Generalized Transportability: Synthesis of Experiments from Heterogeneous Domains

Sanghack Lee and **Juan D. Correa** and **Elias Bareinboim**
Causal Artificial Intelligence Laboratory
Department of Computer Science
Columbia University, USA
{sl4712@,j.d.correa@,eb@cs.}columbia.edu

Abstract

The process of transporting and synthesizing experimental findings from heterogeneous data-collections is central in the empirical sciences. In the causal inference literature, this appears under the rubrics of causal effect identifiability (Pearl 1995) and transportability (Pearl and Bareinboim 2011). In this paper, we generalize these settings and investigate the problem of learning conditional causal effects from an arbitrary combination of observational and experimental distributions collected under different conditions and from heterogeneous domains. Specifically, we introduce a unified graphical criterion that completely characterizes the conditions under which conditional causal effects can be uniquely determined from disparate data collections. Further, we develop an efficient, sound, and complete algorithm that outputs an expression for the conditional effect whenever it exists, which synthesizes available causal knowledge; if the algorithm aborts deriving a formula, then such synthesis is provably impossible, unless further parametric assumptions are made. Finally, we prove that Pearl's do-calculus is complete for this task.

1 Introduction

The ability to translate experimental results of a study conducted in one setting to another is a fundamental process within the scientific method. Science would come to a standstill were it not for the ability to extrapolate results from laboratory experiments to outside the laboratory, where the purported causal claims should ultimately hold, i.e., the real world. In biology, for example, we conduct experiments on *Bonobos* in order to learn more about *Homo Sapiens*, even though the latter is only related, but certainly not the same as the former. The capability of generalizing causal knowledge plays a critical role in machine learning as well, since an intelligent system is trained in one environment — where it is allowed to perform causal interventions — with the goal of operating efficiently, and surgically, in a deployment site, which is almost invariably different (Bareinboim and Pearl 2016; Pearl and Mackenzie 2018).

One natural question that arises in these challenging scenarios is what would allow scientists to believe that experimental studies performed in one species could, at least in principle, be used to make causal claims about another different species? Also, how could engineers expect, or perhaps hope, that an intelligent system trained in one environment would operate successfully when deployed in a possibly different ground? The key observation here is that, while there might exist glaring disparities across domains, some causal mechanisms are shared, and owed to their invariances, they would act as anchors allowing knowledge to be transported, and causal learning to eventually take place (Pearl 2000; Spirtes, Glymour, and Scheines 2001).

The fields of machine learning and artificial intelligence provide the theoretical underpinnings to reason with causal mechanisms so as to tackle the challenge of synthesizing experimental findings in a principled way. In particular, we build on the framework of structural causal models (SCMs) (Pearl 2000) to formalize this setting and systematically leverage the invariant features of the underlying data-generating model. An increasingly large class of problems regarding the synthesis of experimental findings across domains has been studied in the last decades within the SCM framework. For instance, the problem of *identifiability* of causal effects has been investigated, which is concerned with the conditions under which the causal effect of a treatment variable (or set) X on an outcome variable (or set) Y, usually written as $P(Y|do(X))$, can be determined from the combination of the observational distribution and qualitative understanding about the domain encoded in the form of a causal diagram. A criterion known as the *back-door* has been proposed (Pearl 1993), which provides a formal, graphical justification for when causal effects can be identified by the adjustment formula (and then propensity score-IPW estimators). There exist a number of other criteria developed to solve this problem (Galles and Pearl 1995; Pearl and Robins 1995; Kuroki and Miyakawa 1999; Halpern 2000; Spirtes, Glymour, and Scheines 2001). Pearl introduced *do*-calculus as a general algebraic solution to this problem, which is applicable for when observational and/or experimental distributions are available (Pearl 1995). Based on this machinery, more general graphical and algorithmic identifiability conditions were derived, which culminated in complete characterizations (Tian 2002; Tian and Pearl 2002; Shpitser and Pearl 2006b; Huang and Valtorta 2006; Shpitser and Pearl 2006a; Bareinboim and Pearl 2012a; Pearl 2015; Lee, Correa, and Bareinboim 2019).

More recently, the problem of generalizing causal distributions across heterogeneous domains[1] has been formalized within the SCM framework, which appeared under the rubric of *transportability* (Pearl and Bareinboim 2011). Transportability has initially considered whether experiments coming from a *source* domain can be leveraged to answer a query in a *target* domain, where the two domains differ in some of their mechanisms (Bareinboim and Pearl 2012b). The transportability setting has then been generalized to allow multiple source domains, different set of manipulable variables per domain, or both (Bareinboim and Pearl 2014). Transportability has been used in a number of more applied settings, e.g., (Westreich and Edwards 2015; Westreich et al. 2017; Lesko et al. 2017; Keiding and Louis 2018; Zhou et al. 2018); see also (Pearl and Mackenzie 2018; Pearl and Bareinboim 2019).

Despite the many advances achieved in this literature in the last decade, each work addressed one of the following aspects:[2] 1. (conditional) a causal query can be of a *conditional* interventional probability instead of only marginal; 2. (specification) available data can be of an *arbitrary collection* of observational and experimental distributions instead of a restricted class (e.g., all combinations of experiments); and 3. (heterogeneity) the data can come from a number of *heterogeneous* domains.

The goal of this paper is to account for these three aspects simultaneously and ultimately provide a solution to the most general version of transportability. Cohesively combining the disparate machinery (e.g., concepts, conditions, algorithms) developed for these different instances of the transportability problem turns out to be a challenging task since they capture different aspects of the problem and operate at distinct levels of abstraction; the main goal of this paper, technically speaking, will be to put these results together under a general, unifying umbrella. Specifically, our contributions are as follows: (1) We derive a unified, sufficient, and necessary graphical criterion for determining whether conditional interventional distributions (including unconditional and observational distributions) in a target domain can be uniquely determined from a set of observational and experimental distributions spread throughout heterogeneous domains; (2) We develop a sound and complete algorithm for this problem. We then prove that the *do*-calculus is complete for the task of general transportability.

1.1 Preliminaries

We use uppercase letters for variables and lowercase for the corresponding values. We denote by $\mathfrak{X}_V$ the state space of V where $v \in \mathfrak{X}_V$. A bold letter represents a set. Calligraphic letters are for mathematical structures such as graphs and models. We use familial notation for relationships among vertices in a graph: $Pa(\cdot)$, $An(\cdot)$, and $De(\cdot)$ represent parents, ancestors, and descendants of variables (including its argument as well). In this paper, we are interested in graphs, induced from a SCM (to be defined formally), with both directed and bidirected edges. The *root set* of a graph is a set of vertices with no outgoing edge. Given a graph $\mathcal{G}$, we use $\mathbf{V}$ to represent the set of vertices in $\mathcal{G}$ in the current scope if no ambiguity arises. Otherwise, we denote by $\mathbf{V}(\mathcal{G}')$ the set of observed variables in $\mathcal{G}'$. We denote by $\mathcal{G}[\mathbf{W}]$ a subgraph induced on $\mathcal{G}$ by $\mathbf{W}$, which consists of $\mathbf{W}$ and edges among them. We define $\mathcal{G} \setminus \mathbf{Z}$ as $\mathcal{G}[\mathbf{V} \setminus \mathbf{Z}]$. We denote by $\mathcal{G}_{\overline{\mathbf{X}}}$ and $\mathcal{G}_{\underline{\mathbf{X}}}$ edge-subgraphs of $\mathcal{G}$ with incoming edges onto $\mathbf{X}$ and outgoing edges from $\mathbf{X}$, respectively, removed. We adopt set-related symbols for graphs, e.g., $\mathcal{G}' \subseteq \mathcal{G}$ denotes $\mathcal{G}'$ being a subgraph of $\mathcal{G}$, or $\mathcal{T} \cup \mathcal{H}$ stands for the union of two graphs $\mathcal{T}$ and $\mathcal{H}$.

