TECHNICAL REPORT
R-82
May 2022

Counterfactual Transportability: A Formal Approach

Juan D. Correa [1] **Sanghack Lee** [2] **Elias Bareinboim** [3]

Abstract

Generalizing causal knowledge across environments is a common challenge shared across many of the data-driven disciplines, including AI and ML. Experiments are usually performed in one environment (e.g., in a lab, on Earth, in a training ground), almost invariably, with the intent of being used elsewhere (e.g., outside the lab, on Mars, in the real world), in an environment that is related but somewhat different than the original one, where certain conditions and mechanisms are likely to change. This generalization task has been studied in the causal inference literature under the rubric of *transportability* (Pearl and Bareinboim, 2011). While most transportability works focused on generalizing associational and interventional distributions, the generalization of counterfactual distributions has not been formally studied. In this paper, we investigate the transportability of counterfactuals from an arbitrary combination of observational and experimental distributions coming from disparate domains. Specifically, we introduce a sufficient and necessary graphical condition and develop an efficient, sound, and complete algorithm for transporting counterfactual quantities across domains in nonparametric settings. Failure of the algorithm implies the impossibility of generalizing the target counterfactual from the available data without further assumptions.

1. Introduction

Counterfactuals form the basis for different notions across human cognition and decision-making, including credit assignment, regret, responsibility and blame. Counterfactual relations require retrospective thinking, where one must be able to compare what did happen with what would have happened under some alternative hypothesis (Pearl, 2000). Given the impossibility of observing an alternative outcome once an action is taken, counterfactuals evoke "what if?" questions which answer can only be approached by imagining hypothetical conditions usually contrary to the factual evidence. For instance, questions such as "what would be the death rates had the vaccination started two weeks earlier?" or "given that I arrived late, would I have been on time had I taken the subway instead of the taxi?" require us to carry out a mental experiment where we recover some state of affairs, perform a change in the sequence of events, and let a hypothetical situation to play out. More generally, counterfactuals are an important component in the construction of explanations regarding why events occurred the way they did. For instance, the previous questions could be related to "why did the death rate achieve the number it did?" or "was it the way of transportation that caused my late arrival?". (Pearl & Mackenzie, 2018)[Ch. 8]

Formally, a structural account of causation provides suitable semantics for representing counterfactual statements (Pearl, 2000). Each structural causal model (SCM) $\mathcal{M}$ models a generative process and induces a collection of distributions related to the activities of seeing (observational), acting (interventional), and imagining (counterfactual), which together form what is known as the *ladder of causation* (Pearl & Mackenzie, 2018; Bareinboim et al., 2022). In practice, the SCM $\mathcal{M}$ is usually not fully observable, which leads to the inferential challenge of using data from one part of the ladder to make inferences about another. For instance, there exists a plethora of methods allowing for inferences from observational to experimental (i.e., layers 1 to 2 in the ladder) (Pearl, 1995; Tian & Pearl, 2003; Shpitser & Pearl, 2006; Huang & Valtorta, 2006; Bareinboim & Pearl, 2012b; Lee et al., 2019), and from observational and experimental to counterfactual distributions (Pearl, 2001; Avin et al., 2005; Shpitser & Pearl, 2007; Correa et al., 2021).

In practice, obtaining different experimental distributions for the same population is often highly nontrivial. One of the key aspects of human cognition is the ability to generalize concepts from one domain to another. The task of leveraging causal invariances so as to extrapolate and fuse experimental knowledge across settings has been formally studied in the causal inference literature under the rubric

[1] Department of Computer Science, Universidad Autónoma de Manizales, Manizales, Colombia [2] Departament of Data Science, Seoul National University, Seoul, South Korea [3] Department of Computer Science, Columbia University, New York, USA. Correspondence to: Juan D. Correa <jcorrea@autonoma.edu.co>.

Proceedings of the 39^{th} International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022. Copyright 2022 by the author(s).

of *transportability* (Pearl & Bareinboim, 2011). By and large, there are several graphical conditions and algorithms for the transportability of causal effects from a combination of observational and experimental data in various settings (Bareinboim & Pearl, 2012a; Lee & Honavar, 2013a;b; Bareinboim & Pearl, 2013; 2014; 2016; Lee et al., 2020; Correa & Bareinboim, 2019; 2020).

Despite the powerful identifiability and transportability results found in this literature, it is still largely unknown how to transport counterfactual distributions across different environments and changing conditions. In particular, the literature on transportability has been focused on the extrapolation of observational and experimental distributions (layers 1 and 2 of the ladder) but has not addressed how to operate within counterfactual ones (layer 3). For concreteness, consider an example motivated by (Powdthavee et al., 2013).

Example 1.1 (Compulsory education and well-being)**.** Consider an economic study to understand the effects of *compulsory schooling* (X) on people's *subjective well-being* (Y). A researcher group in Australia performed a controlled longitudinal experiment to assess the effect of X on *income* (Z), written as $P(Z \mid do(X))$. A causal model that describes this scenario is shown with the graph in Figure 1, where each variable corresponds to a vertex and the edges describe how variables causally influence one another. The bidirected arrow between X and Z indicates the existence of unmeasured confounders that affect both X and Z (e.g., social status, race, neighborhood).

Another group of researchers in the United States aims to determine how strong is the influence of X on Y by means other than Z. This "influence" can be captured through a quantity known as the *natural direct effect* (NDE), which is written in counterfactual language as $E[Y_{x',Z_x} - Y_{x,Z_x}]$ (Pearl, 2001)[Def. 5]. The first expression is the value Y attains if X is held constant at x' while Z still follows $X = x$. Noteworthy, this is a typical counterfactual quantity since Z and Y consider X as taking different values, while in the real world the variable X can take only one value at a time. The second quantity represents the value of Y when $X = x$ and Z vary accordingly. As the researchers believe that people in the US perceive well-being based on income, which is different than in Australia, they are surveying several people in order to obtain the observational distribution $P^*(X, Z, Y)$. This difference between the populations is represented with a node pointing to Y (■→ Y), as shown in Figure 1. The distributions with superscript indicate the target population, in this case, the US, and those without superscript represent the source population, Australia.

In this setting, the NDE cannot be determined from data from the US alone nor from Australia alone. Still, it can be determined through the following expression:

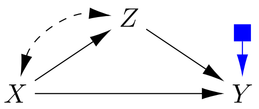


Figure 1: Causal diagram describing the causal structure of a model for studying the effect of compulsory education (X) on perceived well-being (Y) (see Example 1.1 for details).

$$\sum\nolimits_z \left(\underbrace{P^*(y \mid x', z) - P^*(y \mid x, z)}_{\text{from the US}} \right) \underbrace{P(z \mid do(x))}_{\text{from Australia}}, \tag{1}$$

In other words, the first factor in Equation (1) (in parenthesis) is a difference computed from the observational distribution in America, while the second factor (the do distribution) is from the interventional distribution in Australia. ■

In this paper, our goal is to explicate why this extrapolation (and formula) holds in this particular example, and more broadly, understand under what conditions this type of counterfactual inference across domains is allowed. We will investigate the nonparametric transportability of arbitrary counterfactual quantities when the input consists of any combination of observational and interventional distributions, gathered across different heterogeneous domains. More specifically, our contributions are as follows:

1. **Graphical characterization.** We introduce a graphical condition for determining whether a counterfactual quantity is transportable from a collection of datasets. We then prove that this condition is both necessary and sufficient.
2. **Algorithmic solution.** We develop an efficient algorithm to determine the existence of an estimand for a target counterfactual distribution, as a function of available observational and experimental distributions from the different domains. We further show that this algorithm is not only sound but also complete. In other words, the failure of the algorithm in returning an expression implies that the quantity is not transportable without further assumptions.

