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) \( M \) 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 \( M \) 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 exist 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 to extrapolate and fuse experimental knowledge across settings has been formally studied in the causal inference literature under the rubric...
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 selection diagram (to be defined in Section 3) 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 \( E[Y \mid Z = z, X = x] - E[Y \mid Z = z, \text{do}(X)] \) (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 X 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 \( Y (\blacktriangledown \rightarrow Y) \), 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:

\[
\sum_z \left( P^*(y \mid x', z) - P^*(y \mid x, z) \right) \frac{P(z \mid \text{do}(x))}{\text{from Australia}} \text{ from the US}
\]

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 implies that the quantity is not transportable without further assumptions.

Our results generalize the identification results in (Correa et al., 2021) to a transportability setting, where available data comes from different domains and different experimental conditions, also generalizing results in (Correa & Bareinboim, 2020) to counterfactual queries. The remaining of the paper is organized as follows. Section 2 introduces structural causal models and required notions about counterfactuals. Section 3 reviews transportability and the graphical models used for it. It also defines the counterfactual transportability task and a building block to solve it. Section 4 generalizes previous work to the context of the transportability task, including conditional and nested counterfactuals. Section 5 presents a sound and complete algorithm for the proposed task. Finally, Section 6 ends with the conclusions.
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 \( x \) sets of values. The domain of a variable \( X \) is denoted by \( \mathcal{X}_X \). Two values \( x \) and \( z \) are said to be consistent if they share the common values for \( \mathbf{X} \cap \mathbf{Z} \). We also denote by \( x \setminus Z \) the value of \( \mathbf{X} \setminus \mathbf{Z} \) consistent with \( x \) and by \( x \cap \mathbf{Z} \) the subset of \( x \) 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 \( \mathcal{V}(\mathcal{G}) \) the set of vertices (i.e., variables) in a graph \( \mathcal{H} \). Given a graph \( \mathcal{G} \), \( \mathcal{G}_{\mathbf{X}} \) 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 \( Pa(\mathbf{X})_\mathcal{G} := \bigcup_{x \in \mathbf{X}} Pa(x)_\mathcal{G} \). Similarly, we define the set of ancestors \((An(\cdot))\) and descendants \((De(\cdot))\), which include the argument itself.

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 \((U, \mathbf{V}, \mathcal{F}, P(\mathbf{u}))\), where \( U \) 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 \( U_i \subseteq U \) and a set of endogenous variables \( Pa_i \subseteq \mathbf{V} \setminus \{V_i\} \) to the domain of \( V_i \). The uncertainty is encoded through a probability distribution over the exogenous variables, \( P(U) \). An SCM 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 \( V_j \in Pa_i \), and a bidirected edge \((V_i \leftrightarrow V_j)\) for every pair \( V_i, V_j \in \mathbf{V} \) such that \( U_i \cap U_j \neq \emptyset \) (\( V_i \) 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 \textit{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 \( \hat{\mathbf{X}} \) be a collection of functions \( \{X : \mathcal{G}_{\hat{\mathbf{X}}} \rightarrow \mathcal{X}_X\} \) for some \( \mathbf{X} \subseteq \mathbf{V} \) and some set of unobservables \( \hat{\mathbf{X}} \) disjoint from the original \( \mathbf{U} \). Then, an intervention can be described by some \( \hat{\mathbf{X}} \) that induces a derived model \( \mathcal{M}_{\mathbf{X}} \) where each \( f_X \) has been replaced by \( \hat{f}_X \). Then, \( Y_{\mathbf{X}}(u, \hat{u}_X) \) is called the \textit{potential response} of \( \mathbf{Y} \) to \( \mathbf{X} = \hat{\mathbf{X}} \), and is defined as the solution of \( \mathbf{Y} \), for a particular \((u, \hat{u}_X)\), in the derived model \( \mathcal{M}_{\mathbf{X}} \).

The most simple and common type of intervention considered consists on one that fixes \( f_X \) to a constant value \( x \in \mathcal{X}_X \), denoted \( do(X = x) \). Other interventions may replace \( f_X \) with some \( \hat{X} \) which is a function of only observable variables (conditional intervention) or even a function with both observable and unobservable (from \( \hat{\mathbf{U}}_X \)) arguments (stochastic intervention). In any case, we assume that \( \hat{\mathbf{X}} \) is not a function of any of the original \( \mathbf{U} \).

We use \( \mathbf{W}_s \) to denote sets of arbitrary counterfactual variables\(^1\). Let \( \mathbf{W}_s = \{W_1[T_1], W_2[T_2], \ldots\} \) represent a set of counterfactual variables such that \( W_i \in \mathbf{V} \) and \( T_i \subseteq \mathbf{V} \) for \( i = 1, \ldots, l \). Define \( \mathcal{V}(\mathbf{W}_s) = \{W \in \mathbf{V} \mid W \not\in \mathbf{W}_s\} \), that is, the set of observables that appear in \( \mathbf{W}_s \). Let \( \mathbf{w}_s \) represent a vector of values, one for each variable in \( \mathbf{W}_s \).

If all variables in the expression have the same subscript, we could write \( P(W_1[x], W_2[x], \ldots) \) as \( P_x(W_1, W_2, \ldots) \). Also, we will write \( P(Y_* = y_*) \) simply as \( P(y_*) \) 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 ancestry and causal relevance was formalized by (Correa et al., 2021) with the following definition.

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

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

**Example 2.1** (College Degree and Earnings — Counterfactual ancestors). For instance, consider the causal di-

---

\(^1\)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, e.g., \( W_1[x], W_2[x], \ldots \).
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.

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

**Definition 2.2** (Counterfactual Factor (ctf-factor) (Correa et al., 2021)). A ctf-factor is a distribution of the form
\[ P(w_{i_1}[pa_i], w_{i_2}[pa_i], \ldots, w_{i_l}[pa_i]), \]  
where each \( W_i \in V \) and there could be \( W_i = W_j \) for some \( i, j \in \{1, \ldots, l\} \).

**Example 2.2** (ctf-factor). Consider the causal diagram in Figure 2(a), then all of the following are ctf-factors:
\[ P(y_{xw}, w_x), P(w_x, z), \text{ and } P(y_{z0x0w0}, y_{z1x1}) \] \( (3) \)

In contrast, the following are not ctf-factors:
\[ P(y_{zw}, w_x), P(w_x, z), \text{ and } P(y_{z0x0w0}, y_{z1x1}) \] \( (4) \)
as they do not have the required form.

3. Transporting Counterfactual Relationships Across Domains

In this section, we review transportability and selection diagrams, including some examples. Then, we define the counterfactual transportability task and provide a lemma that serves as a basic block to solve it.

Our goal is to assess a counterfactual quantity using assumptions encoded in the form of a graph and observationa-
and Figure 3(b). Formally, this diagram encodes $\Delta^1 = \{Z\}$ and $\Delta^2 = \{W\}$. In the context of Example 2.1, $\Delta^1$ advertises possible differences in the socio-economic conditions ($Z$) between the domains $\pi^1$ and $\pi^*$. Meanwhile, $\Delta^2$ indicates differences regarding occupation ($W$) due, perhaps, to $\pi^2$ being a region where the occupational profile of several college program alumni differs with respect to $\pi^*$.

In this paper, our goal is to assess a counterfactual quantity in a target domain $\pi^*$ 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 $\sigma_X = g(W)$ sets the value of $X$ according to a deterministic function $g : X_W \rightarrow X$ for some set $W \subseteq V \setminus \text{De}(X)$. A stochastic intervention $\sigma_X = \hat{P}(X \mid W)$ sets the value of $X$ according to a given probability distribution conditional on a similar set $W$. A distribution affecting a set of variables $X$ is analogously represented as $\sigma_X$. Depending on the intervention, the causal diagram $G_{\sigma_X}$ corresponds to the SCM $M^t$ after being intervened with $\sigma_X$.

**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 earnings of people in $\pi^*$. We use, however, not interested in the effect that this has on the average person in $\pi^*$ but on those who choose to pursue this degree on their own. To do this, we could consider the following quantity:

$$E[Y_{x_0} \mid x_0] - \sum_{x} \alpha_x E[Y_x \mid x_0].$$  

The first expectation in Equation (5) refers to the expected value of earnings, had a person studied computer science; the second 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, $\alpha_x$ 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 (5)) in mind, suppose the available data consists of an observational study carried out in $\pi^2$ ($P^2(Z, X, W, Y)$) and a study in $\pi^1$ reporting the results of a large-scale scholarship program where students were given the chance to pursue any degree they wanted regardless of socio-economical factors, which we represent with distribution $P^1(Z, X, W, Y; \sigma_X = \hat{P}(X))$. No data from $\pi^*$ is observed.

