TECHNICAL REPORT
R-38
Nov, 2018

Identification of Causal Effect in the Presence of Selection Bias

Juan D. Correa
Computer Science Department
Purdue University
correagr@purdue.edu

Jin Tian
Department of Computer Science
Iowa State University
jtian@iastate.edu

Elias Bareinboim
Computer Science Department
Purdue University
eb@purdue.edu

Abstract

Cause-and-effect relations are one of the most valuable types of knowledge sought after throughout the data-driven sciences since they translate into stable and generalizable explanations as well as efficient and robust decision-making capabilities. Inferring these relations from data, however, is a challenging task. Two of the most common barriers to this goal are known as confounding and selection biases. The former stems from the systematic bias introduced during the treatment assignment, while the latter comes from the systematic bias during the collection of units into the sample. In this paper, we consider the problem of identifiability of causal effects when both confounding and selection biases are simultaneously present. We first investigate the problem of identifiability when all the available data is biased. We prove that the algorithm proposed by [Bareinboim and Tian, 2015] is, in fact, complete, namely, whenever the algorithm returns a failure condition, no identifiability claim about the causal relation can be made by any other method. We then generalize this setting to when, in addition to the biased data, another piece of external data is available, without bias. It may be the case that a subset of the covariates could be measured without bias (e.g., from census). We examine the problem of identifiability when a combination of biased and unbiased data is available. We propose a new algorithm that subsumes the current state-of-the-art method based on the back-door criterion.

Introduction

One prominent challenge shared throughout the empirical disciplines is to infer cause and effect relationships - for instance, one may need to determine how increasing the state's educational budget will bring about change in the average income of the population, whether exposing subjects to a new advertisement campaign would translate into additional sales revenue, or how patients will react to the decrease of the drug's dosage, would they still recover in acceptable health conditions? Despite the disparate nature of these questions in terms of subject matter, they evoke the same set of principles and formal machinery, which comes under the rubric of *causal inference* (Pearl 2000; Spirtes, Glymour, and Scheines 2001).

Causal inference is concerned with the potential mismatch between the inferential power of the collected data and the target inference. In practice, this is particularly relevant since data is almost invariably plagued with various biases, most prominently, confounding and selection. The former refers to the presence of a set of factors that affect both the action (also known as treatment) and the outcome, while the latter arises when the action, outcome, and other factors differentially affect the inclusion of subjects in the data sample (Bareinboim and Pearl 2016).

The problem of *identifiability* gives formal dressing to the issue of confounding (Pearl 2000, Ch. 3). Specifically, it is concerned with determining the effect of a treatment (X) on an outcome (Y), denoted $P(y|do(x))$ (for short, $P_x(y)$), based on the observational, non-experimental distribution $P(\mathbf{v})$ (where $\mathbf{V}$ represents observable variables) and causal assumptions commonly expressed as a directed acyclic graph. The difference between $P(y|do(x))$ and its probabilistic counterpart, $P(y|x)$, is what is called *confounding bias* (Bareinboim and Pearl 2016). This problem has been extensively studied in the literature. A systematic treatment of this problem was given in (Pearl 1995), which introduced *do-calculus*. The do-calculus was shown to be complete for non-parametric identifiability from observational and experimental data (Tian and Pearl 2002a; Huang and Valtorta 2006; Shpitser and Pearl 2006; Bareinboim and Pearl 2012a).

The other source of disparities, *selection bias*, usually appears due to the preferential exclusion of units from the sample. For instance, in a typical study of the effect of grades on college admissions, subjects with higher achievement tend to report their scores more frequently than those who scored lower. In this case, the data-gathering process will reflect a distortion in the sample's proportions and, since the data is no longer a faithful representation of the underlying population, biased estimates will be produced regardless of the number of samples collected (even when the treatment is controlled). The problem of selection bias can also be modeled graphically through the explicit articulation of the sampling mechanism, S. This mechanism can be seen as a binary indicator of entry into the data pool, such that $S{=}1$ if a unit is included in the sample and $S{=}0$ otherwise. Clearly, when the sampling process is entirely random, S is independent of all variables in the analysis. When samples are collected preferentially, the causal effects not only need to be identified but also *recovered* from the distribution

$P(\mathbf{v}|S{=}1)$, instead of $P(\mathbf{v})$ (Bareinboim and Pearl 2012b).

Selection bias has challenged inferences throughout a wide range of disciplines, including AI (Cooper 1995; Elkan 2001; Zadrozny 2004; Cortes et al. 2008), statistics (Whittemore 1978; Little and Rubin 1987; Robinson and Jewell 1991; Kuroki and Cai 2006; Evans and Didelez 2015), and the empirical sciences (e.g., genetics (Pirinen, Donnelly, and Spencer 2012; Mefford and Witte 2012), economics (Heckman 1979; Angrist 1997), and epidemiology (Robins 2001; Glymour and Greenland 2008)).

Even though selection and confounding biases appear together in most of the non-trivial, practical settings, they have been almost invariably treated independently in the literature. There are non-trivial interactions between them, however, which have not been investigated until recently. (Bareinboim, Tian, and Pearl 2014; Bareinboim and Tian 2015) provided sufficient conditions for the non-parametric recoverability of the causal effects from selection bias, and introduced a relaxation of this setting so that external (unbiased) data could be leveraged. (Evans and Didelez 2015) developed an approach for discrete models, where assumptions on the cardinality of the observable variables allow the estimation of the distribution over the sampling mechanism; in turn recovering the marginal distribution. (Correa and Bareinboim 2017) introduced a backdoor-like condition that controls for both biases, while (Correa, Tian, and Bareinboim 2018a) proved completeness for a more general backdoor criterion that allows for external data.

In this paper, we study the simultaneous effect of confounding and selection biases in general non-parametric settings. In particular, our contributions are as follow:

- We prove that the algorithm introduced in (Bareinboim and Tian 2015) is complete for the task of recoverability when all data available is biased. In other words, whenever the algorithm fails to recover a causal effect, the same is provable not recoverable by any other procedure.
- We relax the setting above and allow for the use of unbiased data in the form of a joint distribution over a subset of the observed variables. We develop a new algorithm for this task and prove that the approach is strictly more powerful than the current state-of-the-art method (Correa, Tian, and Bareinboim 2018a).

For the sake of space, the proofs not provided are available in the Appendix (Correa, Tian, and Bareinboim 2018b).

Structural Models, Causal Effects, and Recoverability

The systematic analysis of confounding and selection biases requires a formal language where the characterization of the underlying data-generating model can be encoded explicitly. We use the language of Structural Causal Models (SCMs) (Pearl 2000, pp. 204-207). Formally, a SCM M is a 4-tuple $\langle \mathbf{U}, \mathbf{V}, F, P(\mathbf{u})\rangle$, where $\mathbf{U}$ is a set of exogenous (latent) variables and $\mathbf{V}$ is a set of endogenous (measured) variables. F represents a collection of functions $F = \{f_i\}$ such that each endogenous variable $V_i \in \mathbf{V}$ is determined by a function $f_i \in F$, where f_i is a mapping from the respective domain of $U_i \cup Pa_i$ to V_i, $U_i \subseteq \mathbf{U}$, $Pa_i \subseteq \mathbf{V}\setminus V_i$, and the entire set F forms a mapping from $\mathbf{U}$ to $\mathbf{V}$. The uncertainty is encoded through a probability distribution over the exogenous variables, $P(\mathbf{u})$. Within the structural semantics, performing an action $\mathbf{X}{=}\mathbf{x}$ is represented through the do-operator, $do(\mathbf{X}{=}\mathbf{x})$, which encodes the operation of replacing the original equation of $\mathbf{X}$ by the constant $\mathbf{x}$ and induces a submodel $M_{\mathbf{x}}$. For a detailed discussion of SCMs, causal inference and fusion, we refer readers to (Pearl 2000; Bareinboim and Pearl 2016).

Following the conventions in the field, we denote variables by capital letters and their realized values by small letters. Sets of variables are denoted in bold. We use typical graph-theoretic terminology with the abbreviations $Pa(\mathbf{C})$, $Ch(\mathbf{C})$, $De(\mathbf{C})$, $An(\mathbf{C})$, which stand for the union of $\mathbf{C}$ and respectively the parents, children, descendants, and ancestors of $\mathbf{C}$. The letter $\mathcal{G}$ is used to refer to the causal graph, in which the unobserved common causes are encoded implicitly through the dashed bidirected arrows; $\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ denote the graph resulting from removing all incoming edges to $\mathbf{X}$ and all outgoing edges from $\mathbf{Z}$ in $\mathcal{G}$. For $\mathbf{C} \subseteq \mathbf{V}$, let $\mathcal{G}_{\mathbf{C}}$ be the subgraph of $\mathcal{G}$ composed only of variables in $\mathbf{C}$. Next, we formalize the notion of identifiability.

Definition 1 (Effect Identifiability (Pearl 2000, pp.77))**.** *The causal effect of an action $do(\mathbf{X}{=}\mathbf{x})$ on a set of variables $\mathbf{Y}$ is said to be identifiable from P in $\mathcal{G}$ if $P(\mathbf{y}|do(\mathbf{x}))$ (for short, $P_{\mathbf{x}}(\mathbf{y})$) is uniquely computable from $P(\mathbf{v})$ in any model that induces $\mathcal{G}$. Formally, for every two models M_1 and M_2 compatible with $\mathcal{G}$, $P^{M_1}(\mathbf{v}){=}P^{M_2}(\mathbf{v}){>}0$ implies $P^{M_1}(\mathbf{y}|do(\mathbf{x})){=}P^{M_2}(\mathbf{y}|do(\mathbf{x}))$.*

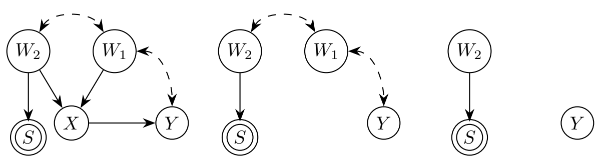
The systematic identification of causal effects calls for the ability to decompose them into easier-to-characterize quantities. For any set $\mathbf{C} \subseteq \mathbf{V}$, we then define $Q[\mathbf{C}](\mathbf{v})$, called *c-factor*, to denote the following function

$$Q[\mathbf{C}](\mathbf{v}){=}P_{\mathbf{v}\setminus\mathbf{c}}(\mathbf{c}){=}\sum_{\mathbf{U}}\prod_{\{i|V_i\in\mathbf{C}\}}P(v_i|pa_i,u_i)P(\mathbf{u}), \quad (1)$$

where pa_i is the set of observable parents of V_i and u_i is the set of unobserved parents. Of special interest are the c-factors associated with the elements of a partition on the observable variables induced by the presence of bidirected arrows, called C-Components (Tian and Pearl 2002a). The set $\mathbf{V}$ is partitioned into c-components by assigning two variables to the same set if and only if they are connected by a path composed entirely of bidirected edges in $\mathcal{G}$.