1.1. Preliminaries

We denote variables by capital letters, X, and values by small letters, x. Bold letters, $\mathbf{X}$ represent sets of variables and $\mathbf{x}$ sets of values. The domain of a variable X is denoted by $\mathfrak{X}_X$. Two values $\mathbf{x}$ and $\mathbf{z}$ are said to be consistent if they share the common values for $\mathbf{X} \cap \mathbf{Z}$. We also denote by $\mathbf{x} \setminus \mathbf{Z}$ the value of $\mathbf{X} \setminus \mathbf{Z}$ consistent with $\mathbf{x}$ and by $\mathbf{x} \cap \mathbf{Z}$ the subset of $\mathbf{x}$ corresponding to variables in $\mathbf{Z}$. We assume the domain of every variable is finite.

We represent qualitative assumptions in the form of causal graphs that are named with a calligraphic letter, e.g., $\mathcal{G}$, $\mathcal{H}$, etc. We denote by $\mathbf{V}(\mathcal{G})$ the set of vertices (i.e., variables) in a graph $\mathcal{H}$. Given a graph $\mathcal{G}$, $\mathcal{G}_{\overline{\mathbf{W}}\underline{\mathbf{X}}}$ is the result of removing edges coming into variables in $\mathbf{W}$ and going out from variables in $\mathbf{X}$. $\mathcal{G}[\mathbf{W}]$ denotes a vertex-induced subgraph, which includes $\mathbf{W}$ and the edges among its elements.We use kinship notation for graphical relationships such as parents, children, descendants, and ancestors of a set of variables. For example, the set of parents of $\mathbf{X}$ in $\mathcal{G}$ is denoted by $Pa(\mathbf{X})_{\mathcal{G}} := \mathbf{X} \cup \bigcup_{X \in \mathbf{X}} Pa(X)_{\mathcal{G}}$. Similarly, we define $Ch()$, $De()$, and $An()$.

To articulate and formalize the generalization of counterfactuals, we require a framework that allows us to reason about multiple domains and *alternative worlds* simultaneously. For this purpose, we use the Structural Causal Model (SCM) paradigm (Pearl, 2000). An SCM $\mathcal{M}$ is a 4-tuple $\langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{u}) \rangle$, where $\mathbf{U}$ is a set of exogenous (latent) variables; $\mathbf{V}$ is a set of endogenous (observable) variables; $\mathcal{F}$ is a collection of functions such that each variable $V_i \in \mathbf{V}$ is determined by a function $f_i \in \mathcal{F}$. Each f_i is a mapping from a set of exogenous variables $\mathbf{U}_i \subseteq \mathbf{U}$ and a set of endogenous variables $\mathbf{Pa}_i \subseteq \mathbf{V} \setminus \{V_i\}$ to the domain of V_i. The uncertainty is encoded through a probability distribution over the exogenous variables, $P(\mathbf{U})$. An SCM $\mathcal{M}$ induces a *causal diagram* $\mathcal{G}$ where $\mathbf{V}$ is the set of vertices, there is a directed edge $(V_j \rightarrow V_i)$ for every $V_i \in \mathbf{V}$ and $V_j \in \mathbf{Pa}_i$, and a bidirected edge $(V_i \dashleftarrow\dashrightarrow V_j)$ for every pair $V_i, V_j \in \mathbf{V}$ such that $\mathbf{U}_i \cap \mathbf{U}_j \neq \emptyset$ (V_i and V_j have a common exogenous parent).

We assume that the underlying model is recursive. That is, there are no cyclic dependencies among the variables. Equivalently, the corresponding causal diagram is acyclic.

The set $\mathbf{V}$ decomposes into subsets called *c-components* (Tian & Pearl, 2002b) according to a diagram $\mathcal{G}$ such that two variables belong to the same c-component if they are connected in $\mathcal{G}$ by a path made entirely of bidirected edges.

2. Structural Causal Models and Counterfactuals

Intervening on a system represented by an SCM $\mathcal{M}$ results in a new model differing from $\mathcal{M}$ only in the mechanisms associated with the intervened variables (Pearl, 1994; Dawid, 2002; 2015). Let $\widehat{\mathbf{X}}$ be a collection of functions $\{\widehat{X} : \mathfrak{X}_{\widehat{\mathbf{U}}_X} \rightarrow \mathfrak{X}_X\}_{X \in \mathbf{X}}$ for some $\mathbf{X} \subseteq \mathbf{V}$. Then, an intervention can be described by some $\widehat{\mathbf{X}}$ that induces a derived model $\mathcal{M}_{\widehat{\mathbf{X}}}$ where each f_X has been replaced by $\widehat{X}$. Then, $\mathbf{Y}_{\widehat{\mathbf{X}}}(\mathbf{u})$ is called the *potential response* of $\mathbf{Y}$ to $\mathbf{X} = \widehat{\mathbf{X}}$, and is defined as the solution of $\mathbf{Y}$, for a particular $\mathbf{u}$, in the derived model $\mathcal{M}_{\widehat{\mathbf{X}}}$.

Following (Correa et al., 2021), we use $\mathbf{W}_*$ to denote sets of arbitrary counterfactual variables[1]. Let $\mathbf{W}_* = \{W_{1[\widehat{\mathbf{T}}_1]}, W_{2[\widehat{\mathbf{T}}_2]}, \ldots\}$ represent a set of counterfactual variables such that $W_i \in \mathbf{V}$ and $\mathbf{T}_i \subseteq \mathbf{V}$ for $i = 1, \ldots, l$. Define $\mathbf{V}(\mathbf{W}_*) = \{W \in \mathbf{V} \mid W_{\widehat{\mathbf{T}}} \in \mathbf{W}_*\}$, that is, the set of observables that appear in $\mathbf{W}_*$. Let $\mathbf{w}_*$ represent a vector of values, one for each variable in $\mathbf{W}_*$.

Then, the probability of any counterfactual event is given by

$$P(\mathbf{Y}_* = \mathbf{y}_*) = \sum_{\{\mathbf{u} \mid \mathbf{Y}_*(\mathbf{u}) = \mathbf{y}_*\}} P(\mathbf{u}), \tag{2}$$

where the predicate $\mathbf{Y}_*(\mathbf{u}){=}\mathbf{y}_*$ means $\bigwedge_{\{Y_{\widehat{\mathbf{X}}} \in \mathbf{Y}_*\}} Y_{\widehat{\mathbf{X}}}(\mathbf{u}){=}y$.

If all variables in the expression have the same subscript, we could write $P(W_{1[\mathbf{x}]}, W_{2[\mathbf{x}]}, \ldots)$ as $P_{\mathbf{x}}(W_1, W_2, \ldots)$. Also, we will write $P(\mathbf{Y}_* = \mathbf{y}_*)$ simply as $P(\mathbf{y}_*)$ when there is no room for confusion.

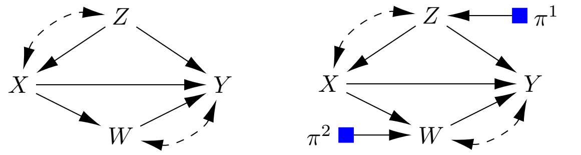
As the variables in the causal diagram have ancestors that causally affect them, counterfactual variables also have causally relevant ancestors. This generalization of the notion of ancestrality and causal relevance was formalized by (Correa et al., 2021) with the following definition.