As mentioned, we use the language of SCMs (Pearl 2000, Ch. 7) as our basic semantical framework, which allows us to represent observational and interventional distributions as well as different domains. Formally, a tuple $\langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$ defines a SCM $\mathcal{M}$ where i) $\mathbf{U}$ is a set of unobserved variables; ii) $\mathbf{V}$ is a set of observed variables; iii) $\mathbf{F}$ is a set of deterministic functions $\{f_V\}_{V \in \mathbf{V}}$ for observed variables, e.g., $v \leftarrow f_V(\mathbf{pa}_V, \mathbf{u}_V)$ where $\mathbf{PA}_V \subseteq \mathbf{V} \setminus \{V\}$ and $\mathbf{U}_V \subseteq \mathbf{U}$; and iv) $P(\mathbf{U})$ is a joint probability distribution over $\mathbf{U}$. Intervening on $\mathbf{X}$ by fixing it to $\mathbf{x}$, denoted by $do(\mathbf{X} = \mathbf{x}) = do(\mathbf{x})$, in $\mathcal{M}$ creates a submodel $\mathcal{M}_{\mathbf{x}} = \langle \mathbf{U}, \mathbf{V}, \mathbf{F}_{\mathbf{x}}, P(\mathbf{U}) \rangle$ where $\mathbf{F}_{\mathbf{x}}$ is $\mathbf{F}$ with f_X replaced by a constant x for every $X \in \mathbf{X}$. The submodel $\mathcal{M}_{\mathbf{x}}$ induces an interventional distribution $P_{\mathbf{x}}$, which is also denoted by $P(\cdot \mid do(\mathbf{x}))$. A SCM induces a causal diagram where its vertices correspond to $\mathbf{V}$, directed edges represent functional relationships as specified in $\mathbf{F}$, and each of bidirected edges portrays the existence of an unobserved confounder (UC) between the two vertices pointed by the edge. We will make extensive use of the *do*-calculus, which is a set of three rules that allow one to reason about invariances across observational and experimental distributions. For all the proofs and appendices, please refer to the full technical report (Lee, Correa, and Bareinboim 2020).

2 Towards General Transportability

In this section, we formalize the notion of general transportability and introduce some basic results. In particular, we will consider heterogeneous domains (i.e., environments, studies, or populations) $\Pi = \{\pi^1, \pi^2, \ldots, \pi^n\}$, where each associates with a SCM compatible with a common causal diagram $\mathcal{G}$. We fix π^1 as a *target* domain in which we are interested in answering a causal query, and others are *source* domains. Through out this paper, let $* = 1$ to emphasize the target domain, e.g., π^* or P^*. The distributions under $do(\mathbf{x})$ associated with π^i will be denoted by $P^i_{\mathbf{x}}$. Following the construction in (Bareinboim and Pearl 2012b), we formally characterize structural heterogeneity across domains:

Definition 1 (Domain Discrepancy). Let π^a and π^b be domains associated, respectively, with SCMs $\mathcal{M}^a$ and $\mathcal{M}^b$

[1]A number of special cases of this general treatment has been studied in the literature in the empirical sciences, including external validity (Campbell and Stanley 1963; Manski 2007), meta-analysis (Hedges and Olkin 1985), quasi-experiment (Shadish, Cook, and Campbell 2002), or heterogeneity (Morgan and Winship 2007).

[2]While it lies outside the scope of this paper to provide a survey of this body of literature, for the sake of clarity, we provide a short summary of the relationship of its main settings in Appendix.

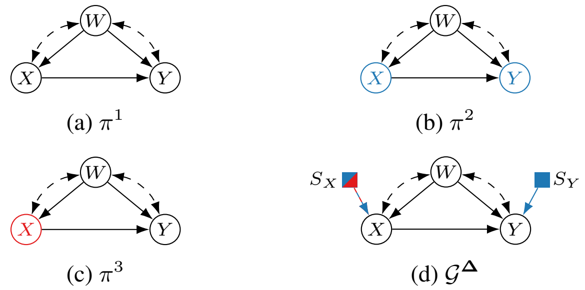


Figure 1: Causal graphs colored to depict the discrepancies between (a) a target domain and (b,c) two source domains where $\boldsymbol{\Delta} = \{\emptyset, \{X, Y\}, \{X\}\}$, which induces $\mathbf{S} = \{S_X, S_Y\}$ where $\mathbf{S}^2 = \{S_X, S_Y\}$ and $\mathbf{S}^3 = \{S_X\}$ and (d) a selection diagram $\mathcal{G}^{\boldsymbol{\Delta}}$.

conforming to a causal diagram $\mathcal{G}$. We denote by $\Delta^{a,b} \subseteq \mathbf{V}$ a set of variables such that, for every $V \in \Delta^{a,b}$, there might exist a discrepancy; either $f_V^a \neq f_V^b$ or $P^a(\mathbf{U}_V) \neq P^b(\mathbf{U}_V)$.

Further, the differences between the target and each of the source domains is represented in $\mathcal{G}$:

Definition 2 (Selection Diagram)**.** Given a collection of domain discrepancies $\boldsymbol{\Delta} = \{\Delta^{*,i}\}_{i=1}^n$ with regard to $\mathcal{G} = \langle \mathbf{V}, \mathbf{E} \rangle$, let $\mathbf{S} = \{S_V \mid \exists_{i=1}^n V \in \Delta^{*,i}\}$ be *selection variables*. Then, a *selection diagram* $\mathcal{G}^{\boldsymbol{\Delta}}$ is defined as a graph $\langle \mathbf{V} \cup \mathbf{S}, \mathbf{E} \cup \{S_V \to V\}_{S_V \in \mathbf{S}} \rangle$.

We shorten $\Delta^{*,i}$ as Δ^i to represent the differences between the target and each source domain. We denote domain-specific selection variables by $\mathbf{S}^i = \{S_V\}_{V \in \Delta^i}$, and the rest by $\mathbf{S}^{-i} = \mathbf{S} \setminus \mathbf{S}^i$. Selection variables work like switches selecting the domain of interest. The state space of $S_V \in \mathbf{S}$ is $\{1\} \cup \{i \mid V \in \Delta^i \in \boldsymbol{\Delta}\}$. Therefore, a selection diagram can be viewed as the causal diagram for a unifying SCM[3] representing heterogeneous SCMs where $P_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w}, \mathbf{s}^i = \mathbf{i}, \mathbf{s}^{-i} = \mathbf{1}) = P^i_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$.

For example, we illustrate in Figs. 1a to 1c a common causal graph $\mathcal{G}$ among three domains with different colors to highlight discrepancies between the target and source domains. This corresponds to $\boldsymbol{\Delta} = \{\emptyset, \{X, Y\}, \{X\}\}$, which entails the selection diagram $\mathcal{G}^{\boldsymbol{\Delta}}$ in Fig. 1d. We are now ready to define the most general transportability instance that will be investigated in this paper, namely:

Definition 3 (g-Transportability)**.** Let $\mathcal{G}^{\boldsymbol{\Delta}}$ be a selection diagram relative to domains $\Pi = \{\pi^i\}_{i=1}^n$ with a target domain π^*. Let $\mathbb{Z} = \{\mathbb{Z}^i\}_{i=1}^n$ be a specification of available experiments, where $\mathbb{Z}^i$ is the collection of sets of variables for π^i in which experiments on each set of variables $\mathbf{Z} \in \mathbb{Z}^i$ can be conducted. Given disjoint sets of variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{W}$, the conditional causal effect $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is said to be g-transportable given $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$ if $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is uniquely computable from $\mathbb{P}^{\Pi}_{\mathbb{Z}} = \{P^i_{\mathbf{z}} \mid \mathbf{z} \in \mathfrak{X}_{\mathbf{Z}}, \mathbf{Z} \in \mathbb{Z}^i \in \mathbb{Z}\}$ in any collection of models that induce $\mathcal{G}^{\boldsymbol{\Delta}}$.