While identification deals with the problem of controlling for confounding bias, an orthogonal problem arises when the observations are not a random sample from the population. This problem is what we referred to as *selection bias* (also called sampling selection bias).

Definition 2 (Effect Recoverability (Bareinboim and Tian 2015))**.** *Given a causal graph $\mathcal{G}$ augmented with the selection mechanism, represented by the S node, the causal effect $P(\mathbf{y}|do(\mathbf{x}))$ is said to be recoverable from selection biased data if the assumptions embedded in $\mathcal{G}$ render the effect expressible in terms of the distribution under selection, $P(\mathbf{v}|S{=}1)$. That is, for any models M_1 and M_2 compatible with $\mathcal{G}$, $P^{M_1}(\mathbf{v}|S{=}1){=}P^{M_2}(\mathbf{v}|S{=}1){>}0$ implies $P^{M_1}(\mathbf{y}|do(\mathbf{x})){=}P^{M_2}(\mathbf{y}|do(\mathbf{x}))$.*



(a) $\mathcal{G}$ (b) $\mathcal{G}_{\{W_2,W_1,Y,S\}}$ (c) $\mathcal{G}_{\{W_2,Y,S\}}$

Figure 1: Subgraphs considered by RC while recovering $P_x(y)$ for the model in (a).

Roughly speaking, the paths between an action $\mathbf{X}$ and an outcome $\mathbf{Y}$ in a causal graph can be partitioned into causal (directed paths) and non-causal (spurious). A path is called *proper* if it contains no variables in $\mathbf{X}$ except at the starting point. The following construction graphically "disables" *proper* causal paths, by cutting the first arrow in such paths, leaving the spurious paths unperturbed.

Definition 3 (Proper Backdoor Graph (van der Zander, Liskiewicz, and Textor 2014))**.** *Let $\mathcal{G}$ be a causal diagram, and $\mathbf{X}, \mathbf{Y}$ be disjoint subsets of variables. The proper backdoor graph, denoted as $\mathcal{G}^{pbd}_{\mathbf{XY}}$, is obtained from $\mathcal{G}$ by removing the first edge of every proper causal path from $\mathbf{X}$ to $\mathbf{Y}$.*

This transformation will allow us to characterize the failing condition for recoverability in the next section.

Recoverability from Biased Data

In this section, we consider the problem of recovering the causal distribution when only biased data is available, namely, evaluating whether $P_{\mathbf{x}}(\mathbf{y})$ is computable from $P(\mathbf{v}|S{=}1)$. First, we consider the state-of-the-art sufficient procedure available in the literature, and then study the conditions under which it fails.

In order to recover a causal effect of the form $P_{\mathbf{x}}(\mathbf{y})$, it is usually wise to express it as a product of c-factors associated with the c-components as follows:

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{V}\setminus\mathbf{Y}} P_{\mathbf{x}}(\mathbf{v}\setminus\mathbf{x}) = \sum_{\mathbf{V}\setminus\mathbf{Y}} Q[\mathbf{V}\setminus\mathbf{X}]$$
$$= \sum_{\mathbf{D}\setminus\mathbf{Y}} Q[\mathbf{D}] = \sum_{\mathbf{D}\setminus\mathbf{Y}} \prod_{i=1}^{l} Q[D_i], \quad (2)$$

where $\mathbf{D}{=}An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}$, and $D_1, \ldots, D_l$ are the c-components of $\mathcal{G}_{\mathbf{D}}$.

This factorization was employed as the basis for the algorithm RC (Bareinboim and Tian 2015), shown in Alg. 1. RC attempts to recover each $Q[D_i]$, by Lemma 3 in (Bareinboim and Tian 2015), every $Q[C_i]$ in line 2 is recoverable, and the function IDENTIFY$(\mathbf{E}, C_i, Q[C_i])$ (Tian 2002) (line 4) can be used to determine the identifiability of $Q[\mathbf{E}]$ from $Q[C_i]$, where $\mathbf{E} \subseteq C_i$. If all such factors are successfully recovered, then the effect $P_{\mathbf{x}}(\mathbf{y})$ is recoverable as (2).

To understand the mechanics of the algorithm, we consider the model in Fig. 1(a) and assume our target distribution is $P_x(y)$. In this graph $\mathbf{D}{=}\{Y\}$ hence $P_x(y){=}Q[Y]$, consequently RC$(\{Y\}, P(\mathbf{v}|S{=}1), \mathcal{G})$ will be invoked. Since all variables in $\mathcal{G}$ are ancestors of Y or S, line 1's condition does not apply, and RC

Algorithm 1 Procedure in (Bareinboim and Tian 2015) for recovering $Q[\mathbf{E}]$

function RC$(\mathbf{E}, P, \mathcal{G})$

Input $\mathbf{E}$ a c-component, P a distribution and $\mathcal{G}$ a causal diagram over variables $\mathbf{V}$ and S.
Output Expression for $Q[\mathbf{E}]$ in terms of $P(\mathbf{v}|S{=}1)$ or FAIL

1: If $\mathbf{V} \setminus (An(\mathbf{E}) \cup An(S)) \neq \emptyset$,
return RC$\left(\mathbf{E}, \sum_{\mathbf{V}\setminus(An(\mathbf{E})\cup An(S))} P, \mathcal{G}_{An(\mathbf{E})\cup An(S)}\right)$
2: Let $C_1, \ldots, C_k$ be the c-components of $\mathcal{G}$ that contains no ancestors of S, and let $\mathbf{C} = \bigcup_i C_i$
3: If $\mathbf{C} = \emptyset$, return FAIL
4: If $\mathbf{E}$ is a subset of some C_i,
return IDENTIFY$(\mathbf{E}, C_i, Q[C_i])$
5: Return RC$\left(\mathbf{E}, \frac{P}{\prod_i Q[C_i]}, \mathcal{G}_{\mathbf{V}\setminus\mathbf{C}}\right)$

---

iterates over each c-component of $\mathcal{G}$, adding those with no ancestor of S to the set $\mathbf{C}$. In this example, $\mathcal{G}$ decomposes into three c-components: $\{X\}, \{W_1, W_2, Y\}$, and $\{S\}$, where only $\{X\}$ satisfies the condition to get into $\mathbf{C}$ and $Q[X]$ is recovered as $P(x|w_1, w_2, S{=}1)$. Since $\{Y\}$ is not a subset of $\{X\}$, line 5 recursively calls RC$(\{Y\}, P(\mathbf{v}|S{=}1)/P(x|w_1, w_2, S{=}1), \mathcal{G}_{\{W_2,W_1,Y,S\}})$. This new graph is shown in Fig.1(b). Now that X is not in the graph, the variable W_1 is no longer an ancestor of either Y or S, then line 1 performs a recursive call as RC$(\{Y\}, \sum_{W_1} P(\mathbf{v}|S{=}1)/P(x|w_1,w_2,S{=}1), \mathcal{G}_{\{W_2,Y,S\}})$. In the graph $\mathcal{G}_{\{W_2,Y,S\}}$, shown in Fig. 1(c), there are three c-components: $\{W_2\}, \{Y\}$, and $\{S\}$. Since Y is not an ancestor of S in this graph, line 2 will recover $Q[C_1]$ where $C_1{=}\{Y\}$ as $\sum_{W_1} P(y|x, w_1, w_2|S{=}1)P(w_1, w_2|S{=}1)/P(w_2|S{=}1)$ and make $\mathbf{C}{=}\{Y\}$. Next, because our target $\{Y\}$ is a subset of $C_1{=}\{Y\}$, line 4 recovers $Q[Y] = Q[C_1]$ and returns it, which, as noted before, corresponds to $P_x(y)$.

While RC was shown to be sound, it was not shown to be complete, that is, whether a FAIL triggered by line 3 implies that the target causal effect is not recoverable, or if the algorithm is not powerful enough to recover the expression.

In the following, we first present a necessary condition for the causal effect to be recoverable and then use it to show the completeness of the procedure RC.

Theorem 1. *Let $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$ be two disjoint sets of variables and $\mathcal{G}$ a causal diagram over $\mathbf{V}$ and S. If $(\mathbf{Y} \not\perp\!\!\!\perp S)_{\mathcal{G}^{pbd}_{\mathbf{XY}}}$, then $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable from $P(\mathbf{v} \mid S{=}1)$ in $\mathcal{G}$.*

The necessary condition in Thm. 1 helps us to show that when RC fails, $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable.

Theorem 2. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two disjoint sets of variables and $\mathcal{G}$ a causal diagram over $\mathbf{V}$ and S. Let $\mathbf{D}{=}An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}$ and $D_1, \ldots, D_\ell$ be the c-components of $\mathcal{G}_{\mathbf{D}}$. Then, the effect $P_{\mathbf{x}}(\mathbf{y})$ is recoverable from $P(\mathbf{v}|S{=}1)$ if and only if each $D_i, i = 1, \ldots, \ell$ is recoverable by the function* RC.

Thm. 2 implies that the strategy employed by RC covers all recoverability scenarios, and all other algorithms concerned with this setting will be in some form or shape, at most, equivalent to it. In other words, the recoverability algorithm in (Bareinboim and Tian 2015) is complete.

Recoverability with External Data

Whenever the conditions of Thm. 2 are not satisfied, the target effect is provably not inferable from $P(\mathbf{v}|S{=}1)$. One common strategy to circumvent this challenging situation is to try to find and leverage alternative sources of data. Popular baseline covariates such as age, sex, and ethnicity can be obtained without bias in many cases, for instance, using data from the census or smaller pilot studies.