The expectations in Equation (5) are associated, respectively to the distributions $P^*(Y_{x_0} \mid x_0)$ and $P^*(Y_x \mid x_0)$ for $x \in X \setminus X^2$. For the reasons mentioned before, the first quantity is simply $P^*(Y \mid x_0)$, yet we have not observed data directly from $\pi^*$, 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:

$$P^*(y \mid x_0) = \sum_z P^1(x \mid x_0, z; \sigma_X)P^2(x_0, z).$$  

(6)

The second probability can be similarly obtained from available data as

$$P^*(y_x \mid x_0) = \sum_z P^1(y \mid x, z; \sigma_X)P^2(x_0, z),$$  

(7)

which together with Equation (6) allow us to estimate Equation (5) 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 $\sigma_X$ is shown in Section 3. Compared to the original causal diagram (Figure 2(a)), the edges $Z \rightarrow X$ and $Z \leftrightarrow X$ have disappeared as they are removed by $\sigma_X$.

Data available in each domain is specified by $Z = \{Z_i \mid \pi^i \in \Pi\}$, where each $Z_i = \{\sigma_{Z_1}, \sigma_{Z_2}, \ldots\}$, $Z_i \subseteq V$, corresponds to domain $\pi^i$. This means that distributions $\{P^i(V; \sigma_{\emptyset}) > 0 \mid Z_j \in Z^i\}$ are assumed to be available. Notice that $P^i(V; \sigma_{\emptyset}) = P^i(V)$ describes the observational (non-interventional) distribution in domain $\pi^i$. In this example, $Z = \{Z^* = \emptyset, Z^1 = \{\sigma_{Z_1}\}, Z^2 = \{\sigma_{W}\}\}$.

We are now ready to formally define the task to be solved in this paper.

**Definition 3.3** (Counterfactual Transportability). A query $P^*(y_x \mid x_0)$ is said to be transportable from $(G^\Delta, Z)$, if $P^*(y_x \mid x_0)$ is uniquely computable from the set of distributions $Z$ for every assignment $(y, x)$ and every set of models $\{M^t\}_{\pi^i \in \Pi}$ inducing $G^\Delta$ and $Z$.  

Since we are trying to say something about $\pi^*$, we use the superscript * to indicate the domain of the distributions of interest.
Counterfactual Transportability: A Formal Approach

To understand the conditions for counterfactual transportability to be feasible, as well as devising a systematic approach to solve it, we establish transportability at the level of a ctf-factor, that serves as our unit of analysis.

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

Lemma 3.1 allows us to determine whether the ctf-factor needed to estimate the query in the target domain can be obtained from other domains, based on the assumptions encoded in the selection diagram.

**Example 3.3** (Ctf-factor transportability). Consider again the ctf-factors in Equation (3). We have,

$$P^*(y_{xuw}, w_x) = P^1(y_{xuw}, w_x) \quad \text{and} \quad P^*(y_{z0x0w0}, y_{z1x1w1}) = P^2(y_{z0x0w0}, y_{z1x1w1}).$$

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. Counterfactual Transportability**

To solve counterfactual transportability systematically, we will generalize results for the identifiability of counterfactuals in a single domain setting based on ctf-factors. The goal is to decompose any given counterfactual query in terms of ctf-factors which transportability from available data can be judged independently.

For this purpose, we follow the strategy in (Correa et al., 2021) and use the notion of ancestrality introduced in Definition 2.1, to write the query in terms of all the variables relevant for the analysis. Given an arbitrary query $P^*(y_*)$, let $D_* = An(Y_*)$ then

$$P^*(y_*) = \sum_{d_* \not\rightarrow y_*} P^*(d_*) .$$

Next, we characterize the relationship between the transportability of $P^*(d_*)$ and $P^*(y_*)$.

**Lemma 4.1** (Transportability of the sum over an ancestral set). Let $P^*(d_*)$ be a ctf-factor and let $Y_* \subseteq D_*$ be such that $D_* = An(Y_*)$. Then, $\sum_{d_* \not\rightarrow y_*} P^*(d_*)$ is transportable from $Z$ iff $P^*(d_*)$ is transportable from $Z$.

Then, $P^*(d_*)$ can be written in ctf-factor-form as

$$P^*(d_*) = P^* \left( \bigwedge_{D_i \in D_*} D_{pa_i} = d \right) ,$$

where each $d = d_* \cap \{D_k\}$ and $pa_i$ is determined for each $D_k \in D_*$ as the union of $t \cap (Pa_i \cap T)$ and $d_i \cap (Pa_i \cap \overline{T})$.

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

$$P^*(y_{x}, x_0) = \sum_{z,w} P^*(y_{z}, w_{z}, x_0, z) \quad \text{and} \quad \sum_{z,w} P^*(y_{xwz}, w_{x0[z]}, z) ,$$

where Equation (13) is in ctf-factor form.

The next step in the solution strategy is to factorize the righthand-side of Equation (11) according to the c-component structure of the graph $G'[V(D_*)]$. Let $C_1, \ldots, C_k$ be the c-components of this graph and define $C_{js} = \{D_{pa_i} \in D_* | D \in C_j\}$ and $c_{js}$ as the values in $d_*$ corresponding to $C_{js}$. Then $P^*(d_*)$ decomposes as

$$P^*(d_*) = \prod_j P^*(c_{js}) .$$

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 (13), the query factorizes as

$$P^*(y_{x}, x_0) = \sum_{z,w} P^*(y_{xwz}, w_{x} P^*(x_{0[z]}, z) .$$

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 $Z$ and $G$.

**Definition 4.1** (Inconsistent ctf-factor). $P(w_*)$ is an inconsistent ctf-factor if it is a ctf-factor, $G'[V(W_*])$ has a single c-component, and one of the following situations hold:

(i) there exist $W_t \in W_*, Z \in \overline{T} \cap V(W_*)$ such that $z \in t, z' \in w_*$ and $z \neq z'$, or
(ii) there exist $W_{t(i)}, W_{t[j]} \in W_*$ and $T \in \overline{T} \cap T_j$ such that $t \in t_i, t' \notin t_j$ and $t \neq t'$. 

Figure 4: Causal diagram corresponding to the regime $\sigma_X$, denoted as $G_{\sigma_X}$. 

**Counterfactual Transportability**
Theorem 4.1 (Transportability from Z). A ctf-factor $P^*(w_s)$ is transportable from Z only if it is consistent. If consistent, let $W = V(W_s)$ and $W' = Pa(W) \backslash W$; then $P^*(w_s)$ is equal to $P_{w_s}(w) \cap W'$ where $w$ and $w'$ are consistent with $w_s \cup \{w_{pa_w} \in W_s \}$. $P^*(w_s)$ is transportable from Z iff $P_{w_s}(w)$ is transportable from 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 $P^1(Z, X, Y; x \leftarrow P(X))$ using Lemma 3.1 since $Y, W \notin \Delta^1$:

$$P^*(y_{xwz}, w_x) = P^1(y_{xwz}, w_x) = P^1(y, w) = P(y, w | x, y; x \leftarrow P(X)). \hspace{1cm} (16)$$

Moreover, $P^*(x_{t|z}, z)$ can be transported from $P^2(Z, W, Y, X)$, because $X, Z \notin \Delta^2$, as

$$P^*(x_{t|z}, z) = P^2(x_{t|z}, z) \hspace{1cm} (17)$$

Algorithm 1 \textbf{Simplify}$(Y_s, y_s)$

\textbf{Input}: $Y_s$ a set of counterfactual variables in V and $y_s$ a set of values for $Y_s$.

\textbf{Output}: An interventional minimal event $Y_s = y_s$ without redundant subscripts or 0 if the counterfactual event is guaranteed to have probability 0.