The problem can be seen as asking about the existence of a functor g that outputs a universal formula given $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$, which takes $\mathbb{P}^{\Pi}_{\mathbb{Z}}$ and returns $P^*_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w})$, i.e., $\exists_g P^*_{\mathbf{x}}(\mathbf{y} \mid \mathbf{w}) = g(\mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z})(\mathbb{P}^{\Pi}_{\mathbb{Z}})$. Again, considering the selection diagram in Fig. 1d with $\mathbb{Z} = \{\emptyset, \{\{Y\}\}, \{\{X\}\}\}$, we can derive

$$P^*_x(y|w) = \frac{P^*_x(y, w)}{P^*_x(w)} = \frac{P^3_x(y, w)}{P^*_x(w)} = \frac{P^3_x(y, w)}{P^2_y(w)} \qquad (1)$$

To witness, note that the first equality follows from the definition of conditional probability, the second one is due to the irrelevance of the different X mechanisms between π^* and π^3 under $do(x)$, and the last one is based on Rule 3 (removing $do(x)$ and adding $do(y)$) together with W being indifferent to the disparities on f_X and f_Y between π^* and π^2. The following lemma provides a way to determine whether a query $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is g-transportable given $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$ based on the selection diagram.

Lemma 1. *A causal effect $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is g-transportable with respect to $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$ if the expression $P_{\mathbf{x}}(\mathbf{y}|\mathbf{w}, \mathbf{S})$ is reducible to an expression in which every term of the form $P_{\mathbf{a}}(\mathbf{b}|\mathbf{c}, \mathbf{S}')$ satisfies $(\mathbf{S} \setminus \mathbf{S}' \perp\!\!\!\perp \mathbf{B} \mid \mathbf{C})$ in $\mathcal{G}^{\boldsymbol{\Delta}} \setminus \mathbf{A}$, $\mathbf{S}^i \cap \mathbf{S}' = \emptyset$, and $\mathbf{A} \in \mathbb{Z}^i$ for some domain $\pi^i \in \Pi$.*

Proof. The condition implies that $P_{\mathbf{x}}(\mathbf{y}|\mathbf{w}, \mathbf{s}{=}\mathbf{1})$ can be written as an expression with terms, e.g. $P_{\mathbf{a}}(\mathbf{b}|\mathbf{c}, \mathbf{s}'{=}\mathbf{1})$, and further entails that $P_{\mathbf{a}}(\mathbf{b}|\mathbf{c}, \mathbf{s}'{=}\mathbf{1}) = P_{\mathbf{a}}(\mathbf{b}|\mathbf{c}, \mathbf{s}^{-i}{=}\mathbf{1}, \mathbf{s}^i{=}\mathbf{i}) = P^i_{\mathbf{a}}(\mathbf{b}|\mathbf{c})$ for any π^i such that $\mathbf{S}^i \cap \mathbf{S}' = \emptyset$. Since $P^i_{\mathbf{a}} \in \mathbb{P}^{\Pi}_{\mathbb{Z}}$, the expression uniquely computes $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ with $\mathbb{P}^{\Pi}_{\mathbb{Z}}$. $\square$

The previous example in Eq. (1) on Fig. 1 can be rewritten by explicitly employing the selection variables to articulate the applications of *do*-calculus and axioms of probability:

$$P_x(y|w, \mathbf{S}) = \frac{P_x(y, w|\mathbf{S})}{P_x(w|\mathbf{S})} = \frac{P_x(y, w|S_Y)}{P_x(w)} = \frac{P^3_x(y, w)}{P^2_y(w)}$$

For instance, $P_x(y, w|S_Y) = P^3_x(y, w)$ due to $\{S_Y\} \subseteq \mathbf{S}^{-3} = \{S_X, S_Y\} \setminus \{S_X\}$. We next characterize non-g-transportability of a conditional causal effect:

Lemma 2. *A causal effect $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is not g-transportable with respect to $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$, if there exist two SCMs compatible with $\mathcal{G}^{\boldsymbol{\Delta}}$ where both agree on $\mathbb{P}^{\Pi}_{\mathbb{Z}}$ while disagreeing on $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$.*

Proof. Having two different values for the query $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ rules out the existence of a valid function mapping from $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$ to the conditional causal effect. $\square$

The conditional causal effect $P^*_x(y|w)$ shown in Fig. 1 would not be g-transportable if π^3 associates with an observational distribution without an experiment on X, i.e., $\mathbb{Z}^3 = \{\emptyset\}$; or if its mechanism on W disagrees with π^*, i.e., $\Delta^3 = \{W\}$. We will provide a graphical criterion for the non-g-transportability of a query in Sec. 3 based on Lemma 2, and devise a sound and complete algorithm for the problem of g-transportability in Sec. 4 grounded on Lemma 1 and the results in Sec. 3.

[3] One can construct a SCM $\mathcal{M} = \langle \cup_i \mathbf{U}^i, \mathbf{V} \cup \mathbf{S}, \mathbf{F}, \prod_i P^i(\mathbf{U}^i) \rangle$ where $\mathbf{F}$ is the same as the one in $\mathcal{M}^1$ except $X \in \mathbf{V}$ such that $S_X \in \mathbf{S}$. For such a variable X, adopt $X = f_X^{S_X}(\mathbf{PA}_X, \mathbf{U}_X^{S_X})$, which selects the given domain's function as specified by S_X.

3 A Graphical Criterion for Non-g-transportability

We present a graphical criterion which can tell whether, a conditional causal effect is not g-transportable. We first examine the case of an unconditional causal effect (Sec. 3.1). The results established for the unconditional case are foundational in investigating the conditional one (Sec. 3.2).

3.1 Non-g-transportability of an Unconditional Interventional Distribution

We investigate a graphical characterization of non-g-transportability of an unconditional causal effect given $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$. We formally introduce essential notions devised in the identifiability literature (Tian and Pearl 2002; Shpitser and Pearl 2006b) with slight revisions. A subgraph of $\mathcal{G}$ is called a *C-component* (Tian 2002; Tian and Pearl 2002) if its bidirected edges form a spanning tree over all vertices in the subgraph. A graph $\mathcal{G}$ can be decomposed into a set of maximal C-components. We denote by $\mathcal{C}(\mathcal{G})$ the decomposition of $\mathbf{V}$ with respect to maximal C-components. An $\mathbf{R}$-rooted *C-forest* is a C-component whose root set is $\mathbf{R}$ and edges are minimal such that every vertex other than $\mathbf{R}$ has one child and bidirected arcs form a spanning tree. A pair of C-forests with an inclusive relationship, often denoted by $\langle \mathcal{F}, \mathcal{F}' \rangle$ such that $\mathcal{F}' \subseteq \mathcal{F}$, sharing the same roots is called a *hedge*. If there exists an $\mathbf{R}$-rooted hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ in $\mathcal{G}$ with $\mathbf{R} \subseteq An(\mathbf{Y})_{\mathcal{G} \setminus \mathbf{X}}$, $\mathbf{X} \cap \mathcal{F} \neq \emptyset$, and $\mathbf{X} \cap \mathcal{F}' = \emptyset$, then we say that $\langle \mathcal{F}, \mathcal{F}' \rangle$ is formed for $P^*_{\mathbf{x}}(\mathbf{y})$, which implies that the same effect is not identifiable in $\mathcal{G}$ from P (Shpitser and Pearl 2006b). For example, $\mathcal{F}_a$ in Fig. 2b is a $\{Y_1, R, Y_2\}$-rooted C-forest. The subgraph made of this root-set alone is also a $\{Y_1, R, Y_2\}$-rooted C-forest. That is, the pair $\langle \mathcal{F}_a, \mathcal{F}_a[\{Y_1, R, Y_2\}] \rangle$ is a hedge, which is formed for $P^*_{x_1}(y_1, y_2)$ in $\mathcal{G}$ (but not for $P^*_{x_1}(y_1)$).