Definition 2.1 (Ancestors of a counterfactual)**.** Let $Y_{\mathbf{x}}$ be such that $Y \in \mathbf{V}, \mathbf{X} \subseteq \mathbf{V}$. Then, the set of (counterfactual) ancestors of $Y_{\mathbf{x}}$, denoted $An(Y_{\mathbf{x}})$, consist of each $W_{\mathbf{z}}$, such that $W \in An(Y)_{\mathcal{G}_{\underline{\mathbf{X}}}}$ (which includes Y itself), and $\mathbf{z} = \mathbf{x} \cap An(W)_{\mathcal{G}_{\overline{\mathbf{X}}}}$. ■

For a set of variables $\mathbf{W}_*$, $An(\mathbf{W}_*)$ is defined as the union of the ancestors of each variable in the set. That is, $An(\mathbf{W}_*) = \bigcup_{W_{\mathbf{t}} \in \mathbf{W}_*} An(W_{\mathbf{t}})$.

Example 2.1 (College Degree and Earnings — Counterfactual ancestors)**.** For instance, consider the causal diagram in Figure 2(a) and suppose X represents college degree, W occupation, Z socio-economic factors, and Y earnings. Let x_0 = "computer science", then the counterfactual variable Y_{x_0} represents earnings had college degree been fixed to computer science ($X = x_0$). The set of ancestors, $An(Y_{x_0}) = \{Y_{x_0}, W_{x_0}, Z\}$, represents the set of random variables that causally affect Y_{x_0}. Specifically, Y_{x_0} is not affected by X or W, only by Z and W_{x_0}, which is not necessarily equal to W. Similarly, we can compute the ancestors of other counterfactual variables such as $An(W_{yz}) = \{W_z, X_z\}$ and $An(Y_w) = \{Y_w, X, Z\}$. ■

[1] When subscripts are used to enumerate variables such as $W_1, W_2, \ldots$, square brackets are added around the part of the subscript denoting interventions.



(a) A causal diagram $\mathcal{G}$ (b) A selection diagram $\mathcal{G}^{\mathbf{\Delta}}$

Figure 2: A causal diagram and a selection diagram over four variables. The selection diagram summarizes the differences between two source domains and a target domain.

3. Transporting Counterfactual Relationships Across Domains

Our goal is to assess a counterfactual quantity using assumptions encoded in the form of a graph and observational and/or experimental distributions arising from different domains. Let $\Pi=\{\pi^*, \pi^1, \pi^2, \ldots\}$ be the set of domains/populations involved in the analysis, and π^* is the *target* domain where the query is to be inferred.

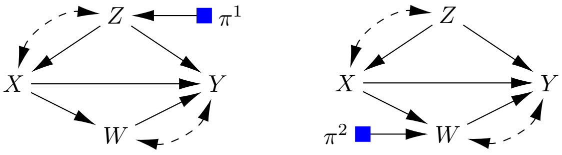
We assume that each domain has an underlying SCM that produces the samples observed. A distribution generated by a domain π^i is denoted with a superscript as P^i. For instance, the observational distribution in the domain π^* is denoted as $P^*(\mathbf{V})$. Moreover, each domain is associated with a causal diagram $\mathcal{G}^i$ describing the qualitative assumptions made for the SCM in that domain.

The ability to generalize pieces of data from one domain to another depends on the commonalities and differences between domains and the quantity of interest. The differences between domains are called "domain discrepancies" (Lee et al., 2020), formally defined next.

Definition 3.1 (Domain Discrepancy)**.** Let π^a and π^b be domains associated, respectively, with SCMs $\mathcal{M}^a$ and $\mathcal{M}^b$ conforming to a causal diagrams $\mathcal{G}^a$ and $\mathcal{G}^b$. We denote by $\Delta^{a,b} \subseteq \mathbf{V}$ a set of variables such that, for every $V_i \in \Delta^{a,b}$, there might exist a discrepancy if $f_i^a \neq f_i^b$ or $P^a(\mathbf{U}_i) \neq P^b(\mathbf{U}_i)$. ■

We will write $\Delta^{*,i}$ simply as Δ^i to represent the differences between the target and each source domain, with $\Delta^*{=}\emptyset$. Domain discrepancies can be represented graphically by augmenting the causal diagram for a domain π^i with extra nodes that represent changes in a mechanism. This type of diagram, called *selection diagram*, was first proposed in (Pearl & Bareinboim, 2011).

While previous approaches assume that all domains have the same causal structure (i.e., the arguments of the corresponding functions in every SCM match), we allow for a function to have different arguments in different domains as long as no cyclic dependencies are present. Therefore, we consider one selection diagram per domain, as defined next.



(a) $\mathcal{G}^{\Delta_1}$ (b) $\mathcal{G}^{\Delta_2}$

Figure 3: Two selection diagrams showing the possible differences between domains π^1, π^2, and π^*.

Definition 3.2 (Selection Diagram)**.** Given a causal diagram $\mathcal{G}^i = \langle \mathbf{V}, \mathbf{E} \rangle$ and domain discrepancies Δ^i, let $\mathbf{S} = \{S_v \mid \exists_{i=1}^{n} V \in \Delta^i\}$ be called *selection variables*. Then, a *selection diagram* $\mathcal{G}^{\Delta_i}$ is defined as a graph $\langle \mathbf{V} \cup \mathbf{S}, \mathbf{E} \cup \{S_v \to V\}_{S_v \in \mathbf{S}} \rangle$. ■

Let $\mathcal{G}^{\mathbf{\Delta}} = \{\mathcal{G}^{\mathbf{\Delta} \boldsymbol{i}} \mid \pi \in \Pi\}$ denote the collection of selection diagrams, including one per domain. Here, $\mathcal{G}^{\mathbf{\Delta *}}$ is always the same $\mathcal{G}^*$.

Example 3.1 (Selection diagram)**.** The selection diagram in Figure 2(b) augments the causal diagram in Figure 2(a) with two selection nodes pointing, respectively, to Z and W. These nodes are labeled with the name of the domain (or domains) for which they advertise differences with respect to the target domain π^*. This single selection diagram is a summary of the two selection diagrams shown in Figure 3(a) and Figure 3(b). Formally, this diagram encodes $\Delta^1 = \{Z\}$ and $\Delta^2 = \{W\}$. In the context of Example 2.1, Δ^1 advertises possible differences in the socio-economic conditions (Z) between the domains π^1 and π^*. Meanwhile, Δ^2 indicates differences regarding occupation (W) due, perhaps, to π^2 being a region where the occupational profile of several college program alumni differs with respect to π^*. ■

In this paper, our goal is to assess a counterfactual quantity in a target domain π^* using observational and experimental data from one or more source domains. We represent experimental distributions using regime indicators (Dawid, 2002; Tian, 2008; Correa & Bareinboim, 2019) to indicate an atomic, conditional, or stochastic policy. An atomic policy $\sigma_X = do(x)$ fixes the value of a variable X to x. A conditional intervention $\sigma_X = g(\mathbf{W})$ sets the value of X according to a deterministic function $g : \mathfrak{X}_{\mathbf{W}} \to \mathfrak{X}_X$ for some set $\mathbf{W} \subseteq \mathbf{V} \setminus De(X)$. A stochastic intervention $\sigma_X = \widehat{P}(X \mid \mathbf{W})$ sets the value of X according to a given probability distribution conditional on a similar set $\mathbf{W}$. A distribution affecting a set of variables $\mathbf{X}$ is analogously represented as $\sigma_{\mathbf{X}}$. Depending on the intervention, the causal diagram $\mathcal{G}^i_{\sigma_{\mathbf{X}}}$ corresponds to the SCM $\mathcal{M}^i$ after being intervened with $\sigma_{\mathbf{X}}$.