We supplement Def. 2 to formally account for the availability of a new source of data, i.e.,

Definition 4 (Recoverability from Selection Bias with External Data)**.** *Given a causal graph $\mathcal{G}$ augmented with the selection mechanism, represented by the S node, the causal effect $P_{\mathbf{x}}(\mathbf{y})$ is said to be recoverable from selection bias with external data $P(\mathbf{t})$ if for any two models M_1 and M_2 compatible with $\mathcal{G}$, $P^{M_1}(\mathbf{v}|S{=}1) = P^{M_2}(\mathbf{v}|S{=}1) > 0$ and $P^{M_1}(\mathbf{t}) = P^{M_2}(\mathbf{t}) > 0$ implies $P^{M_1}_{\mathbf{x}}(\mathbf{y}) = P^{M_2}_{\mathbf{x}}(\mathbf{y})$.*

In other words, Def. 4 requires the causal effect to be uniquely computable from the available data (under selection bias and from the external source) and the assumptions embodied in the causal model.

We consider unbiased external data in the form of a distribution $P(\mathbf{t^0})$, where $\mathbf{T^0} \subset \mathbf{V}$ is a set of variables measured (jointly) without bias. As shown in the next lemma, additional information can be inferred from the external data and model assumptions.

Lemma 1. *Given $P(\mathbf{t^0})$, let $\mathbf{T}'$ be a set of variables such that $(S \perp\!\!\!\perp \mathbf{T}' \mid \mathbf{T^0})$, and let $\mathbf{T} = \mathbf{T^0} \cup \mathbf{T}'$, then $P(\mathbf{t})$ is recoverable.*

Proof. $P(\mathbf{t}) = P(\mathbf{t}' \mid \mathbf{t^0}, S{=}1)P(\mathbf{t^0})$. □

From this point on, we will use $\mathbf{T}$ to denote a set of variables such that $P(\mathbf{t})$ is available (following from $P(\mathbf{t^0})$), and let $\mathbf{R} = \mathbf{V} \setminus \mathbf{T}$ be the rest of the variables. Let $PJ(\mathcal{G}, \mathbf{T})$ denote the graph derived from the original graph $\mathcal{G}$ by representing the variables in $\mathbf{R}$ as unobservables (with bidirected edges), known as the *projection* of $\mathcal{G}$ on the set $\mathbf{T}$ (Verma 1993) (see also Def. 1 in (Tian and Pearl 2002b)). Accordingly, we can define c-factors $Q_R[.]$ in this projection, denoting the following function

$$Q_R[\mathbf{C}] = P_{\mathbf{t}\setminus\mathbf{c}}(\mathbf{c}) = P_{\mathbf{v}\setminus(\mathbf{c}\cup\mathbf{r})}(\mathbf{c}) = \sum_{\mathbf{U},\mathbf{R}} \prod_{\{i|V_i\in\mathbf{C}\cup\mathbf{R}\}} P(v_i \mid pa_i, u_i)P(\mathbf{u}). \quad (3)$$

In other words, the function $Q_R[.]$ represents a c-factor in $\mathcal{G}$ when the variables in $\mathbf{R}$ are treated as latent variables [1].

The next result delineates the new c-factors that can be recovered from $P(\mathbf{t})$:

Lemma 2. *Let $\mathbf{T} \subseteq \mathbf{V}$, $\mathbf{R} = \mathbf{V} \setminus \mathbf{T}$, and $T_1, \ldots, T_m$ be the c-components of $PJ(\mathcal{G}, \mathbf{T})$, then all $Q_R[T_k]$ are recoverable from $P(\mathbf{t})$.*

Proof. We have that

$$P(\mathbf{t}) = \sum_{\mathbf{R}} P(\mathbf{v}) = \prod_{k=1}^{m} Q_R[T_k]. \quad (4)$$

By (Tian and Pearl 2002b, Lemma 2), all $Q_R[T_k]$ are recoverable from $P(\mathbf{t})$. □

Building on Lemmas 1 and 2, we now state the main result of this section:

Theorem 3. *Let $\mathbf{H} \subseteq \mathbf{V} \cup \{S\}$, such that $\mathbf{H}$ is partitioned into c-components $H_1, \ldots, H_l, H_s$ in the subgraph $\mathcal{G}_{\mathbf{H}}$, where $S \in H_s$. Assume*

$$f(P(\mathbf{v} \mid S{=}1)) = \frac{Q[H_s](\mathbf{v}, S{=}1)}{P(S{=}1)} \prod_i Q[H_i], \quad (5)$$

where $f(P(\mathbf{v}|S{=}1))$ is some recoverable quantity, and $P(\mathbf{t})$ is available. Let $\mathbf{T}^0_{\mathbf{H}} = \mathbf{T} \setminus De(\mathbf{V}\setminus\mathbf{H})_{\mathcal{G}}$ and $\mathbf{T}'$ be the set of all variables in $\mathbf{H}$ such that $(\mathbf{T}' \perp\!\!\!\perp S \mid \mathbf{T}^0_{\mathbf{H}})_{\mathcal{G}_{\mathbf{H}}}$. Also, let $\mathbf{T_H} = \mathbf{T}^0_{\mathbf{H}} \cup \mathbf{T}'$ and let $\mathbf{R_H} = \mathbf{H} \setminus \mathbf{T_H}$. Then, for $j{=}1, \ldots, l$, $Q[H_j]$ is recoverable if H_j contains no variables that are both ancestors of H_s and belong to $\mathbf{R_H}$ or its children (i.e. $H_j \cap An(H_s) \cap Ch(\mathbf{R_H}) = \emptyset$) in $\mathcal{G}_{\mathbf{H}}$.

Proof. (sketch, see Appendix C for details) Let a topological order of the variables in $\mathbf{H}$ be $V_{h_1} < \cdots < V_{h_k}$ in $\mathcal{G}_{\mathbf{H}}$. Let $H^{\leq i} = \{V_{h_1}, \ldots, V_{h_i}\}$ be the set of variables in $\mathbf{H}$ ordered before V_{h_i} (including V_{h_i}), and $H^{>i} = \mathbf{H} \setminus H^{\leq i}$ for $i = 1, \ldots, k$, and $H^{\leq 0} = \emptyset$. The assumptions of the theorem allow us to recover $P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})$ from $f(P(\mathbf{v}|S{=}1))$ and $P(\mathbf{t})$. For any H_j that satisfies the condition of the theorem, the associated c-factor can be recovered as:

$$Q[H_j] = \prod_{\{i|V_{h_i}\in H_j\cap An(H_s)\}} \frac{\sum_{h^{>i}\cap\mathbf{T_H}} P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})}{\sum_{h^{>i-1}\cap\mathbf{T_H}} P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})} \times \prod_{\{i|V_{h_i}\in H_j\setminus An(H_s)\}} \frac{\sum_{h^{>i}} f(P(\mathbf{v}|S{=}1))}{\sum_{h^{>i},V_{h_i}} f(P(\mathbf{v}|S{=}1))}. \quad (6)$$

□

Thm. 3 will be the main driving force for recovering causal effects from combined biased data $P(\mathbf{v}|S{=}1)$ and unbiased data $P(\mathbf{t})$. To give an example of how this result can be used, consider the model in Fig 2(a) and assume we have external data over $\mathbf{T^0}{=}\{Z\}$. Then, $\mathbf{T}{=}\{Z, X, Y\}$ because $(S \perp\!\!\!\perp X, Y|Z)$, $\mathbf{R}{=}\mathbf{V}\setminus\mathbf{T}{=}\{R, W\}$, and $H_s{=}\{S, Z\}$ which is the c-component that contains S. Also, the biased distribution factorizes as follows:

$$P(\mathbf{v} \mid S{=}1) = \frac{Q[S, Z]}{P(S{=}1)} Q[W]Q[R]Q[X]Q[Y]. \quad (7)$$

Thm. 3 would allow us to recover $Q[X]$ and $Q[Y]$ since they do not contain any ancestor of H_S.

Recovering Causal Effects Systematically

In order to recover the causal distribution $P_{\mathbf{x}}(\mathbf{y})$ systematically, (Bareinboim and Tian 2015) proposed a strategy that recovers each $Q[D_i]$ in Eq. (2) one by one. It turns out that when external data $P(\mathbf{t})$ is available, each $Q[D_i]$ being recoverable is no longer necessary for the overall recoverability of $P_{\mathbf{x}}(\mathbf{y})$. To witness, let us follow up on the example

[1] C-components with arbitrary variables as latent variables are defined in (Tian and Pearl 2002b).

from Fig 2(a), introduced at the end of the last section. Following the strategy dictated by Eq. (2), we note that

$$P_x(y) = \sum_{Z,R} Q[Y, Z, R] = \sum_{Z,R} Q[Y]Q[Z]Q[R]. \tag{8}$$

Thm. 3 licenses the recoverability of $Q[Y]$, but it is not difficult to shown that neither $Q[R]$ nor $Q[Z]$ is recoverable. Perhaps surprisingly, however, $P_x(y)$ can be recovered as

$$\sum_Z Q[Y] \sum_R Q[R]Q[Z] = \sum_Z P(y|x, z, S{=}1)P(z). \tag{9}$$

The key observation here is that while $Q[R]$ and $Q[Z]$ are not recoverable individually, $\sum_R Q[R]Q[Z]$ is, in fact, a function of Z and equal to $Q_R[Z]$ (see Eq. 3), which can be recovered from $P(\mathbf{t}) = P(z, x, y)$ as $P(z)$ via Lemma 2.

To formally account for this situation, we re-write the causal effect in Eq. (2) by splitting $\mathbf{D} \setminus \mathbf{Y}$ into two parts: $\mathbf{A} = (\mathbf{D}\setminus\mathbf{Y})\cap\mathbf{T}$ and $\mathbf{B} = (\mathbf{D}\setminus\mathbf{Y})\cap\mathbf{R}$ where $\mathbf{R} = \mathbf{V}\setminus\mathbf{T}$, and then we treat elements in $\mathbf{B}$ as latent variables while defining c-factors $Q_B[.]$ in the resulting projected graph $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B})$, as follows:

$$P_\mathbf{x}(\mathbf{y}) = \sum_{\mathbf{D}\setminus\mathbf{Y}} \prod_{i=1}^{l} Q[D_i] = \sum_{\mathbf{A}} \prod_{j=1}^{\ell} Q_B[C_j], \tag{10}$$

where $\mathbf{D}{=}An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}$, $D_1, \ldots, D_l$ are the c-components of $\mathcal{G}_\mathbf{D}$, $C_1, \ldots, C_\ell$ are the c-components of $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D}\setminus\mathbf{B})$, and c-factors $Q_B[C_j]$ are defined as

$$Q_B[C_j] = \sum_{\mathbf{U}\cup\mathbf{B}} \prod_{\{i|V_i\in C_j\cup\mathbf{B}\}} P(v_i|pa_i, u_i)P(\mathbf{u}). \tag{11}$$

$Q_B[C_j]$ could also be expressed in terms of $Q[D_i]$ in the following form:

$$Q_B[C_j] = \sum_{B_j} \prod_{\{i|D_i\in F_j\}} Q[D_i], \tag{12}$$

where B_j are disjoint and possibly empty sets such that $\cup_{j=1}^{\ell} B_j{=}\mathbf{B}$, and $F_1, \ldots, F_\ell$ form a partition of $\{D_1, \ldots, D_l\}$.