Thicket is a graphical structure which affirms the non-identifiability of $P^*_{\mathbf{x}}(\mathbf{y})$ with $\langle \mathcal{G}^{\{\emptyset\}}, \{\mathbb{Z}^*\} \rangle$ (i.e., a single domain with an arbitrary collection of experiments) (Lee, Correa, and Bareinboim 2019). We introduce the notion of *s-thicket*, a generalization of a thicket to a heterogeneous setting by taking selection variables into account:

Definition 4 (s-Thicket)**.** Given $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$, an s-thicket $\mathcal{T}$ is a minimal non-empty $\mathbf{R}$-rooted C-component of $\mathcal{G}$ such that for each $\mathbf{Z} \in \mathbb{Z}^i \in \mathbb{Z}$, either (a) $\Delta^i \cap \mathbf{R} \neq \emptyset$, (b) $\mathbf{Z} \cap \mathbf{R} \neq \emptyset$, or (c) there exists $\mathcal{F} \subseteq \mathcal{T} \setminus \mathbf{Z}$ where $\langle \mathcal{F}, \mathcal{T}[\mathbf{R}] \rangle$ is a hedge. If $\mathbf{R} \subseteq An(\mathbf{Y})_{\mathcal{G} \setminus \mathbf{X}}$ and every *hedgelet* of the hedges intersects with $\mathbf{X}$, we say an s-thicket $\mathcal{T}$ is formed for $P^*_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}^{\boldsymbol{\Delta}}$ with respect to $\mathbb{Z}$.

Definition 5 (hedgelet decomposition)**.** The hedgelet decomposition $\mathbb{H}(\langle \mathcal{F}, \mathcal{F}' \rangle)$ of a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ is the collection of hedgelets $\{\mathcal{F}(\mathbf{T})\}_{\mathbf{T} \in \mathcal{C}(\mathcal{F} \setminus \mathcal{F}')}$ where each hedgelet $\mathcal{F}(\mathbf{T})$ is a subgraph of $\mathcal{F}$ made of (i) $\mathcal{F}[\mathbf{V}(\mathcal{F}') \cup \mathbf{T}]$ and (ii) $\mathcal{F}[De(\mathbf{T})_{\mathcal{F}}]$ without bidirected edges.

An s-thicket is a superimposition of hedges sharing a common root-set, where each hedge is also a superimposition of hedgelets. Intuitively speaking, if we encounter an s-thicket $\mathcal{T}$ for $P^*_{\mathbf{x}'}(\mathbf{y}')$ in $\mathcal{G}$, g-transporting $P^*_{\mathbf{x}'}(\mathbf{r})$, where $\mathbf{X}' = \mathbf{X} \cap \mathcal{T}$, is hindered because every existing experimental distribution either (a) exhibits discrepancies, (b) is based on an intervention on the variables we wish to measure, or (c) is not sufficient to pinpoint $P^*_{\mathbf{x}'}(\mathbf{r})$. Further, $P^*_{\mathbf{x}}(\mathbf{y})$ is not g-transportable since the negative result for $P^*_{\mathbf{x}'}(\mathbf{r})$ can be mapped to that for $P^*_{\mathbf{x}'}(\mathbf{y}')$ where $\mathbf{Y}' \subseteq \mathbf{Y}$ and $\mathbf{R} \subseteq An(\mathbf{Y}')_{\mathcal{G} \setminus \mathbf{X}}$.

Consider, for example, the causal graph $\mathcal{G}$ in Fig. 2a where $\boldsymbol{\Delta} = \{\emptyset, \{B\}\}$ and $\mathbb{Z} = \{\{\{C\}\}, \{\{X_1\}, \{X_3, R\}\}\}$. $\mathcal{G}$ without $R \to Y_2$ is an s-thicket for $P^*_{\mathbf{x}}(\mathbf{y})$ with respect to $\langle \mathcal{G}^{\boldsymbol{\Delta}}, \mathbb{Z} \rangle$. First, an experiment on $\{X_3, R\}$ matches (b) in Def. 4. Since the other two experiments do not match (a) nor (b) in Def. 4, there should be two hedges which do not intersect with C and X_1, respectively (Fig. 2b and Fig. 2c). The former, which disjoints with $\{C\}$, is also its only hedgelet. The latter, which does not contain $\{X_1\}$, is composed of two hedgelets based on the C-component decomposition of its top (i.e., the subgraph induced by removing its root-set) $\mathcal{C}(\mathcal{F}_b[\{B, C, D, X_2, X_3\}]) = \{\{B, C, X_3\}, \{D, X_2\}\}$. Now, we formally establish a connection between an s-thicket and the non-g-transportability of a query:

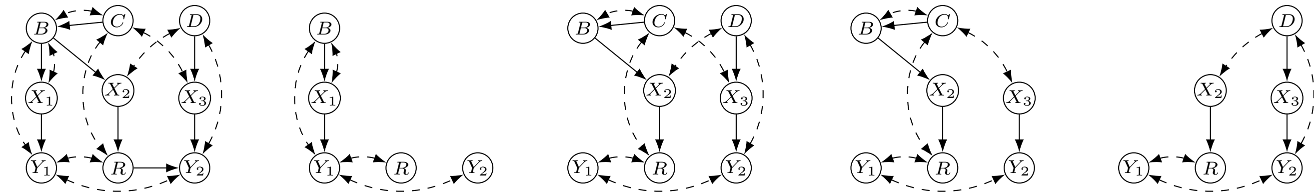
Lemma 3. *With respect to $\mathcal{G}^{\boldsymbol{\Delta}}$ and $\mathbb{Z}$, a causal effect $P^*_{\mathbf{x}}(\mathbf{y})$ is not g-transportable if there exists an s-thicket $\mathcal{T}$ formed for the causal effect.*

Proof sketch. Treating multiple domains as if they are *homogeneous*, the existence of $\mathcal{T}$ entails the existence of two models witnessing the non-g-transportability of $P^*_{\mathbf{x}'}(\mathbf{r})$, for some $\mathbf{X}' \subseteq \mathbf{X}$, from $\mathcal{G}^{\{\emptyset\}}$ and $\{\bigcup_i \mathbb{Z}^i\}$ (Lee, Correa, and Bareinboim 2019). However, the same models will not necessarily agree on some of distributions available in source domains. We incorporate selection variables into the parametrization to make the two models agree on $\mathbb{P}^{\Pi}_{\mathbb{Z}}$ while still disagreeing on $P^*_{\mathbf{x}'}(\mathbf{r})$. The parametrization (Lee, Correa, and Bareinboim 2019) is designed to produce the same distributions for the two models if at least one $R \in \mathbf{R}$ becomes independent to the UCs among $\mathbf{R}$, which isn't the case for $do(\mathbf{x})$. We modify each function for $R \in \mathbf{R}$ to return 0 when $S_R \neq 1$.[4] Consequently, the two models witness the non-g-transportability of $P^*_{\mathbf{x}'}(\mathbf{r})$, and the result will entail the same for $P^*_{\mathbf{x}'}(\mathbf{y}')$ in $\mathcal{T}'$, a graph where $\mathcal{T}$ is extended by adding directed paths from $\mathbf{R}$ to $\mathbf{Y}' \subseteq \mathbf{Y}$. $\square$

At this point, the non-existence of an s-thicket is a necessary condition for the g-transportability of an unconditional causal effect. In Sec. 4 we will further show that this is sufficient too, by presenting an algorithm that returns a valid formula for the target effect whenever no s-thicket exists (Thm. 3). For the sake of a better presentation of the completeness of the graphical criterion for the conditional case in the next section, we put a corollary below based on Lemma 3 and Thm. 3 in the next section:

Corollary 1. *With respect to $\mathcal{G}^{\boldsymbol{\Delta}}$ and $\mathbb{Z}$, a causal effect $P^*_{\mathbf{x}}(\mathbf{y})$ is not g-transportable if and only if there exists an s-thicket $\mathcal{T}$ formed for the causal effect.*

[4] One can replace the constant 0 to an R-specific unobserved variables, which can be an (un)fair coin.