Example 3.2 (Experimental input distributions)**.** Recall the story in Example 2.1 and suppose we are interested in assessing the impact of studying computer science (x_0) on the earning for people in π^*. We are, however, not interested in the effect that this has on the average person in π^* but on those who choose to pursue this degree on their own. To do this, we could consider the following quantity:

$$E[Y_{x_0} \mid x_0] - \sum\nolimits_x \alpha_x E[Y_x \mid x_0]. \tag{3}$$

The first expectation in Equation (3) refers to the expected value of earnings, had a person studied computer science given that this was the person's choice. Naturally, by the axioms of counterfactuals (and common sense), this is the same as $E[Y \mid x_0]$; the interesting aspect is the contrast produced by considering the expectations in the sum that comes after. The quantity $E[Y_x \mid x_0]$ evokes a counterfactual that considers the expected earnings had a person who chose computer science studied some other degree x. Moreover, α_x is some weight assigned to a particular college degree x for the sake of comparison. While one could use uniform weights, a sensible choice could be the distribution of second choices made by people who in the end studied computer science.

With the target quantity (Equation (3)) in mind, suppose the available data consists of an observational study carried out in π^2 ($P^2(Z, X, W, Y)$) and a study in π^1 reporting the results of a large-scale scholarship program where students were given the change to pursue any degree they wanted regardless of socio-economical factors, which we represent with distribution $P^1(Z, X, W, Y; \sigma_X = \widehat{P}(X))$. No data from π^* is observed.

The expectations in Equation (3) are associated, respectively to the distributions $P^*(Y_{x_0} \mid x_0)$ and $P^*(Y_x \mid x_0)$ for $x \in \mathfrak{X}_X$[2]. For the reasons mentioned before, the first quantity is simply $P^*(Y \mid x_0)$, yet we have not observed data directly from π^*, hence we need to be clever about how to use the available data. As the conditions in Section 4 and the algorithm introduced in Section 5 (and some simplification) will allow us to derive:

$$P^*(y \mid x_0) = \sum\nolimits_z P^1(y \mid x_0, z; \sigma_X) P^2(x_0, z). \tag{4}$$

The second probability can be similarly obtained from available data as

$$P^*(y_x \mid x_0) = \sum\nolimits_z P^1(y \mid x, z; \sigma_X) P^2(x_0, z), \tag{5}$$

which together with Equation (4) allow us to estimate Equation (3) using the available distributions. ■

In general, it is useful to represent graphically the effect of an intervention. For instance, the causal diagram that represents the intervention σ_X is shown in Section 3. Compared to the original causal diagram (Figure 2(a)), the edges $Z \rightarrow X$ and $Z \dashleftarrow\!\dashrightarrow X$ have disappeared as they are removed by σ_X.

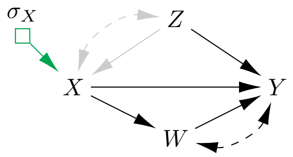


Figure 4: Causal diagram corresponding to the regime σ_X, denoted as $\mathcal{G}_{\sigma_X}$.

Data available in each domain is specified by $\mathbb{Z} = \{\mathbb{Z}^i \mid \pi^i \in \Pi\}$, where each $\mathbb{Z}^i = \{\sigma_{\mathbf{Z}_1}, \sigma_{\mathbf{Z}_2}, \ldots\}$, $\mathbf{Z}_j \subseteq \mathbf{V}$, corresponds to domain π^i. This means that distributions $\{P^i(\mathbf{V}; \sigma_{\mathbf{Z}_j}) \mid \mathbf{Z}_j \in \mathbb{Z}^i\}_{\mathbb{Z}^i \in \mathbb{Z}}$ are assumed to be available. Notice that $P^i(\mathbf{V}; \sigma_\emptyset) = P^i(\mathbf{V})$ describes the observational (non-interventional) distribution in domain π^i. In this example, $\mathbb{Z} = \{\mathbb{Z}^* = \emptyset, \mathbb{Z}^1 = \{\sigma_Z\}, \mathbb{Z}^2 = \{\sigma_W\}\}$.

The ability to infer the target counterfactual from the input distributions is formalized next.

Definition 3.3 (Counterfactual Transportability)**.** A query $P^*(\mathbf{y}_* \mid \mathbf{x}_*)$ is said to be transportable from $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$, if $P^*(\mathbf{y}_* \mid \mathbf{x}_*)$ is uniquely computable from the set of distributions $\mathbb{Z}$ for every assignment $(\mathbf{y}, \mathbf{x})$ and every set of models $\{\mathcal{M}^i\}_{\pi^i \in \Pi}$ inducing $\mathcal{G}^{\boldsymbol{\Delta}}$ and $\mathbb{Z}$. ■

There are counterfactual distributions with a special form that exploits the local structure described by the causal diagram called *counterfactual factors*, defined as follows:

Definition 3.4 (Counterfactual Factor (ctf-factor) (Correa et al., 2021))**.** A ctf-factor is a distribution of the form

$$P(w_{1[\mathbf{pa}_1]}, w_{2[\mathbf{pa}_2]}, \ldots, w_{l[\mathbf{pa}_l]}), \tag{6}$$

where each $W_i \in \mathbf{V}$ and there could be $W_i = W_j$ for some $i, j \in \{1, \ldots, l\}$. ■

Example 3.3 (ctf-factor)**.** Consider the causal diagram in Figure 2(a), then all of the following are ctf-factors:

$$P(y_{zxw}, w_x), \quad P(w_x, z), \text{ and } P(y_{z_0x_0w_0}, y_{z_1x_1w_1}). \tag{7}$$

In contrast, the following are not ctf-factors:

$$P(y_{zw}, w_x), \quad P(w_z, z), \text{ and } P(y_{z_0x_0w_0}, y_{z_1x_1}), \tag{8}$$

as they do not have the required form. ■

Lemma 3.1 (Counterfactual Factor Transportability)**.** *Let $\mathcal{G}^{\boldsymbol{\Delta i}}$ be the selection diagram based on $\mathcal{G}^i$ and Δ^i, and let $P^*(\mathbf{w}_*)$ be a ctf-factor, then $P^*(\mathbf{w}_*) = P^i(\mathbf{w}_*)$ if $\mathcal{G}^{\boldsymbol{\Delta i}}$ does not contain selection nodes S_{v_i} pointing to any variable in $V_i \in \mathbf{W}$, that is, $V_i \notin \Delta^i$.* ■

Lemma 3.1 allow us to determine whether the ctf-factor needed to estimate the query in the target domain can be

[2] Since we are trying to say something about π^*, we use the superscript $*$ to indicate the domain of the distributions of interest.

obtained from other domains, based on the assumptions encoded in the selection diagram.

Example 3.4 (Ctf-factor transportability)**.** Consider again the ctf-factors in Equation (7). We have,

$$P^*(y_{zxw}, w_x) = P^1(y_{zxw}, w_x), \text{ and} \tag{9}$$

$$\begin{aligned} P^*(y_{z_0x_0w_0}, y_{z_1x_1w_1}) &= P^1(y_{z_0x_0w_0}, y_{z_1x_1w_1}) \\ &= P^2(y_{z_0x_0w_0}, y_{z_1x_1w_1}). \end{aligned} \tag{10}$$

However, we cannot guarantee $P^*(w_x, z)$ to be equal to its counterpart $P^1(w_x, z)$ or $P^2(w_x, z)$, as $Z \in \Delta^1$ and $W \in \Delta^2$. ■

4. Graphical Condition for Counterfactual Transportability

Using the notion of ancestrality introduced in Definition 2.1, a counterfactual query can be decomposed in ctf-factors that we could try to transport. Given an arbitrary query $P^*(\mathbf{y}_*)$, let $\mathbf{D}_* = An(\mathbf{Y}_*)$ then

$$P^*(\mathbf{y}_*) = \sum\nolimits_{\mathbf{d}_* \setminus \mathbf{y}_*} P^*(\mathbf{d}_*). \tag{11}$$

Due to their ancestral relationship, $P^*(\mathbf{d}_*)$ is transportable if and only if $P^*(\mathbf{y}_*)$ is transportable, as stated next.