1: let $Y_s \leftarrow |Y_s|$.
2: if there exists $Y_s \in Y_s$ with two or more different values in $y_s \cap Y_s$ or $y_s \in Y_s$ with $y_s \cap Y_s \neq y_s$ then return 0.
3: if there exists $Y_s \in Y_s$ with two consistent values in $y_s \cap Y_s$ or $y_s \in Y_s$ with $y_s \cap Y_s = y_s$ then remove repeated variables from $Y_s$, and values $y_s$.
4: return $Y_s = y_s$.

4.1 Conditional Queries

The counterfactual query of interest could be a conditional one of the form $P^*(Y_s | x_s)$. 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 $X_s(W_t) = V(|X_s| \cap An(W_t))$, that is, the primitive variables in $X_s$, that are ancestors of $W_t$.

Definition 4.2 (Ancestral components). Let $W_s$ be a set of counterfactual variables, $X_s \subseteq W_s$, and $G$ be a causal diagram. Then the ancestral components induced by $W_s$, given $X_s$, are sets $A_1, A_2, \ldots$ that form a partition over $An(W_s)$, made of unions of the ancestors sets $An(W_t) \cap G_{X_s \cap An(W_t)}$ and $An(W_t) \cap G_{X_s \cap An(W_t)}$ are put together if they are not disjoint or there exists a directed arrow in G connecting variables in those sets.

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

Lemma 4.2 (Conditional-marginal transportability reduction). Let $Y_s, X_s$ be two sets of counterfactual variables and let $D_s$ be the set of variables in the same ancestral component, given $X_s$, as any variable in $Y_s$, then

$$P^*(y_s | x_s) = \frac{\sum_{d_s \backslash \{y_s \cup x_s\}} P^*(\nabla D_s \in D_s, d_{pa_d} = d)}{\sum_{d_s \backslash \{x_s\}} P^*(\nabla D_s \in D_s, d_{pa_d} = d)}, \hspace{1cm} (20)$$

where $pa_d$ is consistent with $t, d_s$, and $x_s(D_s)$. for each $D_s \in D_s$. Moreover, $P^*(y_s | x_s)$ is transportable from Z iff $P^*(\nabla D_s \in D_s, d_{pa_d} = d)$ is transportable from Z.

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

\[
P^*(y_x | z_x, x') = P^*(y_{xz}), \quad (21)
\]

which is equal to \( P^*_x(y_x) = P^*(y | x, z) \) and could be obtained from a suitable \( Z \).

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 \( \tilde{X}, \tilde{Z} \) be any counterfactual variables (nested or non-nested) sets \( X, Z \subseteq V \). Then, for \( Y \subseteq V \) disjoint from \( X \) and \( Z \) such that \( \tilde{X} \subseteq An(Y)_{\tilde{X}} \) \( P^*(Y_{\tilde{Z}, \tilde{X}} = y, \tilde{X} = x) \) is transportable from \((\tilde{Z}, G^{\Delta})\) iff \( P^*(Y_{\tilde{Z}, \tilde{X}} = y, \tilde{X} = x) \) is transportable from \((\tilde{Z}, G^{\Delta})\) for every \( x \), and given by

\[
P^*(Y_{\tilde{Z}, \tilde{X}} = y) = \sum_{x \in X} P^*(Y_{\tilde{Z}, \tilde{X}} = y, \tilde{X} = x). \quad (22)
\]

**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

\[
P^*(y, x, z_x) = \sum_z P^*(y, x', z_x) \quad (23)
\]

and the problem is reduced to transporting \( P^*(y, x', z_x) \) from \( Z \).

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 \( P^*(y_x | x, x) \) corresponding to a target domain \( \pi^* \) from a collection of observational and experimental distributions \( Z \), and a selection diagram \( G^{\Delta} \). When the query is transportable, the algorithm outputs an expression for \( P^*(y_x | x, x) \) in terms of the specified distributions and fails if the query is not transportable from such input in \( G^{\Delta} \).

In CTFTR, line 1 computes the ancestral components associated with the query and line 2 determines the set \( D_x \). In line 3 invokes the subroutine CTFTR to transport the numerator of Equation (20) that is later returned in line 14.

The subroutine CTFTR 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 \( \sigma-TR \) (Algorithm 4 in Appendix B) to transport \( Q = P_{P(\sigma(x) \backslash C, C)}(C) \) from \( Z \) and \( G^{\Delta} \). 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 \( P^*(C = c) \). 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 \( P^*(y_x | x, x) \) is transportable from \( Z \) and \( G^{\Delta} \) if and only if CTFTR returns an expression for it. Moreover,
Algorithm 3 $\text{CTFTR}(Y_*, y_*, X_*, x_*, Z, g^\Delta)$

**Input:** $g^\Delta$ causal diagram over variables $V$; $Y_*$, $X_*$, a set of counterfactual variables in $V$; $y_*$, $x_*$, a set of values for $Y_*$ and $X_*$; and available distribution specification $Z$.

**Output:** $P^*(Y_* = y_* \mid X_* = x_*)$ in terms of available distributions or $\text{FAIL}$ if not transportable from $(g^\Delta, Z)$.

1. Let $A_{1*}, A_{2*}, \ldots$ be the ancestral components of $Y_* \cup X_*$ given $X_*$.
2. Let $D_*$ be the union of the ancestral components containing a variable in $Y_*$ and $d_*$, the corresponding set of values.
3. Let $Q \leftarrow \text{CTFTRU}(\bigcup_{D \in D_*} Dpa_* \cup d_*, d_*, Z, g^\Delta)$.
4. Return $\sum_{d_* \backslash (y_* \cup x_*)} Q / \sum_{d_* \backslash x_*} Q$.

$\text{CTFTR}$ decides this task in time $O(n^4z)$ where $n = |V|$ and $z = \sum_{Z' \in Z} |Z'|$.

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 is pervasive in the sciences. Many questions in the empirical sciences 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


A. Graphical Condition for Transportability

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

Proof. A ctf-factor can be expressed, as

$$P^*(w_\ast) = \sum_{\{u|\mathcal{W}_\ast(u) = w_\ast\}} P^*(u). \tag{24}$$

The predicate $\mathcal{W}_\ast(u) = w_\ast$ depends on the functions $f_w, W \in \mathcal{W}$ and $\mathcal{W}_\ast(u)$ only depends on $P^*(\mathcal{U}(\mathcal{W}))$, where $\mathcal{U}(\mathcal{W}) = \bigcup_{w \in \mathcal{W}} U_w$. Since no $V_i \in \mathcal{W} \cap \Delta^i$, we have $f_w^\ast = f_w$ and $P^*(\mathcal{U}(\mathcal{W})) = P^i(\mathcal{U}(\mathcal{W}))$. Then, it follows that Equation (24) is also equal to

$$\sum_{\{u|\mathcal{W}_\ast(u) = w_\ast\}} P^i(u) = P^i(w_\ast). \tag{25}$$

Lemma A.1. Suppose $P^*(W_\ast = w_\ast)$ is not transportable from a set of available distributions in a causal diagram $\mathcal{G}$. Let $A_{1[t_1]}, A_{2[t_2]} \in \mathcal{W}_\ast$ such that $A_{2[t_2]}$ is a function of $A_{1[t_1]}$. Then $\sum_{a_1} P^*(W_\ast = w_\ast)$ is not identifiable from the same input either.