(a) $\mathcal{G}$ (b) $\mathcal{F}_a$ (c) $\mathcal{F}_b$ (d) $\mathcal{F}_b(\{B, C, X_3\})$ (e) $\mathcal{F}_b(\{D, X_2\})$

Figure 2: (a) A causal graph $\mathcal{G}$, which, without $R \to Y_2$, forms an s-thicket for $P^*_{\mathbf{x}}(\mathbf{y})$ given $\mathbf{\Delta} = \{\emptyset, \{B\}\}$ and $\mathbb{Z} = \{\{\{C\}\}, \{\{X_1\}, \{X_3, R\}\}\}$. The s-thicket is the superimposition of two hedges (b, c) where the latter further decomposed into two hedgelets (d, e).

3.2 Non-g-transportability of a Conditional Interventional Distribution

We proceed to the graphical criterion for the g-transportation of $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$. We will assume that the query under consideration is *conditionally minimal* in the sense that there is no $W \in \mathbf{W}$ such that $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w}) = P^*_{\mathbf{x}\cup\{w\}}(\mathbf{y}|\mathbf{w} \setminus \{w\})$ by virtue of Rule 2 of *do*-calculus. Otherwise, we can repeatedly apply Rule 2 and obtain an equivalent minimal expression $P^*_{\mathbf{x},\mathbf{w}'}(\mathbf{y}|\mathbf{w} \setminus \mathbf{w}')$ (Cor. 1 (Shpitser and Pearl 2006a)). The conditional minimality is graphically translated to the existence of an active backdoor path from each of $W \in \mathbf{W}$ to some $Y \in \mathbf{Y}$ given $\mathbf{W} \setminus \{W\}$. We present a major theoretical result which authorizes the delegation of the characterization of a conditional causal effect to that of an unconditional one:

Theorem 1. *Let every $W \in \mathbf{W}$ have a backdoor path to $\mathbf{Y}$ in $\mathcal{G} \setminus \mathbf{X}$ active given $\mathbf{W} \setminus \{W\}$. A query $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is g-transportable if and only if $P^*_{\mathbf{x}}(\mathbf{y}, \mathbf{w})$ is g-transportable with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$.*

The sufficiency holds true since $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w}) = P^*_{\mathbf{x}}(\mathbf{y}, \mathbf{w}) / \sum_{\mathbf{y}} P^*_{\mathbf{x}}(\mathbf{y}, \mathbf{w})$. As for the necessity, suppose $P^*_{\mathbf{x}}(\mathbf{y}, \mathbf{w})$ is not g-transportable. If $P^*_{\mathbf{x}}(\mathbf{w})$ is g-transportable, then $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ must be non-g-transportable, otherwise a contradiction arises since $P^*_{\mathbf{x}}(\mathbf{y}, \mathbf{w})$ would be g-transportable as $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w}) P^*_{\mathbf{x}}(\mathbf{w})$. Then, it remains to prove that $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is not g-transportable whenever $P^*_{\mathbf{x}}(\mathbf{w})$ is not g-transportable with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$. Indeed, that is the case, as follows:

Theorem 2. *Let every $W \in \mathbf{W}$ have a backdoor path to $\mathbf{Y}$ in $\mathcal{G} \setminus \mathbf{X}$ active given $\mathbf{W} \setminus \{W\}$. A query $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is not g-transportable if $P^*_{\mathbf{x}}(\mathbf{w})$ is not g-transportable with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$.*

Proof sketch. Let $\mathcal{T}'$ be a subgraph of $\mathcal{G}$ parametrized to demonstrate the non-g-transportability of $P^*_{\mathbf{x}'}(\mathbf{w}')$ given $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$ (Lemma 3). Pick some $W \in \mathbf{W}'$ that is also in the root-set of $\mathcal{T}'$, and fix a minimal subgraph $\mathcal{P} \subseteq \mathcal{G} \setminus \mathbf{X}$ witnessing an active backdoor path from W to some $Y \in \mathbf{Y}$ given $\mathbf{W} \setminus \{W\}$. $\mathcal{P}$ also includes any directed path from an active collider in $\mathcal{P}$ to its descendant in $\mathbf{W} \setminus \{W\}$. We construct two models for $\mathcal{T}' \cup \mathcal{P}$ while preserving the mechanisms in Lemma 3. We augment the exclusive-or-based parametrization for variables in $\mathcal{P}$ so that W and Y are correlated given $(\mathbf{W} \cap \mathcal{P}) \setminus \{W\}$. In the augmented models, the value of W is determined as the exclusive-or of two Ws computed in $\mathcal{T}'$ and in $\mathcal{P}$. The resultant models will disagree on $P^*_{\mathbf{x}'}(y|\mathbf{w}'')$ where $\mathbf{W}''$ is the subset of $\mathbf{W}$ in $\mathcal{T}' \cup \mathcal{P}$. Therefore, $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ is not g-transportable with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$. $\square$

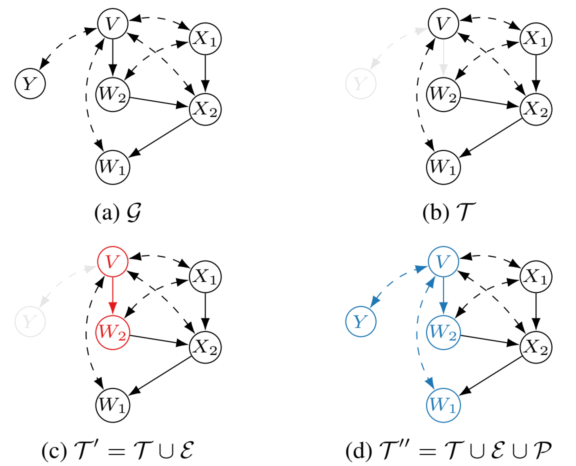


Figure 3: A causal diagram $\mathcal{G}$, and causal diagrams illustrating the phases of a non-g-transportability parametrization for $P^*_{\mathbf{x}}(y|\mathbf{w})$. (b) an s-thicket for $P^*_{\mathbf{x}}(\mathbf{w})$ given P^*, (c) the s-thicket with an extension in red, and (d) a path-witnessing subgraph (blue) augmented extended s-thicket.

We provide an illustrative example in Fig. 3. For the sake of brevity, we assume a single domain setting with P^* available. Given a causal graph $\mathcal{G}$ (Fig. 3a) and P^*, an s-thicket $\mathcal{T}$ is formed for $P^*_{\mathbf{x}}(\mathbf{w})$ (Fig. 3b). Two models are first constructed to disagree on $P^*_{\mathbf{x}}(v, w_1)$. Then, the result is mapped to $P^*_{\mathbf{x}}(\mathbf{w})$ via a graph $\mathcal{E}$ (red), resulting in a parametrization for $\mathcal{T}' = \mathcal{T} \cup \mathcal{E}$ (Fig. 3c). Pick $W_1 \in \mathbf{W}$, which is the only $\mathbf{W}$ in the root set of $\mathcal{T}'$, then find a backdoor path to Y given $\mathbf{W} \setminus \{W_1\}$. The path-witnessing subgraph $\mathcal{P} \in \mathcal{G}$ is shown in blue (Fig. 3d). A separate parametrization for $\mathcal{P}$ is merged with that for $\mathcal{T}'$ via an exclusive-or on W_1. Then, the two models disagree on $P^*_{\mathbf{x}}(y|\mathbf{w})$.

Algorithm 1 GTR and GTRU, sound and complete g-transportability algorithms.