Lemma 4.1 (Non-transportability of the sum over an ancestral set)**.** *Let $P^*(\mathbf{d}_*)$ be a ctf-factor and let $\mathbf{Y}_* \subseteq \mathbf{D}_*$ be such that $\mathbf{D}_* = An(\mathbf{Y}_*)$. Then, $\sum_{\mathbf{d}_* \setminus \mathbf{y}_*} P^*(\mathbf{d}_*)$ is transportable from $\mathbb{Z}$ iff $P^*(\mathbf{d}_*)$ is transportable from $\mathbb{Z}$.* ■

Then, $P^*(\mathbf{d}_*)$ can be written in ctf-factor-form as

$$P^*(\mathbf{d}_*) = P^* \left(\bigwedge\nolimits_{D_{\mathbf{t}} \in \mathbf{D}_*} D_{\mathbf{pa}_d} = d \right), \tag{12}$$

where each $d = \mathbf{d}_* \cap \{D_{\mathbf{t}}\}$ and $\mathbf{pa}_d$ is determined for each $D_{\mathbf{t}} \in \mathbf{D}_*$ as the union of $\mathbf{t} \cap (\mathbf{Pa}_d \cap \mathbf{T})$ and $\mathbf{d}_* \cap (\mathbf{Pa}_d \setminus \mathbf{T})$.

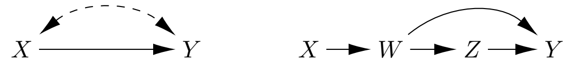
Example 4.1 (Ancestral set and ctf-factor)**.** Consider again the selection diagram in Figure 2(b) and the target quantity $P(y_x \mid x_0) = P(y_x, x_0)/P(x_0)$ in Example 3.2. The corresponding ancestral set is $\mathbf{D} = An(Y_x, W_x, X, Z)$ and the query can be written as

$$\begin{aligned} P^*(y_x, x_0) &= \sum\nolimits_{z,w} P^*(y_x, w_x, x_0, z) &(13) \\ &= \sum\nolimits_{z,w} P^*(y_{xwz}, w_x, x_{0[z]}, z), &(14) \end{aligned}$$

where Equation (14) is in ctf-factor form. ■

Moreover, Equation (12) can be factorized based on the c-component structure of the graph $\mathcal{G}^*[\mathbf{V}(\mathbf{D}_*)]$. Let $\mathbf{C}_1, \dots, \mathbf{C}_k$ be the c-components of this graph and define $\mathbf{C}_{j*} = \{D_{\mathbf{pa}_d} \in \mathbf{D}_* \mid D \in \mathbf{C}_j\}$ and $\mathbf{c}_{j*}$ as the values in $\mathbf{d}_*$ corresponding to $\mathbf{C}_{j*}$. Then $P^*(\mathbf{d}_*)$ decomposes as

$$P^*(\mathbf{d}_*) = \prod\nolimits_j P^*(\mathbf{c}_{j*}). \tag{15}$$



(a) Diagram where the ctf-factor $P(y_x, x')$ is inconsistent.

(b) Diagram where the ctf-factor $P(w_x, w'_{x'})$ is inconsistent.

Figure 5: Examples of causal diagrams and inconsistent ctf-factors derived from them.

Once the query of interest is in ctf-factor-form, the transportability question reduces to determining the transportability of smaller ctf-factors.

Example 4.2 (Ctf-factor factorization)**.** Following from Example 4.1 and Equation (14), the query factorizes as

$$P^*(y_x, x_0) = \sum\nolimits_{z,w} P^*(y_{xwz}, w_x) P^*(x_{0[z]}, z). \tag{16}$$

■

The question becomes whether ctf-factors corresponding to individual c-components can be transported from the available input. The following definition and theorem characterize the factors that can be transported from $\mathbb{Z}$ and $\mathcal{G}$.

Definition 4.1 (Inconsistent ctf-factor)**.** $P(\mathbf{w}_*)$ is an inconsistent ctf-factor if it is a ctf-factor, $\mathcal{G}[\mathbf{V}(\mathbf{W}_*)]$ has a single c-component, and one of the following situations hold:

(i) there exist $W_{\mathbf{t}} \in \mathbf{W}_*, Z \in \mathbf{T} \cap \mathbf{V}(\mathbf{W}_*)$ such that $z \in \mathbf{t}, z' \in \mathbf{w}_*$ and $z \neq z'$, or

(ii) there exist $W_{i[\mathbf{t}_i]}, W_{j[\mathbf{t}_j]} \in \mathbf{W}_*$ and, $T \in \mathbf{T}_i \cap \mathbf{T}_j$ such that $t \in \mathbf{t}_1, t' \in \mathbf{t}_2$ and $t \neq t'$.

■

Example 4.3 (Some inconsistent ctf-factors)**.** First, consider the causal diagram in Figure 5(a), where the ctf-factor $P(y_x, x')$ is inconsistent due to condition (i) in Definition 4.1. For another example, consider the ctf-factor $P(w_x, w'_{x'})$ is inconsistent due to condition (ii). ■

Theorem 4.1 (Transportability from $\mathbb{Z}$)**.** *A ctf-factor $P^*(\mathbf{w}_*)$ is transportable from $\mathbb{Z}$ only if it is consistent. If consistent, let $\mathbf{W} = \mathbf{V}(\mathbf{W}_*)$ and $\mathbf{W}' = Pa(\mathbf{W}) \setminus \mathbf{W}$; then $P^*(\mathbf{w}_*)$ is equal to $P^*_{\mathbf{w}'}(\mathbf{w})$ where $\mathbf{w}$ and $\mathbf{w}'$ are consistent with $\mathbf{w}_* \cup \bigcup_{\{W_{\mathbf{pa}_w} \in \mathbf{W}_*\}} \mathbf{pa}_w$, and $P^*(\mathbf{w}_*)$ is transportable from $\mathbb{Z}$ iff $P^*_{\mathbf{w}'}(\mathbf{w})$ is transportable from $\mathbb{Z}$.* ■

Example 4.4 (Ctf-factor transportability)**.** Consider each of the two factors in Example 4.2. First, $P^*(y_{xwz}, w_x)$ can be transported from $P^1(Z, X, W, Y; \sigma_X = \widehat{P}(X))$ using Lemma 3.1 since $Y, W \notin \Delta^1$:

$$\begin{aligned} P^*(y_{xwz}, w_x) = P^1(y_{xwz}, w_x) &= P^1_{xz}(y, w) &(17) \\ &= P^1(y, w \mid x, z; \sigma_X). &(18) \end{aligned}$$