Under certain conditions, a c-factor $Q_B[C_j]$ may be equal to the c-factor $Q_R[C_j]$, defined in $PJ(\mathcal{G}, \mathbf{T})$, which is potentially recoverable in terms of the unbiased distribution $P(\mathbf{t})$.

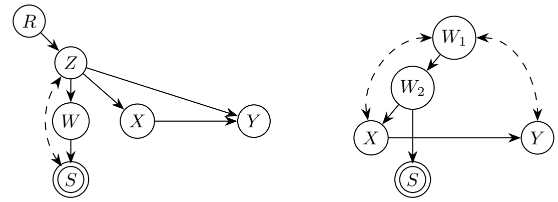
Lemma 3. *Let C_j be a c-component of $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B})$. If $\mathbf{B} \cap Pa(C_j) = \mathbf{R} \cap Pa(C_j)$, then*

(i) $Q_B[C_j] = Q_R[C_j]$, where $Q_R[C_j]$ is a c-factor in $PJ(\mathcal{G}, \mathbf{T})$ as defined in Eq. (3)*; and*

(ii) Let $T_1, \ldots, T_m$ be the c-components of $PJ(\mathcal{G}, \mathbf{T})$, then C_j must be a subset of some T_k.

Proof. *(i)* Let $\hat{B}_j = \mathbf{B} \cap Pa(C_j)$. Any variable in $\mathbf{B}$ that is not in $\hat{B}_j$ does not appear in the expression in (11), and can be summed out, leading to

$$Q_B[C_j] = \sum_{\mathbf{U}\cup\hat{B}_j} \prod_{\{i|V_i\in C_j\cup\hat{B}_j\}} P(v_i|pa_i, u_i)P(\mathbf{u}). \tag{13}$$



(a) $\mathbf{T^0} = \{Z\}$ (b) $\mathbf{T^0} = \{W_2\}$

Figure 2: Examples of recoverability tasks for the effect $P_x(y)$. Model in (a) can be recovered with external data on Z. Model in (b) is recoverable with external data on W_2 or W_1.

Similarly, from (3) we have:

$$Q_R[C_j] = \sum_{\mathbf{U}\cup\mathbf{R}} \prod_{\{i|V_i\in\mathbf{C}_j\cup\mathbf{R}\}} P(v_i|pa_i, u_i)P(\mathbf{u}).$$

Let $\hat{R}_j = \mathbf{R} \cap Pa(C_j)$. Then any variable in $\mathbf{R}$ that is not in $\hat{R}_j$ can be summed out, leading to

$$Q_R[C_j] = \sum_{\mathbf{U}\cup\hat{R}_j} \prod_{\{i|V_i\in\mathbf{C}_j\cup\hat{R}_j\}} P(v_i|pa_i, u_i)P(\mathbf{u}). \tag{14}$$

It is clear that if $\hat{R}_j = \hat{B}_j$, then (14) is equal to (13).

(ii) Since $\mathbf{D} \subseteq \mathbf{V}$ and $\mathbf{B} \subseteq \mathbf{R}$, a c-component of $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B})$ must be a subset of a c-component of $PJ(\mathcal{G}, \mathbf{T})$. □

The importance of Lemma 3 stems from the fact that $Q_R[C_j]$ is potentially identifiable in $PJ(\mathcal{G}, \mathbf{T})$ from the unbiased distribution $P(\mathbf{t})$ based on Lemma 2. Specifically, we can use IDENTIFY$(C_j, T_k, Q_R[T_k])$ to try to recover $Q_B[C_j] = Q_R[C_j]$. If $Q_R[C_j]$ is not identifiable from $P(\mathbf{t})$, then we further attempt to recover $Q_B[C_j]$ by recovering each $Q[D_i]$ in Eq. (12) factor by factor.

To recover an individual $Q[D_i]$, it turns out the RC algorithm (Alg. 1) is not complete anymore in our setting (even if line 2 of RC is enhanced with Thm. 3). Extending RC, we develop a new algorithm called RCE (Alg. 2) to recover any target c-component $Q[\mathbf{E}]$. RCE attempts to systematically recover $Q[\mathbf{E}]$ by recovering, using Thm. 3 (line 2), the c-component $Q[C_i]$ of $\mathcal{G}$ that contains $\mathbf{E}$, and then call the function IDENTIFY to recover $Q[\mathbf{E}]$ from $Q[C_i]$ (line 3a). To facilitate this, RCE reduces the problem to simpler subgraphs, by removing irrelevant non-ancestors (line 1) or other recoverable c-components (line 3b and line 4) from the current graph. These other c-components are recovered either by Thm. 3 (line 2) or by recursively calling RCE (line 4). Due to the recursive nature of the process, RCE may try to compute a c-component more than once, which can be avoided by keeping track of the previous queries. For simplicity we omit these practical details.

Putting these results together, we develop a general, systematic procedure for recovering causal effects called IDSB. The function IDSB in Alg. 3 accepts as input two disjoint sets $\mathbf{X}, \mathbf{Y}$, distributions $P(\mathbf{v}|S{=}1), P(\mathbf{t^0})$, and a causal diagram $\mathcal{G}$; it outputs an expression for $P_\mathbf{x}(\mathbf{y})$ in terms of the input distributions or FAIL. IDSB starts by simplifying the

Algorithm 2 Recursive function used to recover an arbitrary c-component

function RCE$(\mathbf{E}, \mathcal{P}, \mathcal{G})$

Input $\mathbf{E}$ a set of variables such that $\mathbf{E}$ is a c-component, $\mathcal{P}$ a distribution over $\mathbf{V}$, $\mathcal{G}$ a causal diagram over variables $\mathbf{V}$ and S. $P^*(\mathbf{t})$ a distribution over $\mathbf{T}$ and $\mathcal{G}^*$ the original graph over variables $\mathbf{V}^*$ and S are defined globally.
Output Expression for $Q[\mathbf{E}]$ or FAIL

1: Let $\mathbf{W} = An(\mathbf{E}) \cup An(S)$. If $\mathbf{V}\backslash\mathbf{W} \neq \emptyset$, return RCE$\left(\mathbf{E}, \sum_{\mathbf{V}\backslash\mathbf{W}} \mathcal{P}, \mathcal{G}_{\mathbf{W}}\right)$
2: Let $C_1, \ldots, C_k$ be the c-components of $\mathcal{G}$ that are recoverable by Thm. 3 (with $f(P(\mathbf{v}|S{=}1)) = \mathcal{P}$ and $P(\mathbf{t}) = P^*(\mathbf{t})$). Let $\mathbf{C} = \bigcup_i C_i$
3: If $\mathbf{C} \neq \emptyset$,
  (a) If $\mathbf{E}$ is a subset of some C_i, then return IDENTIFY$(\mathbf{E}, C_i, Q[C_i])$
  (b) Return RCE$\left(\mathbf{E}, \frac{\mathcal{P}}{\prod_i Q[C_i]}, \mathcal{G}_{(\mathbf{V}\cup\{S\})\backslash\mathbf{C}}\right)$
4: For each c-component B_i of $\mathcal{G}$ that does not contain $\mathbf{E}$ such that $\mathbf{Z} = \mathbf{V} \setminus (An(S) \cup An(B_i)) \neq \emptyset$:
   $Q[B_i] = \text{RCE}(B_i, \sum_{\mathbf{Z}} \mathcal{P}, \mathcal{G}_{(\mathbf{V}\cup\{S\})\backslash\mathbf{Z}})$
   If $Q[B_i] \neq$ FAIL, return RCE$\left(\mathbf{E}, \frac{\mathcal{P}}{Q[B_i]}, \mathcal{G}_{(\mathbf{V}\cup\{S\})\backslash B_i}\right)$
5: Return FAIL

Algorithm 3 Algorithm capable of recovering $P_{\mathbf{x}}(\mathbf{y})$ from selection bias with external data

function IDSB$\left(\mathbf{X}, \mathbf{Y}, P, P(\mathbf{t^0}), \mathcal{G}\right)$

Input $\mathbf{X}, \mathbf{Y}$ disjoint sets of variables, $P(\mathbf{v}|S{=}1)$ a distribution, $P(\mathbf{t^0})$ distribution over a set of variables $\mathbf{T^0}$, and $\mathcal{G}$ a causal diagram over variables $\mathbf{V}$ and S
Output Expression for $P_{\mathbf{x}}(\mathbf{y})$ in terms of $P(\mathbf{v}|S{=}1)$ and $P(\mathbf{t^0})$ or FAIL