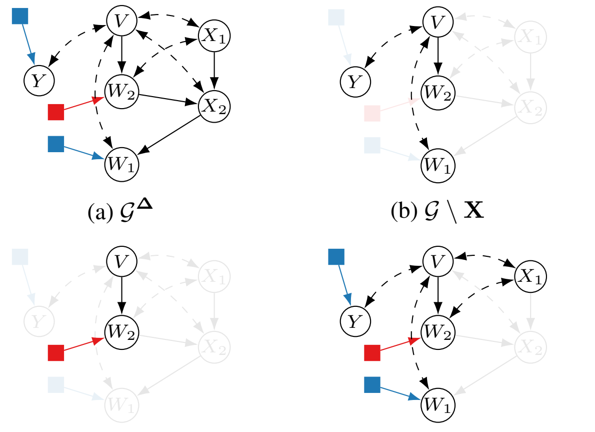
1: **function** GTR($\mathbf{y}, \mathbf{x}, \mathbf{w}, \mathcal{G}, \mathbf{\Delta}$)
input: $\mathbf{y}$, $\mathbf{x}$, $\mathbf{w}$: values for a query $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$; $\mathcal{G}$: causal diagram; $\mathbf{\Delta}$: domain discrepancies. A specification of available experiments $\mathbb{Z}$, and the distributions for those experiments, $\mathbb{P}^{\Pi}_{\mathbb{Z}}$, are globally defined.
output: an estimator computing $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$.
2: **if** $\exists_{W \in \mathbf{W}}(W \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{W} \setminus \{W\})_{(\mathcal{G}\setminus\mathbf{X})_{\underline{W}}}$ **then**
return GTR($\mathbf{y}, \mathbf{x} \cup \{w\}, \mathbf{w} \setminus \{w\}, \mathcal{G}, \mathbf{\Delta}$).
3: **else**
return $Q / \sum_{\mathbf{y}} Q$ **where** $Q \leftarrow$ GTRU($\mathbf{y} \cup \mathbf{w}, \mathbf{x}, \mathcal{G}, \mathbf{\Delta}$).

4: **function** GTRU($\mathbf{y}, \mathbf{x}, \mathcal{G}, \mathbf{\Delta}$)
output: an estimator computing $P^*_{\mathbf{x}}(\mathbf{y})$.
5: **if** $\exists_{\mathbf{Z} \in \mathbb{Z}^i \in \mathbb{Z}}(\mathbf{X} = \mathbf{Z} \cap \mathbf{V}) \wedge (\mathbf{S}^i \perp\!\!\!\perp \mathbf{Y})_{\mathcal{G}^{\mathbf{\Delta}}\setminus\mathbf{X}}$ **then**
return $P^i_{\mathbf{z}\setminus\mathbf{V},\mathbf{x}}(\mathbf{y})$.
6: **if** $(\mathbf{V}' \leftarrow \mathbf{V} \setminus An(\mathbf{Y})_{\mathcal{G}}) \neq \emptyset$ **then**
return GTRU($\mathbf{y}, \mathbf{x} \setminus \mathbf{V}', \mathcal{G} \setminus \mathbf{V}', \{\Delta^i \setminus \mathbf{V}' \mid \Delta^i \in \mathbf{\Delta}\}$).
7: **if** $(\mathbf{V}' \leftarrow (\mathbf{V} \setminus \mathbf{X}) \setminus An(\mathbf{Y})_{\mathcal{G}_{\overline{\mathbf{X}}}}) \neq \emptyset$ **then**
return GTRU($\mathbf{y}, \mathbf{x} \cup \mathbf{v}', \mathcal{G}, \mathbf{\Delta}$).
8: **if** $|\mathcal{C}(\mathcal{G} \setminus \mathbf{X})| > 1$ **then**
return $\sum_{\mathbf{v}\setminus(\mathbf{y}\cup\mathbf{x})} \prod_{\mathbf{C} \in \mathcal{C}(\mathcal{G}\setminus\mathbf{X})}$ GTRU($\mathbf{c}, \mathbf{v} \setminus \mathbf{c}, \mathcal{G}, \mathbf{\Delta}$).
9: **for** $\pi^i \in \Pi$ such that $(\mathbf{S}^i \perp\!\!\!\perp \mathbf{Y})_{\mathcal{G}^{\mathbf{\Delta}}\setminus\mathbf{X}}$, **for** $\mathbf{Z} \in \mathbb{Z}^i$ such that $\mathbf{Z} \cap \mathbf{V} \subseteq \mathbf{X}$ **do**
10: **return** ID($\mathbf{y}, \mathbf{x} \setminus \mathbf{Z}, P^i_{\mathbf{z}\setminus\mathbf{V},\mathbf{x}\cap\mathbf{Z}}, \mathcal{G} \setminus (\mathbf{Z} \cap \mathbf{X})$) **unless** FAIL is returned.
11: **throw** FAIL

4 A Sound and Complete Algorithm for g-Transportability

We present GTR (Alg. 1), a sound and complete algorithm capable of solving g-transportability instances, which smoothly and effectively combines ideas underlying in ID, IDC, MZTR, and GID (algorithms in (Shpitser and Pearl 2006b; 2006a; Bareinboim and Pearl 2014; Lee, Correa, and Bareinboim 2019)) and outputs an estimator for a given conditional interventional query $P^*_{\mathbf{x}}(\mathbf{y}|\mathbf{w})$ in a target domain with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$, if feasible. The experiment specification $\mathbb{Z}$ and the corresponding distributions $\mathbb{P}^{\Pi}_{\mathbb{Z}}$ are defined globally, and do not change with the specific invocation of the algorithm. In contrast, variables $\mathbf{V}$ and selection variables $\mathbf{S}$ reflect graph $\mathcal{G}$ and discrepancies $\mathbf{\Delta}$, respectively, relative to the arguments passed to the current execution of the procedure.

We give a line by line description where symbols such as $\mathcal{G}$, $\mathbf{V}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{W}$ are to be interpreted relative to the current arguments of the algorithm. At Line 2, GTR, recursively transforms the given query using Rule 2 of *do*-calculus to guarantee it is conditionally minimal (and satisfies the requirement for Thm. 1). With this guarantee, the algorithm (Line 3) delegates the identification of the query, based on the definition of conditional probability, to GTRU, which handles unconditional queries. Overall, GTRU transforms the given unconditional query and divides the problem into the identification of (simpler) subqueries. Each subproblem is delegated to ID with a distribution $P^i_{\mathbf{z}}$ under some constraints on the domain π^i and the experiments on $\mathbf{Z} \in \mathbb{Z}^i$. Line 5, which is optional, checks whether an available distribution can be used to answer the query directly, i.e., $P^i_{\mathbf{z}}(\mathbf{y}) = P^*_{\mathbf{x}}(\mathbf{y})$, so as to return an estimator at an early stage. Line 6 narrows the scope of the problem by excluding variables that do not affect $\mathbf{Y}$ (Rule 3). Domain discrepancies are updated accordingly, since selection variables outside the scope have no effect on $\mathbf{Y}$. Line 7 maximizes the intervention set, which helps solving the problem, based on Rule 3. Line 8 breaks down the query into queries where $\mathbf{Y}$ in each subquery forms a C-component (Tian and Pearl 2002). Line 9 examines whether some experimental distribution $P^i_{\mathbf{z}} \in \mathbb{P}^{\Pi}_{\mathbb{Z}}$ can be used to identify the query. If valid, GTRU passes the query to ID with a slight modification of it and graph, taking into account the shared intervention between $\mathbf{Z}$ and $\mathbf{X}$. GTR runs in $O(v^4 z)$ where $v = |\mathbf{V}|$ and $z = \sum_i |\mathbb{Z}^i|$ (see Appendix for details).