Moreover, $P^*(x_{0[z]}, z)$ can be transported from

Algorithm 1 SIMPLIFY$(\mathbf{Y}_*, \mathbf{y}_*)$

```
Input: $\mathbf{Y}_*$ a set of counterfactual variables in $\mathbf{V}$ and $\mathbf{y}_*$ a set of values for $\mathbf{Y}_*$.
Output: An interventionally minimal event $\mathbf{Y}_* = \mathbf{y}_*$ without redundant subscripts or 0 if the counterfactual event is guaranteed to have probability 0.
let $\mathbf{Y}_* \leftarrow \|\mathbf{Y}_*\|$.
if there exists $Y_\mathbf{x} \in \mathbf{Y}_*$ with two or more different values in $\mathbf{y}_* \cap Y_\mathbf{x}$ or $Y_y \in \mathbf{Y}_*$ with $\mathbf{y}_* \cap Y_y \neq y$ then return 0.
if there exists $Y_\mathbf{x} \in \mathbf{Y}_*$ with two consistent values in $\mathbf{y}_* \cap Y_\mathbf{x}$ or $Y_y \in \mathbf{Y}_*$ with $\mathbf{y}_* \cap Y_y = y$ then remove repeated variables from $\mathbf{Y}_*$ and values $\mathbf{y}_*$.
return $\mathbf{Y}_* = \mathbf{y}_*$.
```

$P^2(Z, X, W, Y)$, because $X, Z \notin \Delta^2$, as

$$P^*(x_{0[z]}, z) = P^2(x_{0[z]}, z) = P^2_{wy}(x_0, z) \tag{19}$$

$$= P^2(x_0, z). \tag{20}$$

Both Equation (18) and Equation (20) follow from using σ-TR (Algorithm 4 in Appendix B) (Correa & Bareinboim, 2020) to transport $P^1_{xz}(y, w)$ and $P^2_{wy}(x_0, z)$ from $\mathbb{Z}$. ■

Given a counterfactual variable $Y_\mathbf{x}$ some values in $\mathbf{x}$ may be causally irrelevant to Y once the rest of $\mathbf{x}$ is fixed. In general, a counterfactual $Y_\mathbf{x}$ can be minimized with a process that we denote with the operator $\|\cdot\|$ as $\|Y_\mathbf{x}\| = Y_\mathbf{t}$, where $\mathbf{T} = \mathbf{X} \cap An(Y)_{\mathcal{G}_{\overline{\mathbf{X}}}}$ and $\mathbf{t} = \mathbf{x} \cap \mathbf{T}$. For a set of counterfactual variables $\mathbf{Y}_*$, minimization is done as $\|\mathbf{Y}_*\| = \{\|Y_\mathbf{x}\| \mid Y_\mathbf{x} \in \mathbf{Y}_*\}$. Moreover, such minimization could reveal repeated portions of a counterfactual event or inconsistencies that make the probability of the event to be zero. We capture these ideas in Algorithm 1.

4.1. Conditional Queries

The counterfactual query of interest could be a conditional one of the form $P^*(\mathbf{y}_* \mid \mathbf{x}_*)$. For this case, there exists a tight reduction from such conditional counterfactual to an unconditional one, described in (Correa et al., 2021).

To perform this reduction, one needs to look at the selection diagram paying special attention to variables after the conditioning bar that are also ancestors of those before. Let $\mathbf{X}_*(W_\mathbf{t}) = \mathbf{V}(\|\mathbf{X}_*\| \cap An(W_\mathbf{t}))$, that is, the primitive variables in $\mathbf{X}_*$ that are ancestors of $W_\mathbf{t}$.

Definition 4.2 (Ancestral components)**.** Let $\mathbf{W}_*$ be a set of counterfactual variables, $\mathbf{X}_* \subseteq \mathbf{W}_*$, and $\mathcal{G}$ be a causal diagram. Then the ancestral components induced by $\mathbf{W}_*$, given $\mathbf{X}_*$, are sets $\mathbf{A}_{1*}, \mathbf{A}_{2*}, \ldots$ that form a partition over $An(\mathbf{W}_*)$, made of unions of the ancestral sets $An(W_\mathbf{t})_{\mathcal{G}_{\underline{\mathbf{X}_*(W_\mathbf{t})}}}, W_\mathbf{t} \in \mathbf{W}_*$. Sets $An\big(W_{1[\mathbf{t}_1]}\big)_{\mathcal{G}_{\underline{\mathbf{X}_*(W_{1[\mathbf{t}_1]})}}}$ and $An\big(W_{2[\mathbf{t}_2]}\big)_{\mathcal{G}_{\underline{\mathbf{X}_*(W_{2[\mathbf{t}_2]})}}}$ are put together if they are not disjoint or there exists a bidirected arrow in $\mathcal{G}$ connecting variables in those sets. ■

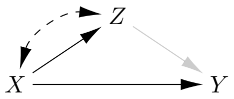


Figure 6: Causal diagram used to compute the ancestral sets of $\mathbf{Y}_* = \{Y_x\}$ given $\mathbf{X}_* = \{Z_x, X\}$, denoted $\mathcal{G}_{\mathbf{X}_*(Y_x)}$.

Next, we extend the result in (Correa et al., 2021) to the transportability task:

Lemma 4.2 (Conditional-marginal transportability reduction)**.** *Let $\mathbf{Y}_*, \mathbf{X}_*$ be two sets of counterfactual variables and let $\mathbf{D}_*$ be the set of variables in the same ancestral component, given $\mathbf{X}_*$, as any variable in $\mathbf{Y}_*$, then*

$$P^*(\mathbf{y}_*|\mathbf{x}_*) = \frac{\sum_{\mathbf{d}_* \setminus (\mathbf{y}_* \cup \mathbf{x}_*)} P^*(\bigwedge_{D_\mathbf{t} \in \mathbf{D}_*} D_{\mathbf{pa}_d} = d)}{\sum_{\mathbf{d}_* \setminus \mathbf{x}_*} P^*(\bigwedge_{D_\mathbf{t} \in \mathbf{D}_*} D_{\mathbf{pa}_d} = d)}, \tag{21}$$

where $\mathbf{pa}_d$ is consistent with $\mathbf{t}$, $\mathbf{d}_*$ and $\mathbf{x}_*(D_\mathbf{t})$, for each $D_\mathbf{t} \in \mathbf{D}_*$. Moreover, $P^*(\mathbf{y}_* \mid \mathbf{x}_*)$ is transportable from $\mathbb{Z}$ iff $P^*(\bigwedge_{D_\mathbf{t} \in \mathbf{D}_*} D_{\mathbf{pa}_d} = d)$ is transportable from $\mathbb{Z}$. ■

Example 4.5 (Conditional query simplification)**.** Recall the selection diagram in Figure 1 and consider the counterfactual $P^*(y_x \mid z_x, x')$. While the unconditional probability $P^*(y_x, z_x, x')$ is not transportable due to the inconsistent ctf-factor $P^*(z_x, x')$, the probability $P^*(y_x \mid z_x, x')$ is still transportable as we can see by using Lemma 4.2. Here $\mathbf{Y}_* = \{Y_x\}$, $\mathbf{X}_* = \{Z_x, X\}$, then $\mathbf{X}_*(Y_x) = \{Z_x\}$ and $\mathbf{D}_*$ can be computed as the ancestors of Y_x in the causal diagram in Figure 6, that is, $\{Y_x\}$. This gives

$$P^*(y_x \mid z_x, x') = P^*(y_{xz}), \tag{22}$$

which is equal to $P^*_{x,z}(y) = P^*(y \mid x, z)$ and could be obtained from a suitable $\mathbb{Z}$. ■

4.2. Nested Queries

The query of interest could involve counterfactuals with interventions that involve other counterfactuals, also called "nested counterfactuals". One example of this is the natural direct effect that we described in Example 1.1. Theorem 4 in (Correa et al., 2021) reduces the identifiability of a nested counterfactual to that of a non-nested one. We extend this reduction to the transportability case in the following.