1: Let $\mathbf{W} = An(\mathbf{Y}) \cup An(S)$, $\mathcal{G} \leftarrow \mathcal{G}_{\mathbf{W}}$, $P \leftarrow \sum_{\mathbf{V}\backslash\mathbf{W}} P$
2: Let $\mathbf{T}' \subset \mathbf{W}$ be the set of all the variables such that $(S \perp\!\!\!\perp \mathbf{T}'|\mathbf{T^0}\cap\mathbf{W})_{\mathcal{G}}$, and $\mathbf{T}{=}\mathbf{T}'\cup(\mathbf{T^0}\cap\mathbf{W})$
3: Recover $P(\mathbf{t})$ by Lemma 1
4: Let $\mathbf{D} = An(\mathbf{Y})_{\mathcal{G}_{\mathbf{W}\backslash\mathbf{X}}}$,
   Let $D_1, \ldots, D_l$ be the c-components of $\mathcal{G}_{\mathbf{D}}$,
   Let $T_1, \ldots, T_m$ be the c-components of $PJ(\mathcal{G}, \mathbf{T})$,
   $\mathbf{R} = \mathbf{W} \setminus \mathbf{T}$, $\mathbf{A} = (\mathbf{D} \setminus \mathbf{Y}) \cap \mathbf{T}$, $\mathbf{B} = (\mathbf{D} \setminus \mathbf{Y}) \cap \mathbf{R}$,
   Let $C_1, \ldots, C_\ell$ be the c-components of $PJ(\mathcal{G}_{\mathbf{D}}, \mathbf{D} \setminus \mathbf{B})$, such that $Q_B[C_j]$ is given by Eq. (12).
5: For each C_j
   If $\mathbf{B} \cap Pa(C_j) = \mathbf{R} \cap Pa(C_j)$ then
     Assume C_j is a subset of T_k
     $Q_B[C_j] = \text{IDENTIFY}(C_j, T_k, Q_R[T_k])$
   If $\mathbf{B} \cap Pa(C_j) \neq \mathbf{R} \cap Pa(C_j)$ or $Q_B[C_j] =$ FAIL, then
     $Q_B[C_j] = \sum_{B_j} \prod_{i, D_i \in F_j} \text{RCE}\,(D_i, P, \mathcal{G})$
     If $Q_B[C_j] =$ FAIL, then return FAIL
6: Return $\sum_{\mathbf{A}} \prod_{j=1}^{\ell} Q_B[C_j]$

model via removing irrelevant non-ancestors (line 1) and recovering $P(\mathbf{t})$ using Lemma 1 (lines 2, 3). IDSB then recovers $P_{\mathbf{x}}(\mathbf{y})$ using Eq. (10) by recovering each $Q_B[C_j]$ (line 5). For each $Q_B[C_j]$, IDSB first attempts to recover it from $P(\mathbf{t})$ based on Lemma 2 by calling the function IDENTIFY if the condition in Lemma 3 is satisfied. If this fails, IDSB tries to recover $Q_B[C_j]$ using (12) by calling RCE for each $Q[D_i]$. The next theorem states that IDSB is sound.

Theorem 4. *The procedure* IDSB *is sound.*

Due to page limits, we provide the proof of Thm. 4 in the Appendix D. Nevertheless, we will illustrate its mechanics using the example from Fig. 3(a) where we assume $P(\mathbf{v}|S{=}1)$ and $P(\mathbf{t^0})$ are given, with $\mathbf{T^0}{=}\{V_2, V_3, V_6\}$, and the goal is to recover $P_x(y)$. Initially, in line 1 $\mathbf{W} = \mathbf{V}$. Line 2 finds that $(S \perp\!\!\!\perp X, V_1|\mathbf{T^0})$, hence $\mathbf{T}{=}\{X, V_1, V_2, V_3, V_6\}$ and line 3 recovers $P(\mathbf{t}){=}P(x, v_1|v_2, v_3, v_6, S{=}1)P(v_2, v_3, v_6)$. At line 4, we have $\mathbf{D}{=}\{V_5, V_6, Y\}$, $\mathbf{R}{=}\{V_4, V_5, Y\}$, $\mathbf{A}{=}\{V_6\}$, and $\mathbf{B}{=}\{V_5\}$. The graphs $\mathcal{G}_{\mathbf{D}}$, $PJ(\mathcal{G}, \mathbf{T})$, and $PJ(\mathcal{G}_{\mathbf{D}}, \mathbf{D}\backslash\mathbf{B})$ (Fig. 3(b), (c), and (d) respectively) are derived from $\mathcal{G}$ (Fig. 3(a)). The table in Fig. 3(i) summarizes the decomposition of these graphs and recalls how each c-component and c-factor are denoted by IDSB in line 4. At this point, we know from Eq. (10) that $P_x(y){=}\sum_{V_6} Q_B[Y]Q_B[V_6]$. Also, $Q_B[Y]{=}Q[Y]$, $Q_B[V_6]{=}\sum_{V_5} Q[V_5, V_6]$, corresponding to $B_1{=}\emptyset, B_2{=}\{V_5\}, F_1{=}\{\{Y\}\}$ and $F_2{=}\{\{V_5, V_6\}\}$. Clearly $\mathbf{B}{=}B_1{\cup}B_2$ and F_1, F_2 constitute a partition over $\{D_1, D_2\}$.

Continuing with line 5, the algorithm considers the first c-component $C_1{=}\{Y\}$, and since $\mathbf{B}{\cap}Pa(Y){=}\emptyset \neq \mathbf{R}{\cap}Pa(Y){=}\{Y\}$, it calls RCE to try to recover $Q[Y]$ (which is equal to $Q_B[Y]$) in the graph $\mathcal{G}$. The recursion induced by this call to RCE is depicted in Fig. 4, where each edge is annotated with the line number (in RCE) that initiates the call and Fig. 3(e)-(h) contain the relevant subgraphs. Each $\mathcal{P}^{(i)}, i{=}0, \ldots, 4$ stands for the distributions associated with the corresponding subgraph, obtained as follows

$$\mathcal{P}^{(0)} = P(\mathbf{v}|S{=}1), \tag{15}$$

$$\mathcal{P}^{(1)} = \mathcal{P}^{(0)} \Big/ Q[V_2] \text{ , where} \tag{16}$$

$$Q[V_2] = \sum\nolimits_{X,V_3,V_6} P(\mathbf{t}) \Big/ \sum\nolimits_{X,V_3,V_6,V_2} P(\mathbf{t}) \text{ ,} \tag{17}$$

$$\mathcal{P}^{(2)} = \sum\nolimits_{V_1} \mathcal{P}^{(1)}, \tag{18}$$

$$\mathcal{P}^{(3)} = \mathcal{P}^{(2)} \Big/ Q[X] \text{ , where} \tag{19}$$

$$Q[X] = \sum\nolimits_{V_4,V_5,Y} \mathcal{P}^{(2)} \Big/ \sum\nolimits_{V_4,V_5,X,Y} \mathcal{P}^{(2)} \text{ , and} \tag{20}$$

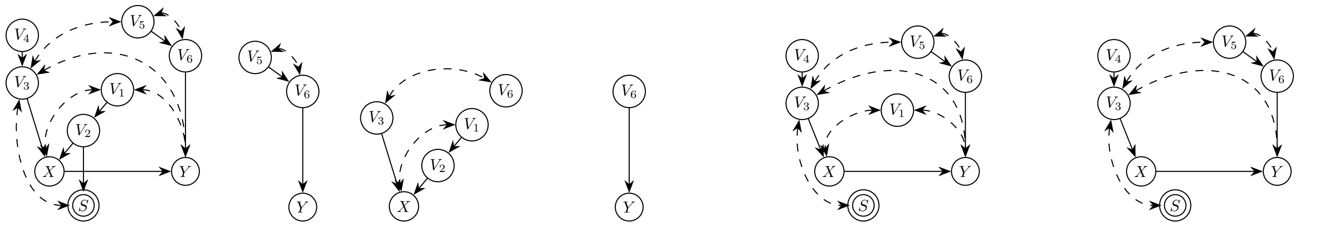
$$\mathcal{P}^{(4)} = \sum\nolimits_{V_3,V_4} \mathcal{P}^{(3)}. \tag{21}$$

Finally, the result returned by RCE is:

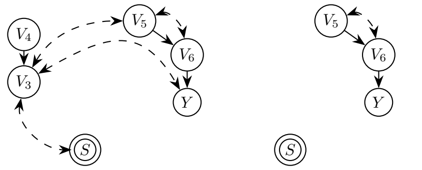
$$Q[Y] = \mathcal{P}^{(4)} \Big/ \sum\nolimits_{Y} \mathcal{P}^{(4)}. \tag{22}$$

After $Q[Y]$ is computed, IDSB moves on to $C_2{=}\{V_6\}$. Since $\mathbf{B}{\cap}Pa(V_6){=}\{V_5\}{=}\mathbf{R}{\cap}Pa(V_6)$, we have that $Q_B[V_6]$ is equal to $Q_R[V_6]$ which is potentially identifiable from $Q_R[T_2]$ where $T_2{=}\{V_3, V_6\}$. Next, IDSB calls IDENTIFY$(\{V_6\},\{V_3,V_6\},Q_R[T_2])$ to obtain $Q_B[V_6]{=}P(v_6)$.

Despite IDSB's generality, it is not clear at this point whether there are positive cases not covered by the algorithm – i.e., cases computable from $P(\mathbf{t^0})$ and $P(\mathbf{v}|S{=}1)$, but where IDSB returns "FAIL". Still, the current state-of-the-art procedure that accepts external data, called General-



(a) $\mathcal{G}$ (b) $\mathcal{G}_{\mathbf{D}}$ (c) $PJ(\mathcal{G},\mathbf{T})$ (d) $PJ(\mathcal{G}_{\mathbf{D}},\mathbf{D}\backslash\mathbf{B})$ (e) $\mathcal{G}_{\mathbf{V}\cup\{S\}\backslash\{V_2\}}$ (f) $\mathcal{G}_{\mathbf{V}\cup\{S\}\backslash\{V_2,V_1\}}$



(g) $\mathcal{G}_{\mathbf{V}\cup\{S\}\backslash\{V_2,V_1,X\}}$ (h) $\mathcal{G}_{\{Y,V_5,V_6,S\}}$

Graph	C-Components	Notation	C-factor
$\mathcal{G}_{\mathbf{D}}$	$\{Y\}$,$\{V_5,V_6\}$	D_i,i=1,2	$Q[.]$
$PJ(\mathcal{G}_{\mathbf{D}},\mathbf{D}\backslash\mathbf{B})$	$\{Y\}$,$\{V_6\}$	C_j,j=1,2	$Q_B[.]$
$PJ(\mathcal{G},\mathbf{T})$	$\{X,V_1\}$,$\{V_3,V_6\}$,$\{V_2\}$	T_k,k=1,2,3	$Q_R[.]$

(i) C-components of the graphs in (b), (c), and (d)

Figure 3: Example of a model and the transformations involved in recovering the target causal effect. We assume $P(\mathbf{v}|S{=}1)$ and $P(v_2, v_3, v_6)$ are given.

RCE($\{Y\}, \mathcal{P}^{(0)}, \mathcal{G}$)
| 3b
RCE($\{Y\}, \mathcal{P}^{(1)}, \mathcal{G}_{\mathbf{V}\cup\{S\}\backslash\{V_2\}}$)
| 1
RCE($\{Y\}, \mathcal{P}^{(2)}, \mathcal{G}_{\mathbf{V}\cup\{S\}\backslash\{V_2,V_1\}}$)
3b |
RCE($\{Y\}, \mathcal{P}^{(3)}, \mathcal{G}_{\mathbf{V}\cup\{S\}\backslash\{V_2,V_1,X\}}$)
| 1
RCE($\{Y\}, \mathcal{P}^{(4)}, \mathcal{G}_{\{Y,V_5,V_6,S\}}$)
| 3a
IDENTIFY($\{Y\}, \{Y\}, Q[Y]$)

Figure 4: Recursion of RCE when used to recover $Q[Y]$ in the model in Fig. 3(a).

ized Adjustment Criterion (GAC) (Correa, Tian, and Bareinboim 2018a), is constrained to backdoor-like expressions. The next proposition compares the power of the two approaches.

