf{a}}(\mathbf{u})), a, \mathbf{u}) \\ &= f_Y(x, a, \mathbf{u}) \end{aligned}$$

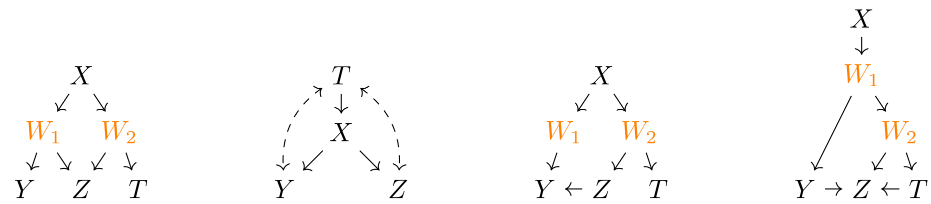
The r.h.s is identical, giving us $Y_{w\mathbf{a}} = Y_{W_{x\mathbf{a}}\mathbf{a}}$. Finally, we argue that $Y_{W_{x\mathbf{a}}\mathbf{a}} = Y_{x\mathbf{a}}$.

The counterfactual $Y_{x\mathbf{a}}$ is evaluated in a submodel of $\mathcal{M}^+$, where f_W receives input x and this value of W_x is an input to f_Y, while $\mathbf{A}$ is fixed to be $\mathbf{a}$. Structurally, this is identical to how the counterfactual $Y_{W_{x\mathbf{a}}\mathbf{a}} = Y_{W_x\mathbf{a}}$ is evaluated. Therefore, it is evident that $Y_{W_x\mathbf{a}} = Y_{x\mathbf{a}}$. ■

Given a variable X, the way an agent actually performs the action CTF-RAND is as follows:

i. **Performing CTF-RAND by eliciting natural decision:** The agent can perform $\text{CTF-RAND}(X \to Ch(X))^{(i)}$ by randomizing the unit's decision. The agent can further perform $\text{READ}(X)^{(i)}$ to elicit the unit's natural decision, which has not been erased. This is described in in [4, 12, 36].

ii. **Performing CTF-RAND using counterfactual mediators:** If X has a counterfactual mediator W in the environment, and $\mathbf{C} \subseteq Ch(X)$ are the children which perceive X via W, then the agent can perform $\text{CTF-RAND}(X \to \mathbf{C})^{(i)}$ by randomizing W. Each value w mimics randomizing X as perceived by $\mathbf{C}$, per Lemma D.4. The agent can still perform $\text{READ}(X)^{(i)}$ by measuring $X^{(i)}$ to get the unit's natural decision, which has not been erased.



(a) (b) (c) (d)

Figure 17: "Expanded" causal diagrams, where counterfactual mediators (labeled W_i) of X have been marked in red. (a) is not permitted by Assumption D.3. Assume the environment in diagram (b) permits the agent to elicit the unit's natural decision X even when randomizing the actual decision.

Having described the structural conditions that permit counterfactual randomization, we want to abstract away the mediators and succinctly describe the agent's physical actions via the definition of CTF-RAND. Given a variable X, and assumptions/knowledge about X in the environment stated in points **[i]** and **[ii]** above, we translate this knowledge into a set of counterfactual randomizations that the agent is physically able to perform in the environment, using Algorithm 4.

Algorithm 4 CTF-PROCEDURES

Input: Causal diagram $\mathcal{G}$ with decision variable X; "expanded" diagram $\mathcal{G}^+$ (Def. D.1) including the counterfactual mediators of X in the environment
Output: $\mathbb{A}_X$ - the set of CTF-RAND actions that can be performed involving X
$\mathbb{A}_X \leftarrow \emptyset$
if environment allows eliciting natural decision X even when randomizing actual decision **then**
  if X can be randomized **then**
    $\mathbb{A}_X \leftarrow \mathbb{A}_X \cup \{\text{CTF-RAND}(X \to Ch(X)_{\mathcal{G}})\}$
  end if
end if
for each counterfactual mediator W of X **do**
  Let $\mathbf{C} := \{C \mid C \in Ch(X)_{\mathcal{G}} \text{ and perceives } X \text{ via } W\}$
  $\mathbb{A}_X \leftarrow \mathbb{A}_X \cup \{\text{CTF-RAND}(X \to \mathbf{C})\}$
end for
Return $\mathbb{A}_X$

Consider Fig. 17(a-d). (a) is not permitted by Assumption D.3. We assume that in the environment represented by (b) X can be randomized for a unit in the target population without erasing the unit's natural decision, satisfying condition **[i]** mentioned earlier. Thus, when applying Algorithm 4 to diagrams (b-d), we get the following resulting set of counterfactual randomization procedures which are permitted by the structural assumptions made (unit superscript i is omitted for legibility):

(a) $\{\text{CTF-RAND}(X \to \{Y, Z\})\}$

(b) $\{\text{CTF-RAND}(X \to Y), \text{CTF-RAND}(X \to \{Z, T\})\}$

(c) $\{\text{CTF-RAND}(X \to \{Y, Z, T\}, \text{CTF-RAND}(X \to \{Z, T\})\}$

D.2 Multiple simultaneous randomizations are possible, for a single unit

For a particular decision variable X, there could be multiple randomization procedures which an agent can perform. Consider the example in Fig. 18. The "expanded" diagram on the left shows two counterfactual mediators, W_1, W_2 in a causal structure which permit an agent to perform all of the following randomization procedures: $\text{RAND}(X)^{(i)}, \text{CTF-RAND}(X \to \{Z, T, B\})^{(i)}$ and $\text{CTF-RAND}(X \to \{T, B\})^{(i)}$ for the same unit i.

However, if all three actions are performed in parallel, randomizing W_1 to enact $\text{CTF-RAND}(X \to \{Z, T, B\})^{(i)}$ will only affect variable Z. This is since the action of randomizing W_2 to further enact $\text{CTF-RAND}(X \to \{T, B\})^{(i)}$ blocks any effect on T, B from the previous action. Similarly,

$\textsc{Rand}(X)^{(i)}$ ends up affecting only variable Y, because $\textsc{ctf-Rand}(X \to \{Z, T, B\})^{(i)}$ blocks any effect from the previous action on Z, T, B. We formalize this observation in Remark D.5.