Lemma 4.3 (Counterfactual Unnesting)**.** *Let $\widehat{\mathbf{X}}, \widehat{\mathbf{Z}}$ be any counterfactual variables (nested or non-nested) sets $\mathbf{X}, \mathbf{Z} \subseteq \mathbf{V}$. Then, for $\mathbf{Y} \subseteq \mathbf{V}$ disjoint from $\mathbf{X}$ and $\mathbf{Z}$ such that $\mathbf{X} \subseteq An(\mathbf{Y})_{\mathcal{G}_{\overline{\mathbf{Z}}}}$, $P^*(\mathbf{Y}_{\widehat{\mathbf{Z}}, \widehat{\mathbf{X}}} = \mathbf{y})$ is transportable from $\langle \mathbb{Z}, \mathcal{G}^{\boldsymbol{\Delta}} \rangle$ iff $P^*(\mathbf{Y}_{\widehat{\mathbf{Z}}, \mathbf{x}} = \mathbf{y}, \widehat{\mathbf{X}} = \mathbf{x})$ is transportable from $\langle \mathbb{Z}, \mathcal{G}^{\boldsymbol{\Delta}} \rangle$ for every $\mathbf{x}$, and given by*

$$P^*(\mathbf{Y}_{\widehat{\mathbf{Z}}, \widehat{\mathbf{X}}} = \mathbf{y}) = \sum_{\mathbf{x} \in \mathfrak{X}_\mathbf{X}} P^*(\mathbf{Y}_{\widehat{\mathbf{Z}}, \mathbf{x}} = \mathbf{y}, \widehat{\mathbf{X}} = \mathbf{x}). \tag{23}$$

■

Algorithm 2 $\text{CTFTRU}(\mathbf{Y}_*, \mathbf{y}_*, \mathbb{Z}, \mathcal{G}^{\boldsymbol{\Delta}})$

Input: $\mathcal{G}^{\boldsymbol{\Delta}} = \{\mathcal{G}^{\boldsymbol{\Delta i}}\}_{\pi \in \Pi}$ selection diagrams over $\mathbf{V}$; $\mathbf{Y}_*$ a set of counterfactual variables in $\mathbf{V}$; $\mathbf{y}_*$ a set of values for $\mathbf{Y}_*$; and available distribution specification $\mathbb{Z}$.
Output: $P^*(\mathbf{Y}_* = \mathbf{y}_*)$ in terms of available distributions or FAIL if not transportable from $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$.

```
 1: $(\mathbf{Y}_*, \mathbf{y}_*) \leftarrow \text{SIMPLIFY}(\mathbf{Y}_*, \mathbf{y}_*)$.
 2: let $\mathbf{W}_* \leftarrow An(\mathbf{Y}_*)$, and let $\mathbf{C}_{1*}, \ldots, \mathbf{C}_{k*}$ be corresponding ctf-factors in $\mathcal{G}^*[\mathbf{V}(\mathbf{W}_*)]$.
 3: if $\exists$ inconsistent $\mathbf{C}_i$ then return FAIL.
 4: for each $\mathbf{C}_i$ do
 5:     $Q \leftarrow \sigma\text{-TR}(\mathbf{C}_i, \mathbb{Z}, \mathcal{G}^{\boldsymbol{\Delta}})$.
 6:     if $Q$ is not FAIL then
 7:         let $P^*_{Pa(\mathbf{C}_i) \setminus \mathbf{C}_i}(\mathbf{C}_i) \leftarrow Q$.
 8:         let $\mathbf{c} \leftarrow (\mathbf{c}_{i*} \cup \bigcup_{C_\mathbf{t} \in \mathbf{C}_{i*}} \mathbf{pa}_c)$.
 9:         let $P^*(\mathbf{C}_{i*} = \mathbf{c}_{i*}) \leftarrow Q|_\mathbf{c}$.
10:         move to the next $\mathbf{C}_i$.
11:     end if
12: end for
13: if any $P^*(\mathbf{C}_{i*} = \mathbf{c}_{i*})$ was not transported from $\mathbb{Z}$ then return FAIL.
14: return $P^*(\mathbf{Y}_* = \mathbf{y}_*) \leftarrow \sum_{\mathbf{w}_* \setminus \mathbf{y}_*} \prod_i P^*(\mathbf{C}_{i*} = \mathbf{c}_{i*})$.
```

Algorithm 3 $\text{CTFTR}(\mathbf{Y}_*, \mathbf{y}_*, \mathbf{X}_*, \mathbf{x}_*, \mathbb{Z}, \mathcal{G}^{\boldsymbol{\Delta}})$

Input: $\mathcal{G}^{\boldsymbol{\Delta}}$ causal diagram over variables $\mathbf{V}$; $\mathbf{Y}_*, \mathbf{X}_*$ a set of counterfactual variables in $\mathbf{V}$; $\mathbf{y}_*, \mathbf{x}_*$ a set of values for $\mathbf{Y}_*$ and $\mathbf{X}_*$; and available distribution specification $\mathbb{Z}$.
Output: $P^*(\mathbf{Y}_*{=}\mathbf{y}_* \mid \mathbf{X}_*{=}\mathbf{x}_*)$ in terms of available distributions or FAIL if not transportable from $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$.

```
1: Let $\mathbf{A}_{1*}, \mathbf{A}_{2*}, \ldots$ be the ancestral components of $\mathbf{Y}_* \cup \mathbf{X}_*$ given $\mathbf{X}_*$.
2: Let $\mathbf{D}_*$ be the union of the ancestral components containing a variable in $\mathbf{Y}_*$ and $\mathbf{d}_*$ the corresponding set of values.
3: let $Q \leftarrow \text{CTFTRU}(\bigcup_{D_\mathbf{t} \in \mathbf{D}_*} \mathbf{D}_{\mathbf{pa}_d}, \mathbf{d}_*, \mathbb{Z}, \mathcal{G}^{\boldsymbol{\Delta}})$.
4: return $\sum_{\mathbf{d}_* \setminus (\mathbf{y}_* \cup \mathbf{x}_*)} Q / \sum_{\mathbf{d}_* \setminus \mathbf{x}_*} Q$.
```

Example 4.6 (Conditional and nested)**.** In Example 1.1, one of the target quantities is $P^*(y_{x', Z_x})$, which is a nested counterfactual. Using Lemma 4.3 we get

$$P^*(y_{x', Z_x}) = \sum\nolimits_z P^*(y_{x'z}, z_x) \tag{24}$$

and the problem is reduced to transporting $P^*(y_{x'z}, z_x)$ from $\mathbb{Z}$. ■

5. A Sound and Complete Algorithm for Counterfactual Transportability

Using the factorization of the query described in the previous section and Lemma 3.1, we propose CTFTR (Algorithm 3), an algorithm that determines the transportability of a probability of the form $P^*(\mathbf{y}_* \mid \mathbf{x}_*)$ corresponding to a target domain π^* from a collection of observational and experimental distributions $\mathbb{Z}$, and a selection diagram $\mathcal{G}^{\boldsymbol{\Delta}}$. When the query is transportable, the algorithm outputs an expression for $P^*(\mathbf{y}_* \mid \mathbf{x}_*)$ in terms of the specified distributions and FAIL if the query is not transportable from such input in $\mathcal{G}^{\boldsymbol{\Delta}}$.

In CTFTR, line 1 computes the ancestral components associated with the query and line 2 determines the set $\mathbf{D}_*$. In line 3 invokes the subroutine CTFTRU to transport the numerator of Equation (21) that is later returned in line 14.