(a) $\mathcal{G}^{\mathbf{\Delta}}$ (b) $\mathcal{G} \setminus \mathbf{X}$ (c) $\mathcal{G}^{\mathbf{\Delta}}[An(W_2)_{\mathcal{G}}]$ (d) $\mathcal{G}^{\mathbf{\Delta}} \setminus \{X_2\}$

Figure 4: (a) A selection diagram $\mathcal{G}^{\mathbf{\Delta}}$ where $\mathbf{\Delta} = \{\emptyset, \{W_1, Y\}, \{W_2\}\}$ and $\mathbb{Z} = \{\emptyset, \{\emptyset\}, \{\{X_2\}\}\}$. (b,c,d) Graphs encountered during the execution of GTR to g-transport $P^*_{\mathbf{x}}(y|\mathbf{w})$.

We offer a running example regarding the identification of $P^*_{\mathbf{x}}(y|\mathbf{w})$ with a causal graph $\mathcal{G}$ (Fig. 3a), $\mathbf{\Delta} = \{\emptyset, \{W_1, Y\}, \{W_2\}\}$ (see $\mathcal{G}^{\mathbf{\Delta}}$ in Fig. 4a with $\mathbf{S}^2$ and $\mathbf{S}^3$ in blue and red), and $\mathbb{Z} = \{\emptyset, \{\emptyset\}, \{\{X_2\}\}\}$, i.e., the target domain has no distribution available while π^2 and π^3 provide an observational distribution and an experiment on X_2, respectively. Given a query $P^*_{\mathbf{x}}(y|\mathbf{w})$, GTR investigates whether there exists any $W \in \mathbf{W}$ that can be moved to the interventional part of the query. Fig. 4b shows $\mathcal{G} \setminus \mathbf{X}$ where the existence of a backdoor path between W and Y is figured out. Since $W_2 \leftarrow V \leftrightarrow Y$ and $W_1 \leftrightarrow V \leftrightarrow Y$ given W_2 as a descendant of the collider (V), it proceeds to identify $P^*_{\mathbf{x}}(y, \mathbf{w})$. GTRU attempts to refine the given graph with the ancestors of $\{Y, W_1, W_2\}$ (Line 6). Then, it checks whether the intervention $\{X_1, X_2\}$ is maximal. Next, it investigates the C-components of $\mathcal{G} \setminus \mathbf{X}$ (Fig. 4b). There are two C-components involving $\{W_2\}$ and $\{Y, V, W_1\}$. Hence, it factorizes the query to $P^*_{y,\mathbf{x},v,w_1}(w_2)$ and $P^*_{\mathbf{x},w_2}(y, v, w_1)$. The first query encounters Line 6 and it is refined, i.e., $P^*_{y,\mathbf{x},v,w_1}(w_2) = P^*_v(w_2)$ (Rule 3) with the graph in

Fig. 4c. The query will reach Line 10 since $\{S_{W_2}\} \subseteq \mathbf{S}^{-2}$ (Lemma 1) and, eventually, ID identifies $P^*_v(w_2) = P^2(w_2|v)$, which corresponds to Rule 2. The second query passes conditions in Lines 5 to 9 since $(\{Y, V, W_1\} \perp\!\!\!\perp S_{W_2})$ in $\mathcal{G}^{\mathbf{\Delta}} \setminus \{X_2\}$ (Fig. 4d). Then, it makes use of $P^3_{x_2}$, since $\{X_2\} \subseteq \mathbf{X} \cup \{W_2\}$, to identify $P^*_{\mathbf{x},w_2}(y, v, w_1)$, which corresponds to identifying $Q^*_{x_1,w_2}(y, v, w_1)$ with $Q^3 = P^3_{x_2}$ in $\mathcal{G}^{\mathbf{\Delta}} \setminus \{X_2\}$ (Bareinboim and Pearl 2012a).

Theorem 3. GTRU *is sound and complete.*

Proof. (soundness) Let a subscript ℓ denote variables and values local to the function. The soundness of the algorithm is partially proved (Lee, Correa, and Bareinboim 2019) excluding the case where distributions from the heterogeneous source domains are utilized. It is sufficient to prove that $P^*_{\mathbf{x}_\ell}(\mathbf{y}_\ell) = P^i_{\mathbf{x}_\ell}(\mathbf{y}_\ell)$ for Lines 5 and 9 where the identification of $P^*_{\mathbf{x}_\ell}(\mathbf{y}_\ell)$ is delegated to that of $P^i_{\mathbf{x}_\ell}(\mathbf{y}_\ell)$ with $P^i_{\mathbf{z}}$ for some $\mathbf{Z} \in \mathbb{Z}^i$. By Lemma 1, $P^*_{\mathbf{x}_\ell}(\mathbf{y}_\ell) = P_{\mathbf{x}_\ell}(\mathbf{y}_\ell \mid \mathbf{S} = \mathbf{1})$. Since $(\mathbf{S}^i_\ell \perp\!\!\!\perp \mathbf{Y}_\ell)$ in $\mathcal{G}^{\mathbf{\Delta}_\ell}_\ell \setminus \mathbf{X}_\ell$ implies $(\mathbf{S}^i \perp\!\!\!\perp \mathbf{Y}_\ell)$ in $\mathcal{G}^{\mathbf{\Delta}} \setminus \mathbf{X}_\ell$, the equality $P_{\mathbf{x}_\ell}(\mathbf{y}_\ell \mid \mathbf{S} = \mathbf{1}) = P_{\mathbf{x}_\ell}(\mathbf{y}_\ell \mid \mathbf{S}^{-i} = \mathbf{1})$ holds true. Therefore, the soundness follows.

(completeness) We show that whenever GTRU fails to g-transport a given query $P^*_{\mathbf{x}}(\mathbf{y})$, there exists an s-thicket for the given query (Lemma 3). Given that GTRU imposes one more condition $(\mathbf{S}^i_\ell \perp\!\!\!\perp \mathbf{Y}_\ell)$ in $\mathcal{G}^{\mathbf{\Delta}_\ell}_\ell \setminus \mathbf{X}_\ell$ at Line 9 compared to GID, those qualified experiments $\mathbf{Z} \in \mathbb{Z}^i \in \mathbb{Z}$ can be considered as experiments conducted in the target domain so that the identification is reducible to GID given $\mathcal{G}$ with the qualified experiments (Lee, Correa, and Bareinboim 2019). Hence, when the algorithm fails to identify the query, there exists a thicket for $P^*_{\mathbf{x}}(\mathbf{y})$ (Thm. 3 (Lee, Correa, and Bareinboim 2019)). If every experiment $\mathbf{Z}$ satisfies items (b) and (c) in Def. 4, then the thicket is an s-thicket. Otherwise, we map the existence of a thicket $\mathcal{T}^\dagger$ to that of an s-thicket $\mathcal{T}$ — it remains to show $\Delta^i \cap \mathbf{R} \neq \emptyset$ (item (a) in Def. 4). First, there exists an $\mathbf{R}^\dagger$-rooted thicket $\mathcal{T}^\dagger \subseteq \mathcal{G}_\ell$ for $P^*_{\mathbf{x}_\ell}(\mathbf{y}_\ell)$, which is also for $P^*_{\mathbf{x}}(\mathbf{y})$. Since $\mathbf{R}^\dagger \subseteq An(\mathbf{Y}_\ell)_{\mathcal{G}_\ell \setminus \mathbf{X}_\ell} = \mathbf{V}_\ell \setminus \mathbf{X}_\ell$ and $\mathcal{G}_\ell[\mathbf{V}_\ell \setminus \mathbf{X}_\ell]$ is a C-component (Line 8), the thicket $\mathcal{T}^\dagger$ with its root set replaced with $\mathbf{V}_\ell \setminus \mathbf{X}_\ell$ is a valid thicket. Then, due to Prop. 1 (below), the modified thicket is an s-thicket for $P^*_{\mathbf{x}}(\mathbf{y})$ with respect to $\langle \mathcal{G}^{\mathbf{\Delta}}, \mathbb{Z} \rangle$. □

Proposition 1. $(\mathbf{S}^i \perp\!\!\!\perp \mathbf{Y})_{\mathcal{G}^{\mathbf{\Delta}} \setminus \mathbf{X}}$ *at Line 9 is equivalent to* $\mathbf{\Delta}^i \cap (\mathbf{V} \setminus \mathbf{X}) = \emptyset$.