Remark D.5 (Superseding action). Given a decision variable X, the action $\textsc{ctf-Rand}(X \to \mathbf{C}')^{(i)}$ can *supersede* the action $\textsc{ctf-Rand}(X \to \mathbf{C})^{(i)}$ if $\mathbf{C}' \subsetneq \mathbf{C}$, where *supersede* means that the former action $\textsc{ctf-Rand}(X \to \mathbf{C}')^{(i)}$ blocks any effect that the latter action has on the variables $\mathbf{C}'$. Additionally, the action $\textsc{ctf-Rand}(X \to \mathbf{C})^{(i)}$ *supersedes* the action $\textsc{Rand}(X)^{(i)}$. ■

Further, Assumption D.3 ensures that all such procedures follow a "nested" structure. I.e., given any two randomization procedures involving the same variable, the sets of children affected by one will be a subset of the set affected by the other, as shown in Fig. 18.

Remark D.6. (Procedure containment) Assumption D.3 implies that if an agent is capable of performing both $\textsc{ctf-Rand}(X \to \mathbf{C})^{(i)}$ and $\textsc{ctf-Rand}(X \to \mathbf{C}')^{(i)}$ s.t. $\mathbf{C} \neq \mathbf{C}'$ and $\mathbf{C} \cap \mathbf{C}' \neq \emptyset$, then either $\mathbf{C} \subseteq \mathbf{C}'$ or $\mathbf{C}' \subseteq \mathbf{C}$. ■

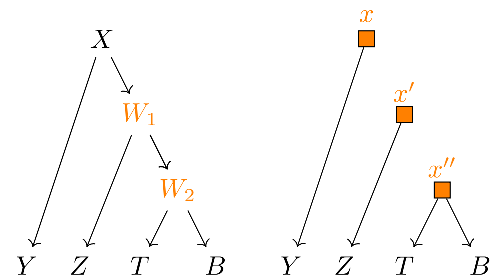


Figure 18: (Left) "Expanded" causal diagram showing counterfactual mediators W_1, W_2 of X; (Right) Agent performing actions $\textsc{Rand}(X)^{(i)}, \textsc{ctf-Rand}(X \to \{Z, T, B\})^{(i)}$ and $\textsc{ctf-Rand}(X \to \{T, B\})^{(i)}$ all together on the single unit i.

D.3 Counterfactual randomization is only possible for direct children of X

Our definition of $\textsc{ctf-Rand}(X \to \mathbf{C})^{(i)}$, is only valid for some $\mathbf{C} \subseteq Ch(X)$ in the causal diagram (Def. 2.2). This action essentially randomizes the value of decision variable X *as an input* to the mechanisms generating its causal children $\mathbf{C}$, while leaving open the possibility of measuring the unit i's natural decision (what it would have normally decided in the $\mathcal{L}_1$ regime), and also the possibility of separately and in parallel randomizing the value of X as an input to other causal children $\mathbf{C}' = Ch(X) \setminus \mathbf{C}$.

However, the notion of "child" is an abstraction w.r.t a specific diagram of the environment under study. Consider Fig. 19(a-Left), where $\mathcal{G}_1$ is the diagram of some environment. Assume there exists a counterfactual mediator W_1 of X (Def. D.2) as shown in Fig. 19(a-Middle), which means an agent is able to perform the physical action $\textsc{ctf-Rand}(X \to A)^{(i)}$, while still being able to measure the natural value of X for unit i.

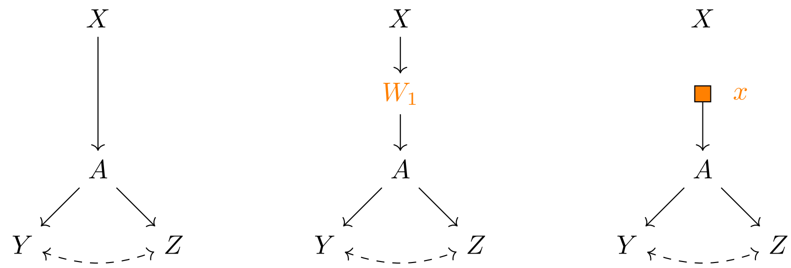
Now consider the diagram $\mathcal{G}_2$ shown in Fig. 19(b-Left). $\mathcal{G}_2$ is a valid projection of $\mathcal{G}_1$ obtained by marginalizing out variable A, and is thus also a valid causal diagram of the environment.

Suppose that there exists a counterfactual mediator W_2 as shown in 19(b-Middle). This means that the agent can also perform $\textsc{ctf-Rand}(X \to Z)^{(i)}$ in the same environment. However, since we are referring to the same environment, this means that the agent is able to perform $\textsc{ctf-Rand}(X \to Z)^{(i)}$ w.r.t the diagram $\mathcal{G}_1$, where Z is not a child node of X! This would translate to even greater experimental power w.r.t graph $\mathcal{G}_1$, where an agent is able to perform counterfactual randomization of X w.r.t further descendants like Z and draw i.i.d samples from queries like $P(A_x, Z_{x'})$ by simultaneously performing both counterfactual randomizations (i.e. by randomizing W_1, W_2 simultaneously).

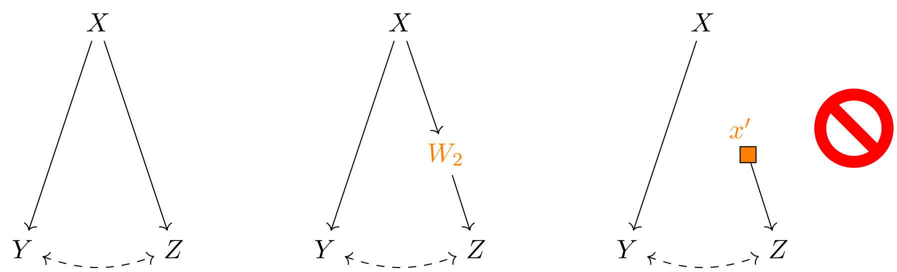
However, this scenario is not possible. Essentially, this would require an "expanded" causal diagram (Def. D.1) like shown in Fig. 20, where W_2 is a counterfactual mediator of X w.r.t Z that comes after another variable A. If A satisfies positivity w.r.t X, i.e., if $P^{\mathcal{M}}(x, a) > 0, \forall x, a$ in $\mathcal{L}_1$, then W_2 cannot be a counterfactual mediator since it cannot be uniquely mapped back to X.