Theorem 5. IDSB *is strictly more powerful than the Generalized Adjustment Criterion for the task of recovering a causal effect* $P_{\mathbf{x}}(\mathbf{y})$ *from a combination of biased distribution* $P(\mathbf{v}|S{=}1)$ *and unbiased distribution* $P(\mathbf{t^0})$ *in* $\mathcal{G}$.

We outline how this statement can be proved (see Appendix D for the formal proof). We first show that whenever IDSB fails to recover $P_{\mathbf{x}}(\mathbf{y})$, then GAC is also unable to recover the effect. Then, to show that IDSB is strictly more general, we present an example where IDSB recovers $P_{\mathbf{x}}(\mathbf{y})$ but GAC fails. Consider the problem of recovering $P_{\mathbf{x}}(\mathbf{y})$ in the model in Fig. 2(b) with external data over $\mathbf{T^0} = \{W_2\}$. GAC asks for the following three conditions:

- Condition (iii) requires a set $\mathbf{Z^T}$ to be available from external data such that the independence $(S \perp\!\!\!\perp Y | \mathbf{Z^T})_{\mathcal{G}^{pbd}_{\mathbf{XY}}}$ holds. For this model $\mathbf{Z^T}{=}\{W_2\}$ suffices.
- Condition (i) requires that no covariate should be a descendant of a variable in a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$, which is also satisfied by $\mathbf{Z}{=}\{W_2\}$.
- However, condition (ii) requires the independence $(X \perp\!\!\!\perp Y | \mathbf{Z}, S)_{\mathcal{G}^{pbd}_{\mathbf{XY}}}$ to hold, which cannot be satisfied in this model by $\mathbf{Z}{=}\{W_2\}$, or $\mathbf{Z}{=}\{W_1,W_2\}$, or any other $\mathbf{Z}$.

Since not all conditions are satisfiable, GAC fails. Nevertheless, IDSB is able to recover $P_x(y)$. To witness, note that $\mathbf{D}{=}An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\backslash\mathbf{X}}}{=}\{Y\}$, hence $P_x(y){=}Q[Y]$. Also $\mathbf{T}{=}\{W_1,W_2,X,Y\}$, $\mathbf{R}{=}\emptyset$, The set $\{Y\}$ is a subset of c-component $T_1{=}\{W_1,X,Y\}$ in $PJ(\mathcal{G},\mathbf{T})$. IDSB will call IDENTIFY($\{Y\}, T_1, Q_R[T_1]$), where $Q_R[T_1]$ is recoverable from $P(\mathbf{t}){=}P(y,x,w_1,w_2){=}P(y,x,w_1|w_2,S{=}1)P(w_2)$ by Lemma 2, and obtain

$$P_x(y) = \frac{\sum_{W_1} P(y, x \mid w_1, w_2)P(w_1)}{\sum_{W_1} P(x \mid w_1, w_2,)P(w_1)}. \tag{23}$$

Conclusions

We investigated the challenges arising due to confounding and selection biases, which come under the rubric of recoverability of causal effects. We first studied the algorithm RC (Alg.1) (Bareinboim and Tian 2015), which takes as input a causal diagram and a biased distribution. We supplemented the algorithm with a necessary condition for recoverability (Thm. 1), and proved that RC is complete for this task, namely, it recovers all effects that are indeed recoverable (Thm. 2). We then relaxed the setting to allow the incorporation of unbiased data (Def. 4). We developed the algorithm IDSB (Alg. 3), which takes as input a combination of biased and unbiased data. We proved that IDSB is strictly more powerful than the current state-of-the-art method available (Thm. 5). Since confounding and selection biases are problems pervasive across disciplines, we hope that the methods developed here should help to understand and alleviate this problem in a broad range of data-intensive applications.

Acknowledgments Juan D. Correa and Elias Bareinboim were in parts supported by grants from IBM Research, Adobe Research, NSF IIS-1704352, and IIS-1750807 (CAREER). Jin Tian was partially supported by NSF grant IIS-1704352 and ONR grant N000141712140.

References

Angrist, J. D. 1997. Conditional independence in sample selection models. *Economics Letters* 54(2):103–112.

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. AUAI Press.

Bareinboim, E., and Pearl, J. 2012b. Controlling Selection Bias in Causal Inference. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AISTATS)*. La Palma, Canary Islands: JMLR. 100–108.

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

Bareinboim, E., and Tian, J. 2015. Recovering Causal Effects from Selection Bias. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* 3475–3481.

Bareinboim, E.; Tian, J.; and Pearl, J. 2014. Recovering from Selection Bias in Causal and Statistical Inference. In *Proceedings of the Twenty-eighth AAAI Conference on Artificial Intelligence*, 2410–2416. Palo Alto, CA: AAAI Press.

Cooper, G. 1995. Causal Discovery from Data in the Presence of Selection Bias. In *Proceedings of the Workshop on Artificial Intelligence and Statistics*, 140–150.

Correa, J. D., and Bareinboim, E. 2017. Causal effect identification by adjustment under confounding and selection biases. In *AAAI*.

Correa, J. D.; Tian, J.; and Bareinboim, E. 2018a. Generalized Adjustment Under Confounding and Selection Biases. In *Proceedings of the 32th Conference on Artificial Intelligence (AAAI 2018)*.

Correa, J. D.; Tian, J.; and Bareinboim, E. 2018b. Identification of causal effects in the presence of selection bias. Technical report, R-38, Purdue AI Lab, Dep. of Computer Science, Purdue University.

Cortes, C.; Mohri, M.; Riley, M.; and Rostamizadeh, A. 2008. Sample Selection Bias Correction Theory. In *Proceedings of the 19th International Conference on Algorithmic Learning Theory*, 38–53.

Elkan, C. 2001. The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*, volume 17, 973–978.

Evans, R. J., and Didelez, V. 2015. Recovering from Selection Bias Using Marginal Structure in Discrete Models. In *Proceedings of the UAI 2015 Conference on Advances in Causal Inference - Volume 1504*, ACI'15, 46–55.

Glymour, M. M., and Greenland, S. 2008. Causal Diagrams. In *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins, 3rd edition. 183–209.

Heckman, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica* 47(1):153.

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

Kuroki, M., and Cai, Z. 2006. On recovering a population covariance matrix in the presence of selection bias. *Biometrika* 93(3):601–611.

Little, R., and Rubin, D. 1987. *Statistical analysis with missing data*. New York, NY, USA: John Wiley & Sons.

Mefford, J., and Witte, J. S. 2012. The covariate's dilemma. *PLoS Genet* 8(11):e1003096.

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

Pearl, J. 2000. *Causality: models, reasoning, and inference*. Cambridge University Press, 2nd edition.

Pirinen, M.; Donnelly, P.; and Spencer, C. 2012. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nature Genetics* 44(8):848–851.

Robins, J. M. 2001. Data, Design, and Background Knowledge in Etiologic Inference. *Epidemiology* 12(3):313–320.

Robinson, L. D., and Jewell, N. P. 1991. Some Surprising Results about Covariate Adjustment in Logistic Regression Models. *International Statistical Review / Revue Internationale de Statistique* 59(2):227–240.

Shpitser, I., and Pearl, J. 2006. Identification of Joint Interventional Distributions in Recursive semi-Markovian Causal Models. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*, volume 21, 1219–1226.

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

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

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

Tian, J. 2002. *Studies in causal reasoning and learning*. Ph.D. Dissertation, Computer Science Department, University of California Los Angeles, CA.

van der Zander, B.; Liskiewicz, M.; and Textor, J. 2014. Constructing separators and adjustment sets in ancestral graphs. In *Proceedings of UAI 2014*, 907–916.

Verma, T. 1993. Graphical aspects of causal models. Technical report, Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles.

Whittemore, A. 1978. Collapsibility of multidimensional contingency tables. *Journal of the Royal Statistical Society. Series B* 40(3):328–340.

Zadrozny, B. 2004. Learning and evaluating classifiers under sample selection bias. In *Twenty-first international conference on Machine learning - ICML '04*, 114–121.

Appendix A. Proof of Theorem 1

The helping lemmas used in this proof are shown after the main proof for convenience. We advise the reader to consult them as needed.

Theorem 1. *Let* $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$ *be two disjoint sets of variables and* $\mathcal{G}$ *a causal diagram over* $\mathbf{V}$ *and* S. *If* $(\mathbf{Y} \not\perp\!\!\!\perp S)_{\mathcal{G}^{pbd}_{\mathbf{XY}}}$, *then* $P_{\mathbf{x}}(\mathbf{y})$ *is not recoverable from* $P(\mathbf{v} \mid S{=}1)$ *in* $\mathcal{G}$.

Proof. Suppose the stated condition is satisfied there exists an active path p between some $Y' \in \mathbf{Y}$ and S in $\mathcal{G}^{pbd}_{\mathbf{XY}}$.

Without loss of generality, let Y' be the element in $\mathbf{Y}$ connected to S with the shortest active path p. Let $\mathbf{X}'$ be the variables in $\mathbf{X}$ that lie in p. Let $\mathcal{G}'$ be the subgraph of $\mathcal{G}$ that contains the same set of variables but only the edges in p. We will show that $P_{\mathbf{x}'}(y')$ is not recoverable from $P(\mathbf{v} \mid S{=}1)$ in $\mathcal{G}'$, when this is the case it's easy to check that $P_{\mathbf{x}}(\mathbf{y}')P_{\mathbf{x}'}(\mathbf{y}')$ is also not recoverable in $\mathcal{G}'$. Then by lemma 5 this will imply in turn that $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable in $\mathcal{G}$.