Proof. Let $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ be the two sets of models witnessing the non-transportability of $P^*(W_\ast = w_\ast)$, they agree on available distributions, but for $w_\ast$, we have $P^{*(1)}(W_\ast = w_\ast) = \alpha, P^{*(2)}(W_\ast = w_\ast) = \beta$ with $\alpha \neq \beta$. Assume, without loss of generality that $\alpha > \beta$. We will extend the strategy used by (Huang & Valtorta, 2008) to construct two sets of models $\mathcal{M}^{(1)'}$ and $\mathcal{M}^{(2)'}$ where the domain of $A_2$ is $\mathbb{X}_{A_2} \times \{0, 1\}$, where $\mathbb{X}_{A_2}$ is the domain of $A_2$ in $\mathcal{M}_1, \mathcal{M}_2$. Let $F(A_1)$ be a probability function from $\mathbb{X}_{A_1}$ to $\{0, 1\}$, such that $P(F(a_1) = k) > 0, k = 0, 1$ and $P(F(a_1) = 0) = 1 - P(F(a_1) = 1)$. In $\mathcal{M}^{(1)'}$, $i = 1, 2$ we define the new function for $A_2$ such that:

$$P^{*(i)'}((a_2, k) | pa_{a_2}, u_{a_2}) = P^{*(i)}(a_2 | pa_{a_2}, u_{a_2})P(F(a_1)=k). \tag{26}$$

For any other $V_j \in \mathcal{V} \setminus \{A_2\}$ we keep the same function such that $P^{*(i)'}(v_j|pa_j, u_j) = P^{*(i)}(v_j|pa_j, u_j)$. In particular, the functions of variables that are children of $A_2$ simply ignore the new bit in $A_2$’s domain. We can verify that for any counterfactual $Z_\ast$ on which $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ agree, we have

$$P^{*(i)'}(Z_\ast = z_\ast) = \sum_{d_\ast \in D_\ast} P^{*(i)'}(D_\ast = d_\ast), \tag{27}$$

where $D_\ast = \text{An}(Z_\ast)$ and $d_\ast$ is the union of $z_\ast$ and the indexing values of the sum. This can be written in the form of a ctf-factor as

$$P^{*(i)'}(Z_\ast = z_\ast) = \sum_{d_\ast \in D_\ast} P^{*(i)'}(\bigwedge_{D_{pa_d} = d} D_{\ast}) \tag{28}$$

and then sum over $U$

$$P^{*(i)'}(Z_\ast = z_\ast) = \sum_{u, d_\ast \in D_\ast} P^{*(i)'}(\bigwedge_{D_{pa_d} = d \mid u} D_{\ast}) P^{*(i)'}(u). \tag{29}$$

Note that $P^{*(i)'}(u) = P^{*(i)}(u)$. Also, once $U$ has been fixed, any $D_{pa_d}$ becomes deterministic and independent of any other variable, as it only depends on $Pa_d$ and $U_d$. Then we can write

$$P^{*(i)'}(\bigwedge_{D_{\ast} \in D_\ast} D_{pa_d} = d \mid u) = \prod_{D_{\ast} \in D_\ast} P^{*(i)'}(D_{pa_d} = d \mid u) = \prod_{D_{\ast} \in D_\ast} P^{*(i)'}(d \mid pa_d, u), \tag{30}$$

---

3see (Correa et al., 2021) for details on computing counterfactual probabilities from the SCM
For any $D \neq A_2$ we have $P^{(i)}(d \mid pa_d, u) = P^{(i)}(d \mid pa_d, u)$. Then $P^{(i)}(Z_\ast = z_\ast) = P^{M_1}(Z_\ast = z_\ast)$, if $D_\ast$ does not contain any $A_2$.

By construction, the factor for $A_2$ is given by Equation (26). Hence if $D_\ast$ does contain some instance of $A_2$ and $d_\ast$ has some $(a_2, k)$,

$$P^{(i)}(Z_\ast = z_\ast) = \sum_{d_\ast \setminus z_\ast} P^{(i)} \left( \bigwedge_{D_k \in D_\ast} D_{pa_d} = d \right) P(F(a_1) = k).$$ (31)

If $(a_2, k) \notin z_\ast$ we can sum out $k$ to obtain $P^{(i)}(Z_\ast = z_\ast) = P^{(i)}(Z_\ast = z_\ast)$. If $a_1 \notin (d_\ast \setminus z_\ast)$ then the expression becomes

$$P^{(i)}(Z_\ast = z_\ast) = P(F(a_1) = k) \sum_{d_\ast \setminus z_\ast} P^{(i)} \left( \bigwedge_{D_k \in D_\ast} D_{pa_d} = d \right) = P(F(a_1) = k) P^{(i)}(Z_\ast = z_\ast).$$ (32)

For any input distribution $P^k(V; \sigma_T)$, we have

$$P^k(V; \sigma_T) = P^k(V \setminus T) \prod_{T \in T} P(T \mid P_{a_1}; \sigma_T) = P^k(\bigwedge_{V_i \in V \setminus T} V_i[t]) \prod_{T \in T} P(T \mid P_{a_1}; \sigma_T),$$ (33)

if $A_1[t]$ appears in the expression and does not have an index in $d_\ast \setminus z_\ast$, or $A_1 \in T$ so that no instance of $A_1$ is an ancestor of any other variable in the expression. Hence any experimental distribution of the form $P(V; \sigma_T)$, given as input $(T \in Z)$, will match in $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ as it matches in $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$.

Consider the assignment $w_\ast \setminus \{a_2, a_2', 0\}$, by construction we have

$$P^{(i)}(W_\ast = (w_\ast \setminus \{a_2\}) \cup \{a_2', 0\}) = P^{(i)}(W_\ast = w_\ast) P(F(a_1) = 0)$$ (34)

For $a_1 \in w_\ast, a_1' \neq a_1$, let $P(F(a_1) = 0) = 1/2$ and $P(F(a_1') = 0) = (\alpha - \beta)/4$. Then

$$\sum_{a_1'} P^{(i)}(W_\ast = (w_\ast \setminus \{a_1, a_2\}) \cup \{a_1', (a_2, 0)\})$$

$$= \sum_{a_1'} P^{(i)}(W_\ast = (w_\ast \setminus \{a_1, a_2\}) \cup \{a_1', a_2\}) P(F(a_1') = 0)$$ (35)

For $\mathcal{M}^{(1)}$ this means

$$\sum_{a_1'} P^{(1)}(W_\ast = (w_\ast \setminus \{a_1, a_2\}) \cup \{a_1', (a_2, 0)\})$$

$$= \frac{1}{2} \alpha + \left( \frac{\alpha - \beta}{4} \right) \sum_{a_1 \neq a_1'} P^{(1)}(W_\ast = (w_\ast \setminus \{a_1, a_2\}) \cup \{a_1', a_2\})$$ (36)

$$> \frac{1}{2} \alpha$$ (37)

As for $\mathcal{M}^{(2)}$:

$$\sum_{a_1'} P^{(2)}(W_\ast = (w_\ast \setminus \{a_1, a_2\}) \cup \{a_1', (a_2, 0)\})$$

$$= \frac{1}{2} \beta + \left( \frac{\alpha - \beta}{4} \right) \sum_{a_1 \neq a_1'} P^{(2)}(W_\ast = (w_\ast \setminus \{a_1, a_2\}) \cup \{a_1', a_2\})$$ (38)

$$< \frac{1}{2} \beta + \left( \frac{\alpha - \beta}{4} \right)$$ (39)

$$< \frac{1}{2} \alpha.$$ (40)

Then, $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ are compatible with $\mathcal{G}$, match in the available distributions and yield different $\sum_{a_1} P^*(W_\ast = w_\ast)$. \qed
Then, \( P^*(W = w) \) is transportable from \( \mathbb{Z} \) only if it is consistent. If consistent, let \( \pi = V(W_\alpha) \) and \( \pi' = Pa(W) \setminus \pi \). Then \( P^*(W = w) \) is equal to \( P_{\pi'}^*(w) \) where \( w \) and \( w' \) are consistent with \( w \cup \{ Pa_w \in W \} \) and \( P^*(W = w) \) is transportable from \( \mathbb{Z} \) if \( P_{\pi'}^*(w) \) is transportable from \( \mathbb{Z} \).

**Proof**. Let \( z = w \cup \{ Pa_w \in W \} \). Since the factor is consistent, \( z \) contains at most one value per observable. We can write \( P^*(W = w) = P^*(\pi'(Pa_w \in W) \pi_w = w) \) where every \( Pa_w \) is consistent with \( z \).

Let \( W \subseteq W \) such that \( Pa_w \cap W = \emptyset \), then \( Pa_w \subseteq W' \) and \( W_{Pa_w} = W_w \) where \( w' \) has any valid value for \( W' \setminus Z \).

For a \( W \) with parents in \( W \), suppose every \( T \in Pa_w \setminus W \) already appears as \( T_w \). Then, we can use composition to make \( W_{Pa_w \cup w} = W_{w'} \). Applying this reasoning from parent to children in \( W \) we get

\[
P^*(w) = P^*\left( \bigwedge_{\{ Pa_w \in W \}} W_{w'} = w \right) = P_{\pi'}^*(w).
\]

The ‘only if’ portion of the statement follows from Lemma A.2 because even if \( P^*(w) \) cannot be identified from any subset of interventional distributions in the target domain \( \pi^* \), it will certainly not be transportable from interventional distributions in other domains, even if they can be transported to \( \pi^* \).