Lemma D.7. *Given a causal diagram $\mathcal{G}$ of a true SCM $\mathcal{M}$ with a variable X and $A \in Desc(X)_{\mathcal{G}}$ where $P(x, a) > 0, \forall x, a$. There cannot be a variable W in an "expanded" SCM $\mathcal{M}^+$ of the environment (Def. D.1) s.t.*

- *$W \in Desc(A)_{\mathcal{G}^+}$, where $\mathcal{G}^+$ is the "expanded" causal diagram of $\mathcal{M}^+$; and*
- *W is invertible to X, i.e. exists f_W^{-1} s.t. $f_W^{-1}(W) = X$.* □



(a) : (Left) Graph $\mathcal{G}_1$; (Middle) enhanced diagram showing counterfactual mediator of X; (Right) agent performing CTF-RAND$(X \to A)^{(i)}$ using W_1



(b) : (Left) Graph $\mathcal{G}_2$; (Middle) enhanced diagram showing counterfactual mediator of X; (Right) agent performing CTF-RAND$(X \to Z)^{(i)}$ using W_2

Figure 19: Given a causal diagram $\mathcal{G}_1$ of the environment, $\mathcal{G}_2$ is a valid projection of $\mathcal{G}_1$ (marginalizing A). If A satisfies positivity w.r.t X, then there *cannot* be a counterfactual mediator W_2 as shown in (b-Middle). Which means an agent *cannot* perform CTF-RAND$(X \to Z)^{(i)}$ as shown in (b-Right).

Proof. If A satisfies positivity w.r.t X, then a given value w cannot be mapped back to a unique x, even if we marginalize out A from the SCM.

Note that, by Assumption D.3, a counterfactual mediator has only one parent in the "expanded" causal diagram (Def. D.1). I.e., if it were a descendant of A, its perception of X is fully mediated by A.

If $f'_W(X, \mathbf{U})$ is invertible from W to X, then so is $f_W \circ f_A(X, \mathbf{U})$. It is evident that $f_W'^{-1}$ is well defined iff $f_A^{-1} \circ f_W^{-1}$ is well defined.

f_A^{-1} is not defined. The positivity condition entails that a given value a could have been generated by any value x (when unit is unknown).

Since f'_W is not invertible, W cannot be a counterfactual mediator. ∎

Lemma D.7 leads to some important conclusions.

Remark D.8. There cannot be an "expanded" causal diagram (such as in Fig. 20), with a counterfactual mediator that bypasses a child-node and directly fixes a descendent-node's perception of X. I.e., an agent cannot perform CTF-RAND$(X \to D)^{(i)}$ for some $D \in Desc(X) \setminus Ch(X) \cup \{X\}$. ∎

Remark D.9. Conversely, given a graph like $\mathcal{G}_2$ in Fig. 19(b), if we are told that the agent can perform the action CTF-RAND$(X \to Z)^{(i)}$, then $\mathcal{G}_2$ cannot be a projection of $\mathcal{G}_1$ (Fig. 19(a)) for the same environment. ∎

The upshot of this discussion is that, in general (i.e. without making further assumptions), counterfactual randomization can only be done via counterfactual mediators (Def. D.2) of a decision variable X, and it can only be performed on the children-nodes of X in the general case.

D.4 Example 1 (Traffic camera) - detailed

A computer vision company's system is being considered by the New York government for detecting speeding cars and automatically issuing them speeding tickets based on footage from traffic cameras. The audit team have a concern about fairness. It is possible, they argue, that the vision company's models are trained on footage from neighborhoods which have a strong correlation between the color of the car X and speeding, maybe since socioeconomic groups have different color preferences, thus leading to certain groups being penalized unfairly. Y represents the AI decision to issue a ticket. This concern amounts to a hypothesis that X affects Y via a "direct" path as opposed to the "indirect" path via Z, an indicator of whether the car was actually speeding (Fig. 21). The indirect path describes the causal effects of, say, how pedestrians and other drivers react to a red car and affect its speeding.

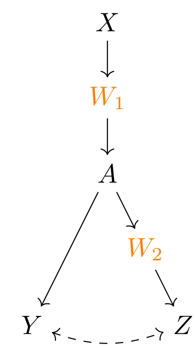


Figure 20: "Expanded" diagram that is needed to sample directly from $P(A_x, Z_{x'})$. This is not possible, per Lemma D.7.

The company claims they have conducted an RCT where they recruited drivers and randomly assigned them cars to use normally for a month, and used the footage solicited from city-wide cameras to gauge speeding of the drivers in the study over the month. Unbeknownst to the drivers they tracked the ground truth Z of whether the car was speeding (using the speedometer and a geo-tagged location of the vehicle), thus begetting the diagram in Fig. 21. Z and Y might be confounded by objects appearing in the car's vicinity (because, say, in the training data, speeding was less frequent on roads with traffic cones).

The government's audit team assess the dataset from this RCT of video clips of the cars in traffic, each labeled with ground truth Z, to test the fairness concern. The direct path is formalized via the *natural direct effect*, or NDE [25], which tracks the effect on outcome Y of changing its perception of X from x to x', while fixing Z to be whatever value it was under x. NDE is defined as follows

$$\text{NDE}_{x,x'}(y) = P(y_{x' Z_x}) - P(y_x) \tag{35}$$

The second term in Eq. 35 can be estimated from experimental data using Lemma C.14. But it is unclear how to estimate first term, even though X could be subjected to a Fisherian RCT. However, the audit team recognizes there exists a special mediator, viz. the features W in the video footage which reveal the car's color to the vision system. W could be the RGB values of the pixels in the frames of the video footage, that reveals the car color X to the model. They use standard video-editing tools to randomly swap the color of the car in the footage. By randomly assigning a particular car $W \leftarrow$ red, they are able to affect the mechanism f_Y's *perception* of X.

Verification of structural assumptions: the audit team verifies that W satisfies Def. D.2 of a counterfactual mediator of X w.r.t Y. **Condition [i]** says that each w (say, a specific RGB range of pixels) belongs to an equivalence class that maps back uniquely to a car color x. This can be verified using the RCT data. **Condition [ii]** is satisfies since they can do a targeted randomization of W. **Condition [iii]** stipulates that f_Y is not affected by any artefacts introduced by the color-editing tool: this can be verified, for instance, by swapping a car's color from x to x' and then swap it back from x' to x, to ensure that the model's decision Y does not change, thus verifying that the mechanism f_Y only cares about what the color features W reveal about X, and not about any image artefacts that may be introduced by editing.