Let $\mathbf{V}$ represent all variables in the graph except for the selection mechanism S, and let $Q = P_{\mathbf{x}'}(y')$. We construct two SCMs M_1 and M_2 compatible with $\mathcal{G}'$, that induce probability distributions $P^{M_1}(\mathbf{v} \mid S{=}1)$ and $P^{M_2}(\mathbf{v} \mid S{=}1)$, respectively, such that

$$P^{M_1}(\mathbf{v} \mid S{=}1) = P^{M_2}(\mathbf{v} \mid S{=}1) \tag{24}$$

$$Q^{M_1} \neq Q^{M_2} \tag{25}$$

Let M_1 be compatible with $\mathcal{G}'$ and M_2 with $\mathcal{G}'_{\underline{S}}$, enforcing $(\mathbf{V} \perp\!\!\!\perp S)_{P^{M_2}}$. Without loss of generality, all variables are assumed to be binary. The construction parametrizes P^{M_1} through its factors (as in lemma 4) and then parametrizes P^{M_2} to enforce (24). As a consequence, $P^{M_2}(\mathbf{v}) = P^{M_2}(\mathbf{v} \mid S{=}1) = P^{M_1}(\mathbf{v} \mid S{=}1)$.

In the sequel we consider every possible form in which p could manifest in $\mathcal{G}$:

case 1 $Y' \in Pa_S$

The causal effect in M_2:

$$\begin{aligned} Q^{M_2} &= P^{M_2}(y') = P^{M_1}(y' \mid S{=}1) \\ &= \frac{P^{M_1}(y', S{=}1)}{\sum_{Y'} P^{M_1}(y', S{=}1)} \\ &= \frac{P^{M_1}(S{=}1|y')P^{M_1}(y')}{P^{M_1}(S{=}1|y')P^{M_1}(y') + P^{M_1}(S{=}1|\overline{y'})P^{M_1}(\overline{y'})} \end{aligned}$$

Using lemma 4, let $P^{M_1}(S{=}1 \mid y') = \alpha$ and $P^{M_1}(S{=}1 \mid \overline{y'}) = \beta$ with $0 < \alpha, \beta < 1$ and $\alpha \neq \beta$ and $P^{M_1}(y') = 1/2$. We obtain:

$$Q^{M_2} = \frac{\alpha}{\alpha + \beta}$$

By the same reasoning $Q^{M_1} = 1/2$ which is never equal to Q^{M_2} given this parametrization.

case 2 There is a directed path p from Y' to S.

Let R be the parent of S in such path and let $\mathbf{W}$ be the set of variables in the path from Y' to R. Note that even if $\mathbf{W} \cap \mathbf{X} \neq \emptyset$, $Q^{M_2} = P(y')$, then, similar to the previous case:

$$Q^{M_2} = P^{M_1}(y' \mid S{=}1) = \frac{P^{M_1}(y', S{=}1)}{P^{M_1}(S{=}1)}$$

The numerator can be rewritten as:

$$\begin{aligned} P^{M_1}(y', S{=}1) &= \sum_R P^{M_1}(y', r, S{=}1) \\ &= \sum_R P^{M_1}(y')P^{M_1}(r \mid y')P^{M_1}(S{=}1 \mid r) \end{aligned}$$

Factorizing the denominator analogously, Q^{M_2} becomes:

$$Q^{M_2} = \frac{P^{M_1}(y')\sum_R P^{M_1}(r \mid y')P^{M_1}(S{=}1 \mid r)}{\sum_{Y'} P^{M_1}(y')\sum_R P^{M_1}(r \mid y')P^{M_1}(S{=}1 \mid r)}$$

Use lemma 7 to set $P^{M_1}(r \mid y') = 1/2 + \epsilon/2, P^{M_1}(r \mid \overline{y'}) = 1/2 - \epsilon/2$ where $\epsilon = (1/5)^k$ (using $p = 3/5, q = 2/5$). Also let $P^{M_1}(S{=}1 \mid r) = 2/3$ and $P^{M_1}(S{=}1 \mid \overline{r}) = 1/2$ and $P^{M_1}(y') = 1/2$. This parametrization leads to $Q^{M_2} = 1/2 + \epsilon/14$ and $Q^{M_1} = 1/2$ which are never equal.

case 3 The path p connecting Y' and S goes through an ancestor of both.

Let N be the common ancestor of Y' and S in p. Let R be the parent of S and Q the parent of Y' in the mentioned path. Let $\mathbf{W_1}$ and $\mathbf{W_2}$ be the nodes in the paths from N to Q and from N to R respectively. Consider an equivalent graph $\mathcal{G}''$ where the arrows in the subpath from N to Q are reversed. Any model constructed for $\mathcal{G}''$ can be translated to a model compatible with $\mathcal{G}'$ using lemma 6. Again we have $Q^{M_2} = P^{M_2}(y')$ Following the same derivation as in *case 2* yields:

$$Q^{M_2} = \frac{P^{M_1}(y', S{=}1)}{\sum_{Y'} P^{M_1}(y', S{=}1)}$$

The numerator of the last expression can be rewritten as:

$$\begin{aligned} P^{M_1}(y', S{=}1) &= \sum_Q P^{M_1}(y', q, S{=}1) \\ &= \sum_Q P^{M_1}(y' \mid q)P^{M_1}(q)P^{M_1}(S{=}1 \mid q) \end{aligned}$$

By rewriting the denominator similarly, and following an analogous process for Q^{M_1}, we have:

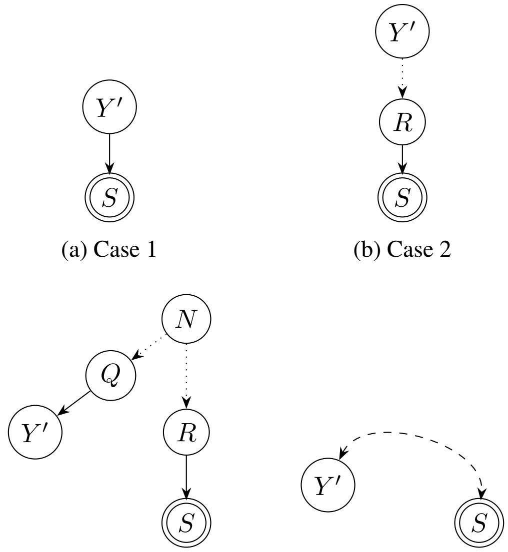
$$Q^{M_1} = \sum_Q P^{M_1}(y' \mid q)P^{M_1}(q)$$

$$Q^{M_2} = \frac{\sum_Q P^{M_1}(y' \mid q)P^{M_1}(q)P^{M_1}(S{=}1 \mid q)}{\sum_{Y',Q} P^{M_1}(y' \mid q)P^{M_1}(q)P^{M_1}(S{=}1 \mid q)}$$

Lemma 7 can be employed to set $P^{M_1}(r \mid q) = 1/2 + \epsilon/2, P^{M_1}(r \mid \overline{q}) = 1/2 - \epsilon/2$ where $\epsilon = (1/5)^k$ (using $p = 3/5, q = 2/5$). Define $P^{M_1}(S{=}1 \mid r) = 2/3$ and $P^{M_1}(S{=}1 \mid \overline{r}) = 1/2$. Calculate $P^{M_1}(S{=}1 \mid q)$ as $\sum_R P^{M_1}(r \mid q)P^{M_1}(S{=}1 \mid r)$. Also let $P^{M_1}(y' \mid q) = 3/4, P^{M_1}(y' \mid \overline{q}) = 1/2$, finally $P^{M_1}(q) = 1/2$. This parametrization leads to:

$$Q^{M_1} = \tfrac{5}{8} \qquad\qquad Q^{M_2} = \tfrac{5}{8} + \tfrac{\epsilon}{56}$$

which are never equal.



(a) Case 1 (b) Case 2

(c) Case 3 (d) Case 4

Figure 5: Graphical representation of the cases stated in Thm. 1

case 4 p is a confounding path between Y' and S consisting of unobservable variables.
The models for this case can be constructed as in *case 3*, then moving the variables in the in the path $Q - R$ (included) from the set of observables to the set of unobservables.

□

In order to prove non-recoverability, it is imperative to construct structural causal models that serve as counterexamples to the recoverability of the causal effect. The following lemmas are useful to construct such models. The first one, lemma 4 licenses the the direct specification of the conditional distributions of any variable given its parents, in accordance to the causal diagram $\mathcal{G}$.

Lemma 4 (Family Parametrization)**.** *Let $\mathcal{G}$ be a causal diagram over a set $\mathbf{V}$ of n variables. Consider also, a set of conditional distributions $P(v_i \mid pa_i), 1 \leqslant i \leqslant n$ such that Pa_i is the set of nodes in $\mathcal{G}$ from which there are outgoing edges pointing into V_i. Then, there exists a model M compatible with $\mathcal{G}$ that induces $P(\mathbf{v}) = \prod_{i=1}^{n} P(v_i \mid pa_i)$.*

Proof. (By construction) For every V_i define any ordering on the values of its domain, and let $v_i^{(j)}$ refer to the j^{th} value in that order. Also, define a continuous unobservable variable $U_i \sim U[0,1]$ (uniformly distributed in the interval $[0,1]$) for every variable $V_i \in \mathbf{V}$. Then, construct a structural causal model $M = \langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{u}) \rangle$ where:

- $\mathbf{V}$ is the same set of observables in $\mathcal{G}$
- $\mathbf{U} = \bigcup_{i=1}^{n} U_i'$
- $\mathcal{F} = \left\{ f_i(pa_i, u_i) = \inf_j \left\{ \sum_{k=1}^{j} P(v_i^{(j)} | pa_i) \geqslant u_i \right\}, 1 \leqslant i \leqslant n \right\}$
- $U_i \sim U[0,1], 1 \leqslant i \leqslant n$

At every variable V_i, given a particular configuration of Pa_i, M simulates its value using the distribution $P(v_i \mid pa_i)$. By the Markov property, the joint distribution will be equal to the product of those distributions. □

The following lemma extends a result from (Tian 2002) to the context of recoverability.

Lemma 5. *Let $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$ be two disjoint sets of variables and $\mathbf{T} \subset \mathbf{V}$ another set of variables. If $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable in $\mathcal{G}$ from $P(\mathbf{v} \mid S{=}1), P(\mathbf{t})$, then $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable in the graph resulted from adding a directed or bidirected edge to $\mathcal{G}$. Equivalently, if $P_{\mathbf{x}}(\mathbf{y})$ is recoverable in $\mathcal{G}$ from $P(\mathbf{v} \mid S{=}1), P(\mathbf{t})$, then $P_{\mathbf{x}}(\mathbf{y})$ is still recoverable in the graph resulted from removing a directed or bidirected edge from $\mathcal{G}$.*

Proof. The proof is analogous to the proof of lemma 8 in (Tian 2002) with the extra consideration of selection bias. For any $V_i \in \mathbf{V} \cup \{S\}$ let Pa_i^+ be defined as the set of observable and unobservable parents of V_i in $\mathcal{G}$.