**A.1. Conditional Counterfactual Transportability**

Before the proof of Lemma 4.2, we prove some auxiliary lemmata.

**Lemma A.3**. Let \( A, B \), and \( C \) be binary random variables causally related as given by the chain \( A \rightarrow B \rightarrow C \). Suppose \( P(B = 1 \mid A = 1) = \alpha \) and \( P(B = 1 \mid A = 0) = 1 - \alpha \), for some \( \alpha \in [0, 1] \). Then, for any \( \beta \) such that \( |1/2 - \beta| \leq |1/2 - \alpha| \), there is always a function \( f_C \) such that \( P(C = 1 \mid A = 1) = \beta \), \( P(C = 1 \mid A = 0) = 1 - \beta \) and \( P(A, B, C) \) is a positive distribution.

**Proof**. Let \( P(C = 1 \mid B = 1) = x \) and \( P(C = 1 \mid B = 0) = 1 - x \), then

\[
\beta = P(C = 1 \mid A = 1) = \sum_b P(C = 1 \mid b)P(b \mid A = 1) = 1 - \alpha + x(2\alpha - 1)
\]

where \( U_c \) is a binary unobservable, \( P(U_c = 0) = x \) and \( \oplus \) is the logical xor operator.
and $m, n$ are nonnegative integers. Then we can define functions for all variables involved such that they are binary and

$$P(a, b) = \begin{cases} \frac{1}{2} \gamma & \text{if } a = b \\ \frac{1}{2} (1 - \gamma) & \text{otherwise} \end{cases}$$

(45)

for any $\gamma \in (0, 1)$.

Proof. First, if $C$ is unobservable set $P(C = 1) = 1/2$, else define an unobservable $U_c$ with $P(U_c = 1) = 1/2$ and let $f_C = U_c$. Let $\alpha, \beta \in (0, 1)$ be parameters to decide later.

If $n = 0$ define $f_A$ such that $P(A = 1 \mid C = 1) = \alpha, P(A = 1 \mid C = 0) = 1 - \alpha$. Similarly, if $m = 0$ define $f_B$ such that $P(B = 1 \mid C = 1) = \alpha, P(B = 1 \mid C = 0) = 1 - \alpha$.

Suppose $n > 0$, then we will define the functions for $A_1, A_2, \ldots, A_n, A$ such that $P(A_i = 1 \mid C = 1)$ gets closer to $\alpha$ as $i$ increases. If $\alpha < 1/2$, set $f_{A_i}$ such that $P(A_i = 1 \mid C = 1) = \alpha/(n + 1)$, $P(A_i = 1 \mid C = 0) = 1 - \alpha/(n + 1)$. Then use lemma A.3 to define $f_{A_i}, i = 1, \ldots, n$ and $f_A$ such that $P(A_i = 1 \mid C = 1) = i\alpha/(n + 1), P(A_i = 1 \mid C = 0) = 1 - i\alpha/(n + 1)$ and finally $P(A = 1 \mid C = 1) = (n + 1)\alpha/(n + 1) = \alpha, P(A = 1 \mid C = 1) = 1 - \alpha$.

If $\alpha > 1/2$ use the same strategy but starting from $P(A_1 = 1 \mid C = 1) = 1 - (1 - \alpha)/(n + 1)$ and decreasing as $P(A_i = 1 \mid C = 1) = 1 - i(1 - \alpha)/(n + 1)$, to obtain $P(A = 1 \mid C = 1) = 1 - (n + 1)(1 - \alpha)/(n + 1) = \alpha$.

The same procedure is applied for $B_1, \ldots, B_m, B$ to obtain $P(B = 1 \mid C = 1) = \beta, P(B = 1 \mid C = 0) = 1 - \beta$.

Finally,

$$P(A = 1, B = 1) = \sum_c P(A = 1 \mid c)P(B = 1 \mid c)P(c)$$

(46)

$$= \frac{1}{2} [\alpha \beta + (1 - \alpha)(1 - \beta)]$$

(47)

$$P(A = 0, B = 0) = \frac{1}{2} [(1 - \alpha)(1 - \beta) + \alpha \beta]$$

(48)

$$P(A = 0, B = 1) = \frac{1}{2} [(1 - \alpha)\beta + \alpha(1 - \beta)]$$

(49)

$$P(A = 1, B = 0) = \frac{1}{2} [\alpha (1 - \beta) + (1 - \alpha)\beta].$$

(50)

If $\gamma < 1/2$ make $\beta = 1 - \alpha$ and $\alpha = \frac{1 + \sqrt{1 - 4\gamma}}{2}$. If $\gamma \geq 1/2$, let $\beta = \alpha$ and $\alpha = \frac{1 + \sqrt{2\gamma - 1}}{2}$. It is just a matter of algebra to verify that $P(A, B)$ results in the intended distribution. □

Lemma A.5. Let $W_s$ be a set of counterfactual variables, $X_s$, and let $A_{1s}, A_{2s}, \ldots, A_{Js}$ be the ancestral components of $W_s$ given $X_s$. Then

$$P(W_s = w_s) = \prod_j \sum_{a_{js} \in A_{js}} P\left(\bigwedge_{A_t \in A_{js}} A_{pa_t} = a\right),$$

(51)

where $pa_t$ is consistent with $t$ and $a_{js}$, for each $A_t \in A_{js}$.

Proof. For sake of simplicity let $X_s = ||X_s||$. First, note that by definition $A_{1s}, A_{2s}, \ldots$ must be mutually disjoint and their union is equal to $A_s = A \cap (W_s)$. Then, we can write

$$P(W_s = w_s) = \sum_{a_{s} \setminus w_s} P\left(\bigwedge_j A_{js} = a_{js}\right).$$

(52)

Furthermore, no variable in two distinct ancestral components could be in the same c-component in $G[V(A_s)]$. Hence, we have

$$P(W_s = w_s) = \sum_{a_{s} \setminus w_s} \prod_j P\left(\bigwedge_{A_t \in A_{js}} A_{pa_t} = a\right).$$

(53)
To prove the result, it suffices to show that this expression can be factorized such that the sum over each \( a_{j*} \setminus w_s \) only affects factor \( j \). For this, we argue that for \( A_t \in A_{j*} \), no value of \( pa_a \) or \( a \) appears in \( a_{k*} \setminus w_s \) for any \( k \neq j \). To see why this is the case, assume for the sake of contradiction that (i) there is some value in \( pa_a \) that appears in such a \( a_{k*} \setminus w_s \) or (ii) \( a \) appears in the same set. For case (i), this implies that there is an ancestor of \( A_t \) (not in \( X_s \)) which is also an ancestor of some variable in component \( A_{k*} \). For case (ii), this means that \( A_t \notin X_s \) is an ancestor of some variable in component \( A_{k*} \). In both cases would imply that \( A_{j*} \) and \( A_{k*} \) are not disjoint, a contradiction. 

\[ \square \]

**Lemma 4.2** (Conditional-marginal transportability reduction). Let \( Y_s, X_s \) be two sets of counterfactual variables and let \( D_s \) be the set of variables in the same ancestral component, given \( X_s \), as any variable in \( Y_s \), then

\[
P^*(y_s \mid x_s) = \frac{\sum_d \sum_{x_s \in (Y_s \cup X_s)} P^*(\bigwedge_{D \in D_s} D_{pa_d} = d) P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a) \prod_{j \in A_{j*} \setminus (Y_s \cup X_s)} \sum_{y_s \in (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a)}{\sum_d \sum_{x_s \in (Y_s \cup X_s)} P^*(\bigwedge_{D \in D_s} D_{pa_d} = d) P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a) \prod_{j \in A_{j*} \setminus (Y_s \cup X_s)} \sum_{y_s \in (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a)}
\]

\( (20) \)

where \( pa_d \) is consistent with \( t, d_s, \) and \( x_s \) for each \( D_t \in D_s \). Moreover, \( P^*(y_s \mid x_s) \) is transportable from \( \mathcal{Z} \) iff \( P^*(\bigwedge_{D \in D_s} D_{pa_d} = d) \) is transportable from \( \mathcal{Z} \).