Thus, they are able to perform the following derivation

$$\begin{aligned}
&P(Y_{W=\text{red}} \mid X = \text{blue}) && \text{Est. via Lemma C.14} && (36)\\
=&P(Y_{W=\text{red},Z} \mid X = \text{blue}) && Z : \text{natural value} && (37)\\
=&P(Y_{W=\text{red},Z_{X=\text{blue}}} \mid X = \text{blue}) && X = \text{blue} \implies Z = Z_{X=\text{blue}} && (38)\\
=&P(Y_{X=\text{red},Z_{X=\text{blue}}} \mid X = \text{blue}) && \text{Lemma D.4} && (39)\\
=&P(Y_{X=\text{red},Z_{X=\text{blue}}}) && && (40)
\end{aligned}$$

Eq. 38 derives from the consistency property [24, Cor. 7.3.2]. Eq. 40 is permitted because of the d-separation implied by the causal diagram in Fig. 4. Thus, the team is able to directly estimate the $\mathcal{L}_3$-quantity $P(y_{x' Z_x})$ via a physical procedure. Using Eq. 35, they can gauge whether a car's color has a *direct* effect on the odds of getting a speeding ticket. ■

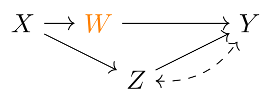


Figure 21: "Enhanced" diagram for *traffic camera* example, where W is *counterfactual mediator* for X

D.5 Example 2 (College admissions)

This example is inspired by an important case study in labor economics [6]. Consider a field experiment where a team of economists generate fake CVs for applications to a particular college. The content of the CVs is generated from a set of generic templates that are quality-controlled. Each unit i is a fictitious applicant. X denotes the applicant's race, fixed uniformly at random for each i and uncorrelated with the contents of the CV. The CV goes to two different departments, the financial scholarship team and the undergrad admissions team. Z, Y denote, respectively, the decision to grant financial aid and the decision to pass the student through the first stage of CV-screening. Importantly, these two departments receive separate copies of the same application and don't interact with each other regarding their decisions, to avoid any interference. The "expanded" causal diagram is shown in Fig. 22(Left).

The team wants to study how qualified students are unable to access higher education because of financial situation, especially across racial groups, and wants to evaluate the counterfactual. In particular, for a fixed set of qualifications (given by the CV) what would the applicant's screening outcome be had they been race x and what would the same applicant's financial assistance outcome had been had they been race x? I.e., of interest is the $\mathcal{L}_3$ distribution $P(Y_x, Z_{x'})$. This quantity is non-identifiable from $\mathcal{L}_2$ data, due to the confounding between Y and Z.

They make the following modeling assumptions: (1) since the CV body is randomly generated for each X, there is no *indirect* effect of X on Z, Y via the body of the CV; (2) for each fictitious applicant of race $X = x$, they randomly choose a name from an equivalence class of names which stereotypically indicate one unique race group x; and (3) Y, Z don't care about which name in particular appears on the CV, the reviewers only care about the race "revealed" by that name. These are strong assumptions needed to make further progress. In particular, (2) can be done using census data to build an equivalence class of, say, names like Lakisha and Jamal for Black applicants, and last names like Nguyen and Xi for Asian applicants (cf. [6]), while ignoring ethnically ambiguous names like John and Jane.

Since Y and Z operate independently, this allows one to identify two counterfactual mediators (Def. D.2), viz., the names W_1, W_2 seen by Y, Z respectively, that mediate the direct effect of X on each department. By separately and simultaneously randomizing W_1, W_2, the agent is thus able to sample from the desired $\mathcal{L}_3$ distribution

$$
\begin{aligned}
&P(Y_{W_1=w}, Z_{W_2=w'}) && \text{Est. via Lemma C.14} \\
=&P(Y_x, Z_{x'}) && \text{Lemma D.4,}
\end{aligned}
$$

where w, w' are randomly chosen names from the equivalence classes of names, s.t. $f_{W_1}^{-1}(w) = x$ and $f_{W_1}^{-1}(w') = x'$, and where x, x' are racial groups that are indicated by names w, w'.

Figure 22: (Left) "Expanded" diagram for the *College Admissions* example, where W_1, W_2 are counterfactual mediators for X; (Right) An agent can compute $P(Y_x, Z_{x'})$ by simultaneously randomizing W_1, W_2 for each unit.

E Experiment details

E.1 Algorithmic Sentencing

The SCM used in this hypothetical scenario is as follows:

$$U_Z \sim Bernoulli(0.2)$$
$$U_A \sim Bernoulli(0.3)$$
$$U_B \sim Bernoulli(0.3)$$
$$\epsilon \sim Bernoulli(0.25)$$

$$Z \leftarrow U_Z$$

Since X, Y are functions of Z,

i. If $Z = 0$

$$X \leftarrow U_A$$
$$Y \leftarrow U_A \oplus X \oplus \epsilon \qquad (\oplus \text{ is the XOR function})$$

ii. If $Z = 1$

$$X \leftarrow \neg U_B$$
$$Y \leftarrow U_B \oplus X \oplus \epsilon$$

Z is a propensity score of the defendant's risk profile, summarizing a high-dimensional set of features like the defendants history, compliance, urine tests etc.

U_A, U_B are latent factors that affect the defendant's recidivism outcome, and which the judge is able to intuit based on years of experience and use to make her natural decision X. These latent factors may be the defendant's speech, mannerism, personality, similarity to previous cases the judge has seen etc. that reveal, say, the trustworthiness of the individual to commit to a rehab plan.

ϵ is an environmental noise factor.

$\mathcal{L}_1$-regime: The observational data is summarized in Fig. 23(a). The natural regime suggests that the judge is rather conservative. For high-risk individuals she makes good decisions ($E[Y \mid Z = 1] = 0.75$), but for low-risk individuals, her outcomes are poor ($E[Y \mid Z = 0] = 0.25$). Overall, the outcome is $E[Y] = 0.35$, since the study population is 80% low-risk.

$\mathcal{L}_2$-regime: Applying the interventions $do(x_0), do(x_1)$ for each context Z, we can compute the expected outcome from the SCM as shown in Table 3.