Corollary 2. GTR *is sound and complete.*

Proof. The soundness of GTR follows from the soundness of GTRU (Thm. 3) and Rule 2. Its completeness follows from the completeness of GTRU (Thm. 3) and Thm. 1. □

Corollary 3. *The rules of do-calculus together with standard probability manipulations are complete for establishing g-transportability of conditional interventional distributions.*

Proof. This is due to: (i) Rule 2 of *do*-calculus and the definition of conditional probability under intervention for transitioning a conditional query to an unconditional one; and (ii) Rule 1 of *do*-calculus to determine whether to utilize the source domains (n.b. the selection variables as a condition as in Lemma 1 is implicit) along with the completeness of *do*-calculus with respect to GID. □

5 Conclusions

We investigated the challenge of learning conditional causal effects through generalizing and synthesizing experimental findings from heterogeneous domains, which unified many threads in the causal identifiability and transportability literature (Tian and Pearl 2002; Shpitser and Pearl 2006b; Huang and Valtorta 2006; Shpitser and Pearl 2006a; Bareinboim and Pearl 2013b; 2013a; 2012a; 2014; Lee, Correa, and Bareinboim 2019). This setting has been called g-transportability (Def. 3). Concretely, we developed a general treatment to the g-transportability problem in two ways — first, as a complete graphical criterion, which leads to a novel parametrization strategy characterizing the g-transportability of any causal query (Lemma 3, Thm. 1, and Thm. 2); second, as an efficient algorithmic procedure (GTR, Alg. 1, Thm. 3, and Cor. 2), which synthesizes quantitative causal knowledge under the guidance of qualitative and transparent assumptions encoded as a causal graph. Further, we proved that Pearl's *do*-calculus is complete for this task (Cor. 3). We hope these new analytical tools can help lower the barrier for the broader research community to advance science through collaborative synthesis of shared datasets and knowledge (Perrino et al. 2013; Pearl and Mackenzie 2018).

Acknowledgments

This research is supported in parts by grants from NSF IIS-1704352, and IIS-1750807 (CAREER), IBM Research, and Adobe Research.

References

Bareinboim, E., and Pearl, J. 2012a. Causal inference by surrogate experiments: z-identifiability. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, 113–120. Corvallis, OR: AUAI Press.

Bareinboim, E., and Pearl, J. 2012b. Transportability of causal effects: Completeness results. In *Proceedings of The Twenty-Sixth Conference on Artificial Intelligence (AAAI 2012)*, 698–704.

Bareinboim, E., and Pearl, J. 2013a. Causal transportability with limited experiments. In *Proceedings of the Twenty-Seventh National Conference on Artificial Intelligence (AAAI 2013)*, 95–101. Menlo Park, CA: AAAI Press.

Bareinboim, E., and Pearl, J. 2013b. Meta-transportability of causal effects: A formal approach. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, 135–143. JMLR W&CP 31.

Bareinboim, E., and Pearl, J. 2014. Transportability from multiple environments with limited experiments: Completeness results. In *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc. 280–288.

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

Campbell, D., and Stanley, J. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Wadsworth Publishing.

Galles, D., and Pearl, J. 1995. Testing identifiability of causal effects. In Besnard, P., and Hanks, S., eds., *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI 1995)*. San Francisco: Morgan Kaufmann. 185–195.

Halpern, J. Y. 2000. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research* 12(1):317–337.

Hedges, L. V., and Olkin, I. 1985. *Statistical Methods for Meta-Analysis*. Academic Press.

Huang, Y., and Valtorta, M. 2006. Identifiability in causal bayesian networks: A sound and complete algorithm. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press. 1149–1156.

Keiding, N., and Louis, T. A. 2018. Web-based enrollment and other types of self-selection in surveys and studies: Consequences for generalizability. *Annual Review of Statistics and Its Application* 5(1):25–47.

Kuroki, M., and Miyakawa, M. 1999. Identifiability criteria for causal effects of joint interventions. *Journal of the Japan Statistical Society* 29:105–117.

Lee, S.; Correa, J. D.; and Bareinboim, E. 2019. General identifiability with arbitrary surrogate experiments. In *Proceedings of the Thirty-Fifth Conference Annual Conference on Uncertainty in Artificial Intelligence*. Corvallis, OR: AUAI Press.

Lee, S.; Correa, J. D.; and Bareinboim, E. 2020. Generalized transportability: Synthesis of experiments from heterogeneous domains. Technical Report R-52, Causal AI Lab, Department of Computer Science, Columbia University.

Lesko, C. R.; Buchanan, A. L.; Westreich, D.; Edwards, J. K.; Hudgens, M. G.; and Cole, S. R. 2017. Generalizing study results: a potential outcomes perspective. *Epidemiology (Cambridge, Mass.)* 28(4):553–561.

Manski, C. 2007. *Identification for Prediction and Decision*. Cambridge, Massachusetts: Harvard University Press.

Morgan, S., and Winship, C. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, New York, NY.

Pearl, J., and Bareinboim, E. 2011. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the Twenty-Fifth Conference on Artificial Intelligence*, 247–254.

Pearl, J., and Bareinboim, E. 2019. Note on ''Generalizability of Study Results''. *Epidemiology* 30(2).

Pearl, J., and Mackenzie, D. 2018. *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Pearl, J., and Robins, J. 1995. Probabilistic evaluation of sequential plans from causal models with hidden variables. In Besnard, P., and Hanks, S., eds., *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. San Francisco: Morgan Kaufmann. 444–453.

Pearl, J. 1993. From Bayesian networks to causal networks. In *Proceedings of the Adaptive Computing and Information Processing Seminar*, 25–27.

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

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pearl, J. 2015. Generalizing experimental findings. *Journal of Causal Inference* 3:259–266.

Perrino, T.; Howe, G.; Sperling, A.; Beardslee, W.; Sandler, I.; Shern, D.; Pantin, H.; Kaupert, S.; Cano, N.; Cruden, G.; Bandiera, F.; and Brown, C. H. 2013. Advancing science through collaborative data sharing and synthesis. *Perspectives on Psychological Science* 8(4):433–444.

Shadish, W.; Cook, T.; and Campbell, D. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin, second edition.

Shpitser, I., and Pearl, J. 2006a. Identification of conditional interventional distributions. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI 2006)*. Corvallis, OR: AUAI Press. 437–444.

Shpitser, I., and Pearl, J. 2006b. Identification of joint interventional distributions in recursive semi-Markovian causal models. In *Proceedings of The Twenty-First National Conference on Artificial Intelligence*, 1219–1226. AAAI Press.

Spirtes, P.; Glymour, C.; and Scheines, R. 2001. *Causation, Prediction, and Search*. Cambridge, MA: MIT Press, 2nd edition.

Tian, J., and Pearl, J. 2002. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, 567–573. Menlo Park, CA: AAAI Press/The MIT Press.

Tian, J. 2002. *Studies in Causal Reasoning and Learning*. Ph.D. Dissertation, Computer Science Department, University of California, Los Angeles, CA.

Westreich, D., and Edwards, J. K. 2015. Invited commentary: every good randomization deserves observation. *American Journal of Epidemiology* 182(10):857–860.

Westreich, D.; Edwards, J. K.; Lesko, C. R.; Stuart, E.; and Cole, S. R. 2017. Transportability of trial results using inverse odds of sampling weights. *American Journal of Epidemiology* 186(8):1010–1014.

Zhou, H. H.; Singh, V.; Johnson, S. C.; and Wahba, G. 2018. Statistical tests and identifiability conditions for pooling and analyzing multisite datasets. *Proceedings of the National Academy of Sciences* 115(7):1481–1486.