**Proof.**

\[
P^*(Y_s = y_s \mid X_s = x_s) = \frac{P^*(Y_s = y_s, X_s = x_s) \sum_{y_s \in (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a)}{\sum_{y_s \in (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a)}
\]

\( (54) \)

Let \( A_{1*}, A_{2*}, \ldots \) be the ancestral components of \( Y_s \cup X_s \), given \( X_s \). Then, by Lemma A.5

\[
P^*(Y_s = y_s, X_s = x_s) = \prod_{j \in A_{j*} \setminus (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a) \prod_{j \in A_{j*} \setminus (Y_s \cup X_s)} \sum_{y_s \in (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a)
\]

\( (55) \)

Let \( Y_r \in Y_s \) and \( y \) its corresponding value in \( y_s \). If there exists any variable \( A_t \in A_{j*} \) such that \( y \in t, Y_r \) and \( A_t \) must be in the same \( A_{j*} \), as the former is then an ancestor of the latter. Let \( A_{Y_r} \) be the union of ancestral components that contain ancestors of \( Y_r \), and let \( A_{Y_r} \) be the rest. Then, if we sum over \( Y_r \) in the right-hand-side of Equation (55), the factors with variables in \( A_{Y_r} \) can take out of the sum.

\[
P^*(Y_s = y_s \mid X_s = x_s) = \frac{\left( \sum_{A_{Y_r} \setminus (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a) \right) \left( \sum_{A_{Y_r} \setminus (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a) \right) \prod_{j \in A_{j*} \setminus (Y_s \cup X_s)} \sum_{y_s \in (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a)}{\left( \sum_{A_{Y_r} \setminus (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a) \right) \prod_{j \in A_{j*} \setminus (Y_s \cup X_s)} \sum_{y_s \in (Y_s \cup X_s)} P^*(\bigwedge_{A \in A_{j*}} A_{pa_a} = a)}
\]

\( (56) \)

After cancelling out the first factor, simplifying the expression, and defining \( D_s = A_{Y_r} \), we obtain Equation (20).

For simplicity, let \( G^\Delta' \) be the same as \( G^\Delta \) after removing all edges out of \( V(X_s \cap D_s) \) and any edge out of \( V(Y_s \cap A^n(X_s)) \). \( G' \) and \( G \) have the same c-component structure and the same ancestral relationships; therefore and by the same reasoning, the query is not transportable from \( \langle G', \mathcal{Z} \rangle \) either. For the rest of the proof, let \( G^\Delta \) denote \( G^\Delta' \) instead. Clearly, if \( P^*(\bigwedge_{D \in D_s} D_{pa_d} = d) \) is transportable, so it is \( P^*(Y_s = y_s \mid X_s = x_s) \). For the second part of the statement, suppose this query is not transportable and let \( M(1), M(2) \) be two sets of models witnessing this fact.

For simplicity, let

\[
\rho(y, x) = \sum_{d_s \setminus (y_s \cup x_s)} P^*(\bigwedge_{D \in D_s} D_{pa_d} = d)
\]

\( (57) \)

then

\[
P^*(Y_s = y_s \mid X_s = x_s) = \frac{\rho(y_s, x_s)}{\sum_{y_s} \rho(y_s, x_s)}
\]

\( (58) \)

If \( \rho(x) = \sum_y \rho(y, x) \) is transportable then \( P^*(Y_s = y_s \mid X_s = x_s) \) must be non-transportable, else

\[
\rho(y, x) = P^*(Y_s = y_s \mid X_s = x_s) \rho(x)
\]

\( (59) \)
Contradicting the assumption that the $\rho(y, x)$ is not identifiable. Therefore, we can further assume $\rho(x)$ is not identifiable, for the rest of the argument.

Let $(y', x')$ be an assignment such that $\rho(y', x')$ differs between $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$, then let

<table>
<thead>
<tr>
<th>$X_*$</th>
<th>$Y_*$</th>
<th>$\rho^{(1)}(X, Y)$</th>
<th>$\rho^{(2)}(X, Y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x'$</td>
<td>$y'$</td>
<td>$a$</td>
<td>$b$</td>
</tr>
<tr>
<td>$x' \neq y'$</td>
<td>$y'$</td>
<td>$c$</td>
<td>$d$</td>
</tr>
<tr>
<td>$\neq x'$</td>
<td>$y'$</td>
<td>$e$</td>
<td>$f$</td>
</tr>
<tr>
<td>$\neq x' \neq y'$</td>
<td></td>
<td>$g$</td>
<td>$h$</td>
</tr>
</tbody>
</table>

Due to the non-transportability of $\rho(y, x)$ we have $a \neq b$ and without loss of generality we can assume $a > b$. Similarly, due to the non-transportability of $\rho(x)$ we have $a + c \neq b + d$. For $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$:

$$\rho^{(1)}(y' \mid x') = \frac{a}{a + c}$$

$$\rho^{(2)}(y' \mid x') = \frac{b}{b + d}.$$  

These probabilities are equal if and only if $ad = bc$. Hence, if they are not equal, we are done as $\rho^{(1)}(y' \mid x') \neq \rho^{(2)}(y' \mid x')$. If they are equal, let $X_t \in X_* \cap D_*; t$ be such that $X_t$ and some $Y_t \in Y_*$ have a common ancestor or they have ancestors sharing a bidirected edge. Assume $X_t$ is such that no other element in $X_*$ is closer to $Y_t$ in the corresponding path in $G$. Let $\overline{p}$ be the path in $G$ that goes from $X$ to $Y$, passing to the common ancestor or bidirected alluded before. Add a bit to every variable in $\overline{p}$ and denote them with subscript $p$. Define independent functions for the bits which we will parametrize later.

We define two new sets of models, $\mathcal{M}^{(1)'}$ and $\mathcal{M}^{(2)'}$, based on $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. For every variable in $\overline{p}$ except for $X$ and $Y$, append the corresponding extra bit, defined in $\overline{p}$, to the original variables in $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. Rename $X$ and $Y$ as $\tilde{X}, \tilde{Y}$ and make them unobservable, then define $X$ in the new models with functions:

$$f_x' = \begin{cases} x' & \text{if } X_p = 1 \\ \tilde{X} & \text{otherwise} \end{cases}$$

where $X_p$ is unobservable too and $x' = x'(X)$, for which the query disagrees in $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. Analogously, define

$$f_y' = \begin{cases} y' & \text{if } Y_p = 1 \\ \tilde{Y} & \text{otherwise} \end{cases}$$

According to our choice of $X$, there exists $Z \in \overline{p} \cap \text{An}(X) \cap \text{An}(Y)$ (possibly $Y$ itself) or there exists $Z_1, Z_2 \in \overline{p}$ with $Z_1 \in \text{An}(X), Z_2 \in \text{An}(Y)$ and there is an edge $Z_1 \leftrightarrow Z_2 \in \overline{p}$. For the parametrization of the extra bits in $\overline{p}$ define a new unobservable $U$ and let $Z_p = U$ if the common ancestor is observable or let $U$ be the unobservable parent of $Z_{1p}$ and $Z_{2p}$ in the second. Notice that $Z_1$ and $Z_2$ may be equal to $X$ and $Y$ themselves.

Using lemma A.4 we will parametrize $\overline{p}$ such that

<table>
<thead>
<tr>
<th>$X_0$</th>
<th>$Y_0$</th>
<th>$P(X_0, Y_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$\frac{1}{2} \gamma$</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$\frac{1}{2} (1 - \gamma)$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>$\frac{1}{2} (1 - \gamma)$</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>$\frac{1}{2} \gamma$</td>
</tr>
</tbody>
</table>

for some $\gamma \in (0, 1)$ that we will pick later.
Claim A.1 (Disagreement on the query). $\mathcal{M}^{(1)'}$ and $\mathcal{M}^{(2)'}$ disagree on the query for any $\gamma$ such that $c - d \neq [(a + c + 1)h - (b + d + 1)g](1 - \gamma)$. ⊢

Proof. For $\mathcal{M}^{(1)'}$ and $\mathcal{M}^{(2)'}$ we have

$$\rho' (x', y') = \sum_{x_p, y_p, \bar{x}, \bar{y}} \rho'(x', y', \bar{x}, \bar{y} | x_p, y_p) P(x_p, y_p).$$

(64)

Going over each possible combination of $X_p$ and $Y_p$, we get

$$\rho' (x', y') = \rho(X = x', Y = y') P(X_p = 0, Y_p = 0)$$

(65)

$$+ \rho(Y = y') P(X_p = 1, Y_p = 0)$$

(66)

$$+ \rho(X = x') P(X_p = 0, Y_p = 1)$$

(67)

$$+ P(X_p = 1, Y_p = 1).$$

(68)