This suggests a sensible strategy of allowing the judge to decide on high-risk cases, and perhaps use an exogenous RL agent to decide on low-risk cases with the action $do(x_1)$. Call this the "naive mixed" regime, which would incur an expected outcome of

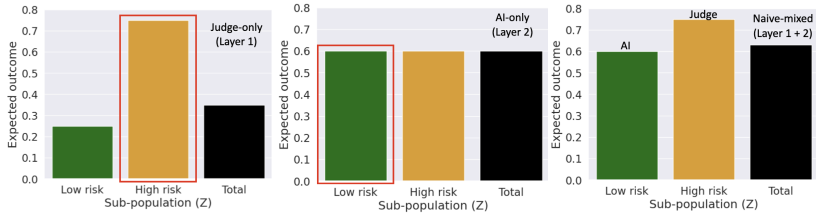
$$E[Y; \text{"naive mixed"}] = P(z_0).E[Y_{x_1} \mid z_0] + P(z_1).E[Y \mid z_1] = (0.8)(0.6) + (0.2)(0.75) = 0.63,$$

which is better than the $\mathcal{L}_1$ (observational) expected outcome of 0.35.

$\mathcal{L}_3$-regime: In contrast, by counterfactual randomization we can sample from the $\mathcal{L}_3$ distribution $P(Y_x, X, Z)$, allowing the exogenous RL agent to observe $X = x', Z = z$ for a particular study unit and execute the actions

$$\textsc{Read}(X) = x'; \textsc{Read}(Z) = z$$

$$\textsc{ctf-Write}(x \to Y), \text{ where } x = \arg\max_x E[Y_x \mid x', z]$$



(a) (b) (c)

Figure 23: Expected outcomes in *Algorithmic Sentencing* experiment under (a) Judge-only $\mathcal{L}_1$ strategy; (b) RCT $\mathcal{L}_2$ strategy; and (c) Naive mixed $\mathcal{L}_1 + \mathcal{L}_2$ strategy of using judge when $Z = 1$ and RCT strategy when $Z = 0$.

	$Z = 0$	$Z = 1$
$do(x_0)$	0.4	0.4
$do(x_1)$	**0.6**	**0.6**

Table 3: Expected outcome $E[Y_x \mid z]$ computed under the interventional regime.

Computing from the SCM directly, this $\mathcal{L}_3$ strategy yields an expected outcome of

$$\sum_{x',z} P(x', z). \max_x \mathbb{E}[Y_x \mid x', z] = 0.75,$$

outperforming both $\mathcal{L}_1$, contextual $\mathcal{L}_2$, and the "naive mixed" $\mathcal{L}_2$ strategies. It turns out the best strategy in this problem is to go with the judge's natural decision X when the defendant is high-risk ($Z = 1$), and to do the opposite of the judge's recommendation when the defendant is low risk ($Z = 0$).

Simulations The experiments compare the performance of three algorithms, where TS and TS^{aug} are the conventional Thompson Sampling algorithms, using context $\{Z = z\}$ and $\{Z = z, X = x'\}$ respectively. Experiments were run for 1000 iterations, 200 epochs.

The third algorithm TS^{ett} doesn't treat $\{X = x'\}$ merely as another context variable. Rather, the causal significance of the natural decision is leveraged via the consistency property

$$E[Y_{x'} \mid x', z] = E[Y \mid x', z]$$

The r.h.s allows us to hot-start the algorithm with observational data for demonstrably lower cumulative regret, as shown in Algorithm 5.

Note on hot-starting It is common practice to hot-start Bandit algorithms with observational data, expecting faster convergence (e.g. hot starting an A/B testing algorithm with historical observational data). If this is done heuristically, oblivious to the causal structure and without warrant from inference rules, there is no guarantee that this speeds up convergence. In fact, *hot-starting can be shown to hurt performance in SCMs where the observational data recommends the wrong arm.*

To see why, imagine a simple 2-variable decision problem $\{X, Y\}$, where the variables are confounded in the natural regime. It is possible to define an SCM where $P(Y = 1 \mid X = 0) > P(Y = 1 \mid X = 1)$ but where $P(Y = 1; do(X = 1)) > P(Y = 1; do(X = 0))$.

Hot-starting a Bandit algorithm with observational data would place heavy weight on arm $X = 0$, which means it takes longer to discover the optimal arm $X = 1$.

Algorithm 5 TS^{ett}: Thompson Sampling ETT (Bernoulli-Beta case)

Input: No. of timesteps, T; Observational data, $P(\mathbf{v})$
for $z \in$ Domain(Z) **do**
  for $x' \in$ Domain(X) **do**
    $\alpha[z][x'] \leftarrow 1$
    $\beta[z][x'] \leftarrow 1$ {Initializing priors}
  end for
end for
$t = 1$
while $t <= T$ **do**
  Perform $\text{READ}(Z) = z$, for unit
  Perform $\text{READ}(X) = x'$, for unit
  for $i \in$ Domain(X) **do**
    if $x_i = x'$ **then**
      $\mu_i \leftarrow E[Y \mid x', z]$ {Hot-start using obs. data}
    else
      $\mu_i \sim Beta(\alpha[z][x_i], \beta[z][x_i])$
    end if
  end for
  Perform $\text{CTF-WRITE}(x \to Y)$ where $x = x_i$ s.t. $i := \arg\max_{i'} \mu_{i'}$
  Perform $\text{READ}(Y) = y$, for unit {Get value of Y_x}
  if $x \neq x'$ **then**
    $\alpha[z][x'] \leftarrow \alpha[z][x'] + y$
    $\beta[z][x'] \leftarrow \beta[z][x'] + 1 - y$ {Update priors}
  end if
  $t \leftarrow t + 1$
end while

E.2 Big Tech surveillance

The name Omega is an homage to the adversary in the (in)famous *Newcomb's Problem* [21], a powerful predictor of the protagonist's future choices. The SCM used in this hypothetical scenario to generate data is as follows:

$$U_1 \sim Bernoulli(0.5)$$
$$U_2 \sim Bernoulli(0.5)$$
$$U_3 \sim Bernoulli(0.5)$$

$$X \leftarrow U_1 \oplus U_2 \qquad (\oplus \text{ is the XOR function})$$
$$D \leftarrow X \oplus U_3$$

Since Y is a function of X, the average outcome is shown below for different realizations of the latents

i. Avg. Y, when $U_3 = 0$

	$U_3 = 0$			
	$U_1 = 0$		$U_1 = 1$	
	$U_2 = 0$	$U_2 = 1$	$U_2 = 0$	$U_2 = 1$
$do(x_0)$	**0.6**	0.9	0.8	**0.5**
$do(x_1)$	0.9	**0.6**	**0.5**	0.8

Natural choice of X marked **bold**

ii. Avg. Y, when $U_3 = 1$

	$U_3 = 1$			
	$U_1 = 0$		$U_1 = 1$	
	$U_2 = 0$	$U_2 = 1$	$U_2 = 0$	$U_2 = 1$
$do(x_0)$	**0.8**	0.7	0.6	**0.7**
$do(x_1)$	0.7	**0.8**	**0.7**	0.6

iii. Avg. Y, with U_3 marginalized (consolidating i. and ii.)