The subroutine CTFTRU can be used to transport non-conditional counterfactuals. Line 1 invokes SIMPLIFY (Algorithm 1) which makes the query interventionally minimal, removes redundant variables, or determines if the event has probability zero. Line 2 computes the ancestral set and ctf-factors corresponding to the query and line 3 checks whether any of them is inconsistent, in which case the algorithm fails. The loop in line 4 tries to transport every ctf-factor from the available input. Line 5 calls σ-TR (Algorithm 4 in Appendix B) to transport $Q = P^*_{Pa(\mathbf{C}_i) \setminus \mathbf{C}_i}(\mathbf{C}_i)$ from $\mathbb{Z}$ and $\mathcal{G}^{\boldsymbol{\Delta}}$. If successful, line 8 creates a set with the values that are used to evaluate Q (in line 9) so that it is equal to the ctf-factor $P^*(\mathbf{C}_{i*} = \mathbf{c}_{i*})$. Line 10 moves on to the next factor when the current one has been transported. Finally, line 13 fails if any of the ctf-factors is not transportable from the input or line 14 the corresponding expression.

Theorem 5.1 (CTFTR completeness)**.** *A counterfactual probability $P^*(\mathbf{y}_* \mid \mathbf{x}_*)$ is transportable from $\mathbb{Z}$ and $\mathcal{G}^{\boldsymbol{\Delta}}$ if and only if* CTFTR *returns an expression for it. Moreover,* CTFTR *decides this task in time $O(n^4 z)$ where $n = |\mathbf{V}|$ and $z = \sum_{\mathbb{Z}^i \in \mathbb{Z}} |\mathbb{Z}^i|$.* ■

6. Conclusions

In this paper, we studied the problem of transporting counterfactual quantities from a combination of observational and experimental distributions obtained from one or more heterogeneous domains. Using a decomposition based on ctf-factors, we characterized the transportability of such factors between domains (Lemma 3.1) and used it to establish a sufficient and necessary graphical condition for the transportability of a given counterfactual query (Lemma 4.1, Theorem 4.1). We considered conditional and nested counterfactuals, and then provided tight reductions for those types of queries (Lemmas 4.2 and 4.3). In Section 5, we developed a sound and complete algorithm (Algorithm 3) for the counterfactual transportability task (Theorem 5.1). In other words, this means that the target counterfactual quantity is not transportable whenever the algorithm returns failure, unless further parametric assumptions are made about the underlying generating model. The problem of generalizing and fusing information across settings in pervasive in the sciences. There are many questions in the empirical sciences that can be formulated as counterfactual queries, with data coming from observations and experiments in heterogeneous domains, which constitute counterfactual transportability tasks. We hope the language, conditions, and algorithms developed in this paper serve as stepping stones in the modeling and the solution of these problems.

ACKNOWLEDGEMENTS

Elias Bareinboim and Juan D. Correa were supported in part by the NSF, ONR, DoE, Amazon, JP Morgan, and The Alfred P. Sloan Foundation. Sanghack Lee was supported by the New Faculty Startup Fund from Seoul National University.

References

Avin, C., Shpitser, I., and Pearl, J. Identifiability of Path-Specific Effects. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence* {*IJCAI-05*}, pp. 357–363, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.

Bareinboim, E. and Pearl, J. Transportability of causal effects: Completeness Results. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, CA, 2012a. Department of Computer Science, University of California, Los Angeles.

Bareinboim, E. and Pearl, J. Causal Inference by Surrogate Experiments: z-Identifiability. In Murphy, N. d. F. and Kevin (eds.), *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pp. 113–120. AUAI Press, 2012b.

Bareinboim, E. and Pearl, J. A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference*, 1(1):107–134, 2013.

Bareinboim, E. and Pearl, J. Transportability from Multiple Environments with Limited Experiments: Completeness Results. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems 27*, pp. 280–288. Curran Associates, Inc., 2014.

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

Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pp. 507–556. Association for Computing Machinery, New York, NY, USA, 1st edition, 2022.

Correa, J. D. and Bareinboim, E. From Statistical Transportability to Estimating the Effects of Stochastic Interventions. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2019.

Correa, J. D. and Bareinboim, E. General Transportability of Soft Interventions: Completeness Results. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Correa, J. D., Lee, S., and Bareinboim, E. Nested Counterfactual Identification from Arbitrary Surrogate Experiments. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Dawid, A. P. Influence diagrams for causal modelling and inference. *International Statistical Review*, 70:161–189, 2002.

Dawid, A. P. Statistical Causality from a Decision-Theoretic Perspective. *Annual Review of Statistics and Its Application*, 2(1):273–303, 2015.

Huang, Y. and Valtorta, M. Pearl's Calculus of Intervention Is Complete. In T.S.˜Richardson, R. D. a. (ed.), *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pp. 217–224, Corvallis, OR, 2006. AUAI Press.

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

Lee, S. and Honavar, V. Causal Transportability of Experiments on Controllable Subsets of Variables: z-Transportability. In Nicholson, A. and Smyth, P. (eds.), *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 361–370. AUAI Press, 2013a.

Lee, S. and Honavar, V. m-Transportability: Transportability of a Causal Effect from Multiple Environments. In desJardins, M. and Littman, M. (eds.), *Proceedings of the Twenty-Seventh National Conference on Artificial Intelligence*, pp. 583–590, Menlo Park, CA, 2013b. AAAI Press.

Lee, S., Correa, J. D., and Bareinboim, E. General Identifiability with Arbitrary Surrogate Experiments. In *Proceedings of the Thirty-Fifth Conference Annual Conference on Uncertainty in Artificial Intelligence*, Corvallis, OR, 2019. AUAI Press.

Lee, S., Correa, J. D., and Bareinboim, E. Generalized Transportability: Synthesis of Experiments from Heterogeneous Domains. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, Menlo Park, CA, 2020. AAAI Press.

Pearl, J. A probabilistic calculus of actions. In de Mantaras, R. L. and D.˜Poole (eds.), *Uncertainty in Artificial Intelligence 10*, pp. 454–462. Morgan Kaufmann, San Mateo, CA, 1994.

Pearl, J. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

Pearl, J. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition, 2000.

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

Pearl, J. and Bareinboim, E. Transportability of Causal and Statistical Relations: A Formal Approach. In *Proceedings of the Twenty-Fifth Conference on Artificial Intelligence (AAAI-11)*, pp. 247–254, Menlo Park, CA, 8 2011.

Pearl, J. and Mackenzie, D. *The Book of Why*. Basic Books, New York, 2018.

Powdthavee, N., Lekfuangfu, W. N., and Wooden, M. The Marginal Income Effect of Education on Happiness: Estimating the Direct and Indirect Effects of Compulsory Schooling on Well-Being in Australia. Technical Report 7365, Institute for the Study of Labor (IZA), Bonn, 2013.

Shpitser, I. and Pearl, J. Identification of Joint Interventional Distributions in Recursive semi-Markovian Causal Models. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*, volume 2, pp. 1219–1226, 2006.

Shpitser, I. and Pearl, J. What Counterfactuals Can Be Tested. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, pp. 352–359. AUAI Press, Vancouver, BC, Canada, 2007.

Tian, J. Identifying Dynamic Sequential Plans. In *In Proceedings of the Twenty-Fourth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-08)*, pp. 554–561, Corvallis, Oregon, 2008. AUAI Press.

Tian, J. and Pearl, J. A General Identification Condition for Causal Effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002)*, pp. 567–573, Menlo Park, CA, 2002a. AAAI Press/The MIT Press.

Tian, J. and Pearl, J. On the Testable Implications of Causal Models with Hidden Variables. *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI-02)*, pp. 519–527, 2002b.

Tian, J. and Pearl, J. A General identification condition for causal effects. Technical Report R-290-A, Department of Computer Science, University of California, Los Angeles, CA, 2003.