Similarly, we have

$$\rho' (x') = \rho(X = x') P(X_p = 0) + P(X_p = 1).$$

(69)

For $\mathcal{M}^{(1)'}$

$$\rho^{(1)'}(y' | x') = \frac{\frac{1}{2}a\gamma + \frac{1}{2}(a + e)(1 - \gamma) + \frac{1}{2}(a + c)(1 - \gamma) + \frac{1}{2}\gamma}{\frac{1}{2}(a + c) + \frac{1}{2}}$$

(70)

$$= \frac{a\gamma + (2a + c + e)(1 - \gamma) + \gamma}{a + c + 1}$$

(71)

$$= \frac{a - (a + c + e)(\gamma - 1) + \gamma}{a + c + 1}$$

(72)

$$= \frac{a + (a + c + e)(1 - \gamma) - (1 - \gamma) + 1}{a + c + 1}$$

(73)

$$= \frac{a - (1 - (a + c + e))(1 - \gamma) + 1}{a + c + 1}$$

(74)

$$= \frac{a - g(1 - \gamma) + 1}{a + c + 1}$$

(75)

Analogously for $\mathcal{M}^{(2)'}$:

$$\rho^{(2)'}(y' | x') = \frac{b - h(1 - \gamma) + 1}{b + d + 1}$$

(76)

Those two are equal if and only if

$$ab - bg(1 - \gamma) + b + ad - dg(1 - \gamma) + d + a - g(1 - \gamma) + 1$$

$$= ab - ah(1 - \gamma) + a + bc - ch(1 - \gamma) + c + b - h(1 - \gamma) + 1$$

$$\iff$$

$$-bg(1 - \gamma) + ad - dg(1 - \gamma) + d + -g(1 - \gamma)$$

$$= -ah(1 - \gamma) + bc - ch(1 - \gamma) + c + -h(1 - \gamma)$$

(77)

(78)

Recall that we have $ad = bc$, which also implies that $c \neq d$, else $a = b$ which is a contradiction. Then, the condition for equality can be further simplified as

$$c - d = [(a + c + 1)h - (b + d + 1)g](1 - \gamma).$$

(79)

The left hand side is nonzero, all $a, b, c, d, g$ and $h$ are fixed, and $\gamma$ is a free parameter. Therefore, as long as we pick a $\gamma$ such that the equality doesn’t hold, we get that $\rho' (y' | x') = P'(Y_* = y' | X_* = x')$ does not match in $\mathcal{M}^{(1)'}$ and $\mathcal{M}^{(2)'}$. □
**Claim A.2** (Agreement on given distributions). Let $Z \subset V$ be any subset of observable variables and let $\sigma_Z \in Z^k \in \mathbb{Z}$. If $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$ agree on $P^k(V; \sigma_Z)$, then $\mathcal{M}^{(1)'}$ and $\mathcal{M}^{(2)'}$ also agree on $P^k(V; \sigma_Z)$.

**Proof.** For simplicity we omit the superscript $k$ for the domain, which is fixed with $\sigma_Z$. The superscript (1) and (2) indicate to which of the sets of models under consideration the expression refers to.

As the input distributions are in layer 2, let us use c-factor notation, where each c-factor corresponds to a causal effect. For any $C \subseteq V$, the quantity $Q[C](v)$ is the c-factor of $C$ and denotes the following function

$$Q[C](v) = \sum_{u(C)} \prod_{v_i \in C} P^k(v_i \mid pa_z, u_i) P^k(v(C)),$$

(80)

where $U(C) = \bigcup_{V \in C} U_i$.

Let $C_1, C_2, \ldots$ be the C-components of $\mathcal{G}_{\sigma_Z}$. By assumption we have $Q^{(1)}[V; \sigma_Z] = Q^{(2)}[V; \sigma_Z]$, and since any $Q[C_j; \sigma_Z]$ is identifiable from $Q[V; \sigma_Z]$ (Tian & Pearl, 2002b), it follows $Q^{(1)}[C_j; \sigma_Z] = Q^{(2)}[C_j; \sigma_Z]$ for any $C_j$ not intersecting $Z$.

$\mathcal{M}^{(k)'}$ is identical to $\mathcal{M}^{(k)}$, $k = 1, 2$, except for the functions of the observables in the path $\overline{p}$. For any variable $T$ not in $\overline{p}$, but with a parent on it, the function $f_T$ remains the same and it simply ignores the extra bit that its parent has in $\mathcal{M}^{(k)'}$.

Let $C_j$ be a c-component containing some set of variables $R$ in $\overline{p}$ different to $X$ and $Y$ (the endpoints of $\overline{p}$). First, by definition

$$Q^{(k)'}[C_j; \sigma_Z](v) = \sum_{u(C_j)} \prod_{v_i \in C_j} P^{(k)'}(v_i \mid pa_z, u_i; \sigma_X) P^{(k)'}(v(C_j)).$$

(81)

For any $S \notin \overline{p}, R \in R$, $X$ and $Y$ that could be in $C_j$ in $\mathcal{M}^{(k)'}$, their corresponding factors in the previous expression can be re-written in terms of probabilities of $\mathcal{M}^{(k)}$, as follows:

$$P^{(k)'}(s \mid pa_z, u_s; \sigma_X) = P^{(k)}(s \mid pa_z, u_s; \sigma_X),$$

(82)

$$P^{(k)'}(r \mid pa_r, u_r; \sigma_X) = P^{(k)}(r \mid pa_r, u_r; \sigma_X) P(r_p \mid (pa_r)_p),$$

(83)

$$P^{(k)'}(y \mid pa_y, u_y; \sigma_X) = P^{(k)}(y \mid pa_y, u_y; \sigma_X) P(Y_p = 0 \mid (pa_y)_p) + 1(y = y') P(Y_p = 1 \mid (pa_y)_p),$$

(84)

$$P^{(k)'}(w \mid pa_w, u_w; \sigma_X) = P^{(k)}(w \mid pa_w, u_w; \sigma_X) P(W_p = 0 \mid (pa_w)_p) + 1[w = w'] P(W_p = 1 \mid (pa_w)_p).$$

(85)

It follows that

$$Q^{(k)'}[C_j; \sigma_Z](v) = \left[ \prod_{R \in R} P(r_p \mid (pa_r)_p) \right]$$

$$Q^{(k)}[C_j; \sigma_Z](v) P(Y_p = 0 \mid (pa_y)_p) P(W_p = 0 \mid (pa_w)_p)$$

$$+ Q^{(k)}[C_j \setminus \{Y\}; \sigma_Z](v) P(Y_p = 1 \mid (pa_y)_p) P(W_p = 0 \mid (pa_w)_p) 1[y = y']$$

$$+ Q^{(k)}[C_j \setminus \{W\}; \sigma_Z](v) P(Y_p = 0 \mid (pa_y)_p) P(W_p = 1 \mid (pa_w)_p) 1[w = w']$$

$$+ Q^{(k)}[C_j \setminus \{W, Y\}; \sigma_Z](v) P(Y_p = 1 \mid (pa_y)_p) P(W_p = 1 \mid (pa_w)_p) 1[w = w', y = y']$$

(86)

Since $X$ and $Y$ have no descendants in $\mathcal{G}$,

$$Q^{(k)}[C_j \setminus \{W, Y\}; \sigma_Z](v) = \sum_{w, y} Q^{(k)}[C_j \setminus \{W, Y\}; \sigma_Z](v)$$

(87)

$$Q^{(k)}[C_j \setminus \{W\}; \sigma_Z](v) = \sum_{w} Q^{(k)}[C_j \setminus \{W\}; \sigma_Z](v)$$

(88)

$$Q^{(k)}[C_j \setminus \{Y\}; \sigma_Z](v) = \sum_{y} Q^{(k)}[C_j \setminus \{W\}; \sigma_Z](v).$$

(89)
all match between $\mathcal{M}^{(1)}$ and $\mathcal{M}^{(2)}$. Consequently, every c-factor in the right-hand side of (86) is the same in those models, and since every other term is also the same in both $\mathcal{M}^{(1)'}$ and $\mathcal{M}^{(2)'}$, we conclude that $Q^{(1)'}[C_j; \sigma_Z](v) = Q^{(2)'}[C_j; \sigma_Z](v)$, which in turn implies our claim since

$$P^{(k)'}(v; \sigma_Z) = \prod_j Q^{(k)'}[C_j; \sigma_Z](v).$$

(90)

In summary, $\mathcal{M}^{(1)'}$ and $\mathcal{M}^{(2)'}$ induce $\mathcal{G}^\Delta$ and matching $Z$, yet they differ on the value for $P'(Y_* = y' \mid X_* = x')$, proving the non-identifiability of the query.