	$U_1 = 0$		$U_1 = 1$	
	$U_2 = 0$	$U_2 = 1$	$U_2 = 0$	$U_2 = 1$
$do(x_0)$	**0.7**	0.8	0.7	**0.6**
$do(x_1)$	0.8	**0.7**	**0.6**	0.7

U_1, U_2, U_3 are latent attributes affecting Alice's decisions each evening. In particular, U_1 indicates whether she is tired, U_2 indicates whether she had a busy day and is distracted, U_3 indicates whether she is hungry, on any given evening.

If Alice is either tired but mentally relaxed ($X = 1 \oplus 0$), or if she is physically energetic but distracted ($X = 0 \oplus 1$), Alice decides to take a walk and use Omega via mobile app. If Alice is neither tired nor distracted she prefers to continue working on her desktop and uses Omega via desktop app during breaks ($X = 0 \oplus 0$). If she is both tired and distracted, she also decides to use Omega on her desktop because she has no energy to take a walk ($X = 1 \oplus 1$).

There are so many possible factors affecting her decisions, Alice is unaware that these are the specific unconscious causes of her natural choices. However, Omega's unscrupulous data scientists surveil U_1, U_2, U_3 (perhaps by tracking Alice's wearable health monitor and calendar) and predict her natural choice. Omega then uses behavioural insights to ping Alice with the precise notifications and content to maximize her time spent on the platform for each realization of U_1, U_2, U_3.

D is the type of ads Alice sees when she logs in to Omega for the day.

$\mathcal{L}_1$-regime: The observational data is contained in Table (iii) in the SCM above, where the bold values correspond to Alice's natural choices. Given that all combinations of latents happen with equal probability, it is easy to see that the expected reward in the observational regime is $E[Y] = (0.25)(0.7 + 0.7 + 0.6 + 0.6) = 0.65$.

$\mathcal{L}_2$-regime: Applying the interventions $do(x_0), do(x_1)$, we can compute the expected outcome from the SCM as shown in Table 4. This is simply the average of all the values in Table (iii) of the SCM above.

An interventional strategy of randomizing ones actions (or fixing a constant action) outperforms the observational $\mathcal{L}_1$ regime of allowing one's actions to be determined by natural inclination.

	$E[Y; do(x)]$
$do(x_0)$	0.7
$do(x_1)$	0.7

Table 4: Expected outcome $E[Y_x]$ computed under the interventional regime.

$\mathcal{L}_3$-regime - ETT: By counterfactual randomization Alice can sample from the $\mathcal{L}_3$ distribution $P(Y_x, X)$. She records her natural choice $X = x'$ on a particular evening (what she would have normally done) and randomizes the choice of X that she actually undertakes, during the explore phase. Using this distribution, she then performs the following action, for the natural $X = x'$ that she observes in the exploit phase:

$$do(X = \arg\max_x E[Y_x \mid x'])$$

We can compute this from Table (iii) of the SCM. Alice simply chooses to do the opposite of what she naturally feels like doing (corresponding to the the non-bold cells of the Table). This "ETT" $\mathcal{L}_3$ strategy yields an expected outcome of

$$\sum_{x'} P(x').\max_x E[Y_x \mid x'] = (0.5)[0.7 + 0.8] = 0.75,$$

outperforming both $\mathcal{L}_1$ and $\mathcal{L}_2$ strategies.

$\mathcal{L}_3$-regime - Optimal: Finally, Alice leverages her ability to perform *path-specific* randomization to sample from the distribution $P(Y_x, X, D_{x'})$ in the explore phase. She then performs the following actions in the exploit phase:

$$\textsc{Read}(X) = x'$$

$$\text{CTF-}\textsc{Write}(x'' \to D), \text{ where } x'' = \arg\max_{x''}\left(\max_{x} E[Y_x \mid x', D_{x''}]\right); \textsc{Read}(D) = d$$

$$\text{CTF-}\textsc{Write}(x \to Y), \text{ where } x = \arg\max_{x} E[Y_x \mid x', d_{x''}]$$

In words, during the exploit phase, Alice first obtains $D_{x''} = d$ and $X = x'$, and then performs the action x which maximizes her outcome Y_x. She chooses x'' that gives her the best global optimum of $E[Y_x \mid x', d_{x''}]$.

Computing this from the SCM, suppose Alice chooses to record $D_{x_0} = 0 \oplus U_3 = U_3$.

- When $D_{x_0} = U_3 = 0$, Alice sees according to Table (i) of the SCM that the optimal strategy is to choose the opposite of what she naturally feels like doing (the values not in bold), giving the expected outcome $E[Y_x \mid x', D_{x_0} = 0]$, where $x \neq x'$, as $(0.5)[0.9 + 0.8] = 0.85$
- When $D_{x_0} = U_3 = 1$, Alice sees according to Table (ii) of the SCM that the optimal strategy is to go with her natural inclination (the values in bold), giving the expected outcome $E[Y_{x'} \mid x', D_{x_0} = 0] = (0.5)[0.8 + 0.7] = 0.75$
- Overall, since both values of D_{x_0} are equally likely, this strategy yields an expected outcome of $0.5[0.85 + 0.75] = 0.8$, which outperforms $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$-ETT strategies.

Repeating this strategy, except by choosing to measure D_{x_1} will yield an identical expected outcome, however Alice's strategy will be vice-versa. When $D_{x_1} = 0$, it is optimal to follow her natural inclination and when $D_{x_1} = 1$ she had best do the opposite of what she feels like.

Simulations The experiment compares the performance of four algorithms, where TS is the conventional Thompson Sampling algorithm, TS^{aug} is Thompson Sampling using context $\{X = x', D_{x''} = d\}$ for some random x'', and TS^{ett} is given in Algorithm 5. Experiments were run for 2000 iterations, 200 epochs.

The fourth algorithm TS^{opt} doesn't treat $\{X = x', D_{x''} = d\}$ merely as extra context variables, but makes use of the consistency axiom to hot-start the Bandit algorithm as shown in Algorithm 6.

Algorithm 6 TS^{opt}: Thompson Sampling OPTIMAL (Bernoulli-Beta case)

Input: No. of timesteps, T; Observational data, $P(\mathbf{v})$
for $x'' \in \text{Domain}(X)$ **do**
  for $x' \in \text{Domain}(X)$ **do**
    $\alpha_D[x''][x'] \leftarrow 1$
    $\beta_D[x''][x'] \leftarrow 1$ {Initializing D-priors}
  end for
end for
for $i \in \text{Domain}(X)$ **do**
  for $j \in \text{Domain}(X)$ **do**
    for $d \in \text{Domain}(D)$ **do**
      for $k \in \text{Domain}(X)$ **do**
        $\alpha_Y[x_i][x_j][d][x_k] \leftarrow 1$
        $\beta_Y[x_i][x_j][d][x_k] \leftarrow 1$ {Initializing Y-priors}
      end for
    end for
  end for
end for
$t = 1$
while $t <= T$ **do**
  Perform $\textsc{Read}(X) = x'$, for unit
  for $j \in \text{Domain}(X)$ **do**
    $\mu_i^D \sim Beta(\alpha_D[x''][x_j], \beta_D[x''][x_j])$
  end for
  Perform $\textsc{ctf-Write}(x' \to D)$ for $x' = x_j; j := \arg\max_{j'} \mu_{j'}^D$
  Perform $\textsc{Read}(D) = d$, for unit {Get value of $D_{x''}$}
  for $k \in \text{Domain}(X)$ **do**
    if $x_k = x' = x''$ **then**
      $\mu_k^Y \leftarrow E[Y \mid x'', d]$ {Hot-start using obs. data}
    else
      $\mu_k^Y \sim Beta(\alpha_Y[x''][x'][d][x_k], \beta_Y[x''][x'][d][x_k])$
    end if
  end for
  Perform $\textsc{ctf-Write}(x \to Y)$ for $x = x_k; k := \arg\max_{k'} \mu_{k'}^Y$
  Perform $\textsc{Read}(Y) = y$, for unit {Get value of Y_x}
  $\alpha_D[x''][x'] \leftarrow \alpha_D[x''][x'] + y$
  $\beta_D[x''][x'] \leftarrow \beta_D[x''][x'] + 1 - y$ {Update D-priors}
  if $\neg(x = x' = x'')$ **then**
    $\alpha_Y[x''][x'][d][x] \leftarrow \alpha_Y[x''][x'][d][x] + y$
    $\beta_Y[x''][x'][d][x] \leftarrow \beta_Y[x''][x'][d][x] + 1 - y$ {Update Y-priors}
  end if
  $t \leftarrow t + 1$
end while

F Signature of realizable distributions

Consider the PCH depicted in Fig. 24. From Corollary 3.7, we know that an $\mathcal{L}_3$-distribution is not realizable given $\mathcal{G}$ and $\mathbb{A}^\dagger(\mathcal{G})$ iff $An(\mathbf{W}_\star)$ contains two potential outcomes of the same variable under different regimes (depicted in the shaded blue region). In this section, we provide a characterization of the converse of this corollary. I.e., what is the signature of $\mathcal{L}_3$-distributions which *are* realizable given $\mathcal{G}$ and $\mathbb{A}^\dagger(\mathcal{G})$?

First, we introduce the notion of a nested counterfactual [8], and extend the definition of realizability to accommodate this notion.

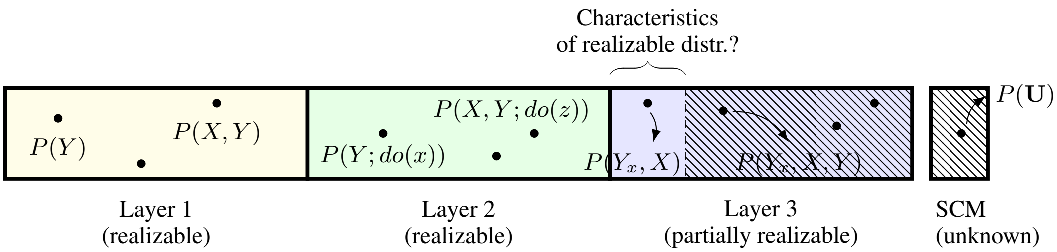


Figure 24: Pearl Causal Hierarchy (PCH) induced by an unknown SCM $\mathcal{M}$. An $\mathcal{L}_3$-distribution is not realizable if $An(\mathbf{W}_\star)$ repeats a variable, per Cor. 3.7 (shaded region). We wish to characterize the $\mathcal{L}_3$-distributions which *are* realizable given $\mathbb{A}^\dagger(\mathcal{G})$.

Definition F.1 (Nested potential response)**.** Given SCM $\mathcal{M}$, a *nested potential response* $Y_{\mathbf{W}_\star}$ is the evaluation of Y in a "derived" model where, for each $W_\mathbf{t} \in \mathbf{W}_\star$, the corresponding structural equation f_W is replaced by the value of f_W evaluated in the sub-model $\mathcal{M}_\mathbf{t}$. $Y_{\mathbf{W}_\star}$ can also be defined recursively where $\mathbf{W}_\star$ itself contains a nested potential response. ■

As an example, in Fig. 25, $Y_{A_x,B_{x'}}$ is the nested potential response according to a "derived" model where f_A is replaced by the value of A evaluated according to $\mathcal{M}_x$ and f_B by the value of B evaluated according to $\mathcal{M}_{x'}$. Conceptually, this is the value of Y had A been what it would have under a fixed x and had B been what it would have under a fixed x'. A *nested counterfactual distribution* is some distribution $P(\mathbf{W}_\star)$ where $\mathbf{W}_\star$ contains (possibly nested) potential responses. The realizability of a nested counterfactual distribution can be defined analogously to that of a regular $\mathcal{L}_3$-distribution (Def. 3.4): the ability to draw samples from the distribution.