**A.2. Nested Counterfactual Transportability**

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

$$P^*(Y_{\hat{Z}, \hat{X}} = y) = \sum_{x \in X} P^*(Y_{\hat{Z}, x} = y, \hat{X} = x).$$

(22)

Proof. (soundness) From the model we have: $P^*(Y_{\hat{Z}, \hat{X}} = y) = \sum_{\{u\}Y_{\hat{Z}, \hat{X}}(u) = y} P^*(u)$. In $M_{\hat{Z}, \hat{X}}$, for any given $u$, the variables in $X$ take a particular value-vector $x = \hat{X}(u)$. Then, we can partition $\{u \mid Y_{\hat{Z}, \hat{X}}(u) = y\}$ as $\bigcup_{x \in X} \{u \mid Y_{\hat{Z}, x} = y\}$. Hence

$$P^*(Y_{\hat{Z}, \hat{X}} = y) = \sum_{x \in X} \sum_{\{u\}Y_{\hat{Z}, x}(u) = y, \hat{X}(u) = x} P^*(u) = \sum_{x \in X} P^*(Y_{\hat{Z}, x} = y, \hat{X} = x).$$

(91)

(necessity) Suppose $P^*(Y_{\hat{Z}, x} = y, \hat{X} = x)$ is not transportable from $Z$. Since $X \subseteq An(Y)_{\mathcal{G}^\Delta}$ there exist $X_r \in \hat{X}$ and a set of variables $D_x$ such that a path $X \rightarrow D_1 \rightarrow D_2 \rightarrow D_k \rightarrow Y$ exists in $\mathcal{G}^*$ for some $Y_Z \in \hat{Y}$, with $k \geq 0$. This implies that $D_{i}[Z]$ is a function of $X_r, D_{i+1}[Z]$ is a function of $D_i[Z]$ and $Y_Z$ is a function of $D_k[Z]$. Then, we can use Lemma A.1 from $X$ to $Y$ following this path to prove the result.

**B. Algorithm**

**B.1. Soundness and Completeness**

**Lemma B.1** (ctfTRU completeness for non-conditional, non-nested queries). A counterfactual probability $P^*(y_*)$ is transportable from $Z$ and $\mathcal{G}^\Delta$ if and only if ctfTRU returns an expression for it.

Proof. (soundness) The algorithm starts by simplifying the target event by removing irrelevant subscripts and looking for inconsistent or redundant events (line 1). Line 2 finds the relevant ctf-factors consisting of a single c-component, licensed by the properties of ancestral sets and ctf-factors. As long as the factors are consistent, and allowed by Theorem 4.1, lines 7-10 try to transport of the causal effect $P^*_{P_{\alpha}(C_i)\mid C_i}(C_i)$ from the available distributions employing the algorithm $\sigma$-TR (Algorithm 4) as a subroutine. Then, the factor $P^*(C_i = c_i)$ is transported by evaluating $P^*_{P_{\alpha}(C_i)\mid C_i}(C_i)$ with the appropriate values according to Theorem 4.1. If all the ctf-factors are consistent and transportable, the algorithm returns the query as in line 14. This expression is equal to the original query.

(necessity) The procedure fails if any of the factors $P^*(C_i = c_i)$ is inconsistent or not transportable from $Z$. If inconsistent, Theorem 4.1 ascertains the non-transportability of the factor from $Z$. On the other hand, $\sigma$-TR (in line 5) failing to obtain $P^*_{P_{\alpha}(C_i)\mid C_i}(C_i)$ from $Z$ implies the non-transportability of $P^*_{P_{\alpha}(C_i)\mid C_i}(C_i)$ from $Z$ (Correa & Bareinboim, 2020). That is, there exist two sets of models that agree on all distributions given by $Z$ but differ on $P^*_{P_{\alpha}(C_i)\mid C_i}(C_i)$ for a particular set
Algorithm 4 $\sigma$-TR($C_i, Z, G^\Delta$)

**Input:** A set of variables $C_i$ that form a single c-component in $G$, a set of distributions $Z$, and a selection diagram $G^\Delta$.

**Output:** An expression for the c-factor $P^*_u(C_i)\setminus C_i(C_i)$ in terms of distributions available in $Z$ or FAIL if this distribution is not transportable.

1: for each $Z \in Z^k \in Z$ s.t. $C_i \cap Z = \emptyset$ and $C_i \cap \Delta^k = \emptyset$ do
2: let $B_i$ be the c-component of $G_{\sigma Z_i}$ such that $C_i \subseteq B_i$.
3: compute $P_{V\setminus B_i}(B_i)$ from $P_Z(V)$.
4: let $Q$ $\leftarrow$ IDENTIFY($C_i, B_i, P_{V\setminus B_i}(B_i), G^k_{\sigma Z_i}$).
5: if $Q$ is not FAIL then
6: return $Q$.
7: end if
8: end for
9: return FAIL

of values $v$. With a proper encoding of the values, consistent with $(c_i \cup \bigcup_{C_i \in C_i} pa_i)$, this is also a counterexample for $P^*_u(C_i = c_{i\ast})$.

If any $P^*(C_i = c_{i\ast})$ is not transportable, then $P^*(W_\ast = w_\ast)$ is also not transportable. This is because if the latter were transportable, the former can be obtained from it; a contradiction. Moreover, Lemma 4.1 then implies that $P^*(Y_\ast = y_\ast)$ is not identifiable.

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

**Proof.** (Soundness and completeness) Lines 1 and 2 determine the set of variables and values $D_\ast$ and $d_\ast$ as specified by Lemma 4.2 to reduce the conditional query to a marginal one. Then, line 3 invokes CTFTRU to identify this quantity. If CTFTRU succeeds, the returned expression is equal to the conditional query. If it fails, Lemma 4.2 implies that the conditional query is also not transportable.

(Efficiency) Let $n = |V|$ and $z = \sum_{Z \in Z} |Z^i|$. Operations such as computing the set of ancestors, ancestral components or c-components in a graph can be done in $O(n^2)$ time. The number of sets of ancestors and c-components is $O(n)$, hence the first two operations in CTFTR take at most $O(n^3)$ time. CTRTRU is called once and the first step, that uses SIMPLIFY, takes at most $O(n^2)$ time because the simplification is done by computing sets of ancestors and then performing checks that take $O(n^2)$ time. There are $O(n)$ ctf-factors to process, so the total number of times the for-loop in CTFTRU could execute, calling $\sigma$-TR is $O(n)$. Besides some $O(n^2)$ operations, $\sigma$-TR may call IDENTIFY up to $z$ times. IDENTIFY recursively reduces the input c-factor at least by one variables each time, and the operations used can be performed in $O(n^2)$ so it takes $O(n^2 z)$ per call to return an expression or FAIL. Consequently, CTRTRU takes $O(n^4 z)$ time, which is also the overall time for CTFTR.

**B.2. Transportability algorithm for interventional distributions**

Our algorithm CTFTRU relies on a subroutine to transport a given interventional distribution from $Z$. For this purpose we employ $\sigma$-TR (Correa & Bareinboim, 2020) which is shown in Algorithm 4. In turn, $\sigma$-TR uses the IDENTIFY algorithm (Tian & Pearl, 2002a), shown in Algorithm 5.
Algorithm 5 \textsc{Identify}(C, T, Q, G)

\textbf{Input:} C ⊆ T ⊆ V, Q = Q[T] and graph G. Assuming G[C] and G[T] are composed of a single c-component.

\textbf{Output:} Expression for Q[C] in terms of Q or Fail.

1: Let A ← An(C)\_G[T].
2: if A = C then return $Q[C] = \sum_{t \in A} Q$.
3: if A = T then return FAIL.
4: if A = C then
5: Let $T'$ be the C-component containing C in G[A].
6: Compute $Q[T']$ from $Q[A] = \sum_{t \in A} Q$.
7: return \textsc{Identify}(C, T', Q[T'], G).
8: end if