Consider an arbitrary SCM $\mathcal{M}$, with a singleton decision variable X and an outcome variable Y. Let $\mathcal{G}$ be the causal diagra of $\mathcal{M}$. We define the following regular expression.

$$\textit{Realizability signature} := P(Y_{\{\text{child-nest}\}}, \mathbf{E}_{\{\text{child-nest}\}}, \mathbf{C}, \mathbf{D}_{\{\text{child-nest}\}}, X), \tag{41}$$

with the following auxiliary definitions:

- $V_{\{\text{child-nest}\}}$ denotes the nested potential response $V_{A_{x\mathbf{Pa}_A}, B_{x''\mathbf{Pa}_B}, \dots}$ where $A, B \in Ch(X) \cap An(V)$;
- $\mathbf{E}_{\{\text{child-nest}\}}$ denotes the set $\{E_{\{\text{child-nest}\}} \mid E \in An(Y)_{\mathcal{G}_{\overline{X}}} \cup Desc(Y)_{\mathcal{G}_{\overline{X}}}\}$;
- $\mathbf{C}$ denotes the set $\{C \mid C \in NDesc(X)\}$; and
- $\mathbf{D}_{\{\text{child-nest}\}}$ denotes the set $\{D_{\{\text{child-nest}\}} \mid D \in Desc(X) \setminus \mathbf{E}\}$;

In words, *child-nest* denotes the pathways from X to V via children of X. It is the nested potential response of V in a derived model where each child of X (that is also an ancestor of V) receives some fixed input x, while all other inputs as naturally received.

The *realizability signature* for an SCM with singleton decision variable is a regular expression that describes realizable counterfactual distributions in the environment. Any (possibly nested) counterfactual distribution that is realizable given $\mathcal{G}$ and $\mathbb{A}^\dagger(\mathcal{G})$ should fit the template in Eq. 41, although some terms might be missing.

Consider the diagram in Fig. 25. The terms in the realizability signature for this diagram are:

- $Y_{\{\text{child-nest}\}} : Y_{A_x B_{x'}}$
- $\mathbf{E}_{\{\text{child-nest}\}} : \{E_{A_x B_{x'}}, W_{A_x B_{x'}}\}$
- $\mathbf{C} : \{C, F\}$
- $\mathbf{D}_{\{\text{child-nest}\}} : \{D_{x''}, G_{D_{x''}}\} = \{D_{x''}, G_{x''}\}$
- X : natural decision

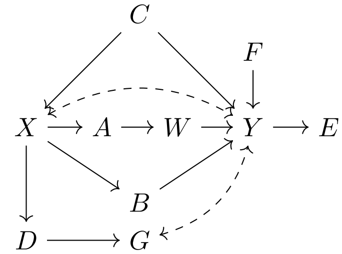


Figure 25

Any realizable (possibly nested) counterfactual distribution in this environment contains some or all of these terms.

Examples of distributions which are realizable given $\mathbb{A}^\dagger(\mathcal{G})$ are as follows:

i. ETT: $P(Y_x, X)$
ii. $P(Y_x, X, D_{x''})$
iii. $P(Y, W, X, D)$

Examples of distributions which are *not* realizable are as follows:

iv. PN/PS: $P(Y_x, X, Y)$
v. $P(Y_x, X, D_{x'}, G_{x''})$

Example [iv] is not realizable because it has multiple Y terms, which only occurs once in the *realizability signature*. And example [v] is not realizable because $\mathbf{D}_{\{\text{child-nest}\}}$ as defined in the *signature* requires $D_{x'}$ and $G_{x'}$ to share the same subscript (which can possibly be empty).

Note: the set of subscripts in the definition of the *realizability signature* in Eq. 41 are limited to $\emptyset, x_1, x_2, ..., x_m$, where $m = |Ch(X)|$. I.e., the subscripts in a realizable distribution are either empty of some x, x'' etc. (of which there can be at most as many as the number of children of X).

G Related works

Decision-making has been studied extensively, but predominantly with an $\mathcal{L}_2$ mindset.

Much work has been done in the area of counterfactual ($\mathcal{L}_3$) identification and estimation, including completeness results in the non-parametric setting [32, 8]. Some works have investigated how to overcome impossibility of identification, by bounding the range of non-identifiable counterfactual quantities [37], or by making parametric assumptions. As clarified in Sections 1 and 3, our work does not involve counterfactual identification from available data, but rather completeness results for when the data can itself be gathered, in the non-parametric setting. An interesting extension to our work would be to investigate the relationship between realizability and identification, viz., which additional $\mathcal{L}_3$-quantities now become identifiable if the environment permits counterfactual randomization?

In the area of causal decision-making, counterfactual strategies have been studied in the growing field of causal reinforcement learning (CRL) [4, 12, 36]. The literature currently focuses on ETT-related strategies where an agent both measures natural decision and randomizes the actual decision for a single unit. A similar notion has also been described using Single World Intervention Graphs (SWIG) [31], where the split node in a SWIG could represent the joint measurement of a natural decision and the randomization of the actual decision for a single unit. We present an important extension by formalizing counterfactual randomization via *counterfactual mediators* (Appendix D). An ETT-based approach only allows one randomization of a variable X, affecting all downstream mechanisms. Our formalization of counterfactual randomization (Def. 2.2) recognizes the possibility of *path-specific* randomization, isolating only certainly causal pathways and granting more granular experimental capabilities. Using this broader definition, we develop an optimal $\mathcal{L}_3$-decision strategy that outperforms the ETT baseline, as shown in Experiment 2 in Section 4.

Ctf-randomization can be used to gather data from $\mathcal{L}_3$-distributions, while identification and bounding can use this data to shed light on non-realizable $\mathcal{L}_3$-quantities, such as the probabilities of causation [23]. In settings where one cares directly about the *canonical type* of a unit (i.e., whether the unit is a "complier", "defier" etc. [2, 3]), such bounds on the probabilities of causation can be informative about the mixture of these types in a population [18, 19]. Realizability, identification and bounding thus work synergistically to aid a decision-maker.

The impossibility of measuring certain counterfactual quantities has been discussed in the literature under the title of the "fundamental problem of causal inference" (FPCI) [15]. We clarify the connection of our main results to the FPCI in Corollary 3.8.

The notion of "randomizing perception" of an attribute has been discussed in [28, 6]. We describe the necessary structural conditions under which randomization of a proxy variable (what we call a counterfactual mediator of the attribute) actually mimics a perceived intervention on the attribute, in Appendix D, and in particular, in Def. D.2. For a discussion of the semantic interpretation of such interventions, and a brief survey of related references, see [29, App. D.1].

Our work also complements the counterfactual editing literature in the fields of NLP and Computer Vision [9, 22].