If $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable in $\mathcal{G}$, then there exist two models with the same causal graph $\mathcal{G}$, M_1 and M_2 such that

$$\begin{aligned} &P^{M_1}(\mathbf{v} \mid S{=}1) = P^{M_2}(\mathbf{v} \mid S{=}1) > 0, \\ &P^{M_1}(\mathbf{t}) = P^{M_2}(\mathbf{t}) > 0 \\ &\text{and } P_{\mathbf{x}}^{M_1}(\mathbf{y}) \neq P_{\mathbf{x}}^{M_2}(\mathbf{y}) \end{aligned} \tag{26}$$

where

$$P^{M_k}(\mathbf{v}|S{=}1) \tag{27}$$

$$= \sum_{\mathbf{U}} \frac{P(S{=}1|pa_s^+)}{P(S{=}1)} \prod_{V_i \in \mathbf{V}} P(v_i|pa_i^+)P(\mathbf{u}) \quad , k = 1,2 \tag{28}$$

$$P^{M_k}(\mathbf{t}) = \sum_{\mathbf{U}} \sum_{\mathbf{V}\setminus\mathbf{T}} \prod_{V_i \in \mathbf{V}} P(v_i \mid pa_i^+)P(\mathbf{u}) \quad , k = 1,2 \tag{29}$$

For a graph $\mathcal{G}'$ with extra edges added to $\mathcal{G}$, we can always construct new models in such a way that the added edges are ineffective.

(i) Let $\mathcal{G}'$ be the graph identical to $\mathcal{G}$ except with an extra edge $W \to V_j$. Then $P(\mathbf{v} \mid S{=}1)$ and $P(\mathbf{t})$ decompose as

$$P(\mathbf{v}|S{=}1) \tag{30}$$

$$= \sum_{\mathbf{U}} \frac{P(S{=}1|pa_S^+)}{P(S{=}1)} P(v_j|pa_{v_j}^+, w) \prod_{\substack{V_i \in \mathbf{V}, \\ V_i \neq V_j}} P(v_i|pa_i^+)P(\mathbf{u}) \tag{31}$$

$$P(\mathbf{t}) = \sum_{\mathbf{U}} \sum_{\mathbf{V}\setminus\mathbf{T}} P(v_j|pa_{v_j}^+, w) \prod_{\substack{V_i \in \mathbf{V}, \\ V_i \neq V_j}} P(v_i|pa_i^+)P(\mathbf{u}) \tag{32}$$

We construct two models M_1' and M_2' with causal graph $\mathcal{G}'$ as:

$$P^{M_k'}(v_i|pa_i^+)=P^{M_k}(v_i|pa_i^+) \qquad ,i\neq j,k=1,2 \tag{33}$$

$$P^{M_k'}(S{=}1|pa_S^+)=P^{M_k}(S{=}1|pa_S^+) \qquad ,k=1,2 \tag{34}$$

$$P^{M_k'}(v_j|pa_{v_j}^+,w)=P^{M_k}(v_j|pa_{v_j}^+) \qquad ,k=1,2 \tag{35}$$

$$P^{M_k'}(\mathbf{u}) = P^{M_k}(\mathbf{u}) \qquad ,k=1,2 \tag{36}$$

Clearly if the pair (M_1, M_2) satisfies (26) so does (M_1', M_2'). Hence $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable in $\mathcal{G}'$.

(ii) Let $\mathcal{G}'$ be the graph identical to $\mathcal{G}$ except with an extra edge $V_l \leftrightarrow V_j$. Then $P(\mathbf{v} \mid S{=}1)$ and $P(\mathbf{t})$ decompose as

$$P(\mathbf{v}|S{=}1) \tag{37}$$

$$=\sum_{u'} P(u') \sum_{\mathbf{U}} \frac{P(S{=}1|pa_S^+)}{P(S{=}1)} P(v_j|pa_{v_j}^+,u') P(v_l|pa_{v_l}^+,u') \prod_{V_i\in\mathbf{V},V_i\neq V_j,V_i\neq V_l} P(v_i|pa_i^+)P(\mathbf{u}) \tag{38}$$

$$P(\mathbf{t}) = \sum_{u'} P(u') \sum_{\mathbf{U}} \sum_{\mathbf{V}\setminus\mathbf{T}} P(v_j|pa_{v_j}^+,u')P(v_l|pa_{v_l}^+,u') \prod_{V_i\in\mathbf{V},V_i\neq V_j,V_i\neq V_l} P(v_i|pa_i^+)P(\mathbf{u}) \tag{39}$$

Where U' represents a new unobservable variable. We construct two models M_1' and M_2' with causal graph $\mathcal{G}'$ as:

$$P^{M_k'}(v_i|pa_i^+)=P^{M_k}(v_i|pa_i^+) \qquad ,i\neq j,i\neq l,k=1,2 \tag{40}$$

$$P^{M_k'}(S{=}1|pa_S^+)=P^{M_k}(S{=}1|pa_S^+) \qquad ,k=1,2 \tag{41}$$

$$P^{M_k'}(v_i|pa_i^+,u')=P^{M_k}(v_i|pa_i^+) \qquad ,i{=}j,l,k=1,2 \tag{42}$$

$$P^{M_k'}(\mathbf{u})=P^{M_k}(\mathbf{u}) \qquad ,k=1,2 \tag{43}$$

Again, if the pair (M_1, M_2) satisfies (26), so does (M_1', M_2'). Hence $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable in $\mathcal{G}'$.

(iii) Let $\mathcal{G}'$ be the graph identical to $\mathcal{G}$ except with an extra edge $W \to S$. Then $P(\mathbf{t})$ is exactly the same and $P(\mathbf{v} \mid S{=}1)$ decomposes as

$$P(\mathbf{v}|S{=}1)=\sum_{\mathbf{U}} \frac{P(S{=}1|pa_S^+,w)}{P(S{=}1)} \prod_{V_i\in\mathbf{V}} P(v_i|pa_i^+)P(\mathbf{u}) \tag{44}$$

We construct two models M_1' and M_2' with causal graph $\mathcal{G}'$ as:

$$P^{M_k'}(v_i|pa_i^+)=P^{M_k}(v_i|pa_i^+) \qquad ,k=1,2 \tag{45}$$

$$P^{M_k'}(S{=}1|pa_S^+,w)=P^{M_k}(S{=}1|pa_S^+) \qquad ,k=1,2 \tag{46}$$

$$P^{M_k'}(\mathbf{u})=P^{M_k}(\mathbf{u}) \qquad ,k=1,2 \tag{47}$$

Since $P(S{=}1) = \sum_{pa_S^+} P(S{=}1 \mid pa_S^+)P(pa_S^+)$, that distribution will remain the same. Then, if pair (M_1, M_2) satisfies (26) so does (M_1', M_2'). Hence $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable in $\mathcal{G}'$.

(iv) Let $\mathcal{G}'$ be the graph identical to $\mathcal{G}$ except with an extra edge $V_j \leftrightarrow S$. Then $P(\mathbf{v} \mid S{=}1)$ and $P(\mathbf{t})$ decompose as

$$P(\mathbf{v}|S{=}1) \tag{48}$$

$$= \sum_{u'} P(u') \sum_{\mathbf{U}} \frac{P(S{=}1|pa_S^+,u')}{P(S{=}1)} P(v_j|pa_{v_j}^+,u') \prod_{V_i\in\mathbf{V},V_i\neq V_j} P(v_i \mid pa_i^+)P(\mathbf{u}) \tag{49}$$

$$P(\mathbf{t}) = \sum_{u'} P(u') \sum_{\mathbf{U}} \sum_{\mathbf{V}\setminus\mathbf{T}} P(v_j|pa_{v_j}^+,u') \tag{50}$$

$$\prod_{V_i\in\mathbf{V},V_i\neq V_j} P(v_i|pa_i^+)P(\mathbf{u}) \tag{51}$$

Where U' represents a new unobservable variable. We construct two models M_1' and M_2' with causal graph $\mathcal{G}'$ as:

$$P^{M_k'}(v_i|pa_i^+)=P^{M_k}(v_i|pa_i^+) \qquad ,i \neq j, k = 1,2 \tag{52}$$

$$P^{M_k'}(S{=}1|pa_S^+,u')=P^{M_k}(S{=}1|pa_S^+) \qquad ,k = 1,2 \tag{53}$$

$$P^{M_k'}(v_j|pa_{v_j}^+,u')=P^{M_k}(v_j|pa_{v_j}^+) \qquad ,k = 1,2 \tag{54}$$

$$P^{M_k'}(\mathbf{u})=P^{M_k}(\mathbf{u}) \qquad ,k = 1,2 \tag{55}$$

Again, if the pair (M_1, M_2) satisfies (26), so does (M_1', M_2'). Hence $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable in $\mathcal{G}'$.

□

The following lemma permits the construction of a structural causal model M compatible with a causal diagram $\mathcal{G}$, using another model compatible with a related, but different, causal diagram $\mathcal{G}'$ where some arrows in a chain of variables have the reverse direction.

Lemma 6 (Chain Reversal)**.** *Consider a causal diagram $\mathcal{G}$ and a probability distribution $P(\mathbf{v})$ induced by any SCM M compatible with $\mathcal{G}$. If $\mathcal{G}$ contains a chain of vertices $R_1 \to R_2, \ldots, R_\ell$ where each node represents a binary random variable, for every $1 \leq i \leq \ell$ the only incoming edge into R_i comes from R_{i-1}. Then, there exists another model M' where the direction of the arrows along the chain $R_1 \to R_2, \ldots, R_\ell$ is reversed compatible with same distribution.*

Proof. (By construction) Given M and any probability distribution $P(\mathbf{v})$ induced by it, compute the joint distribution $P(r_1, \ldots, r_\ell, t)$. Construct a new model M' with the same set of observable variables and identical functions for all variables but for $R_1, \ldots, R_\ell, T$. For those, assign the functions $f_{R_i}(r_{i-1}, U_{R_i}), 1 \leqslant i \leqslant \ell - 1$ as in lemma 4. Also, let $f_{R_\ell}(U_{R_\ell}) = U_{R_\ell}, P(U_{R_\ell}) = P(r_\ell)$. By lemma 4 the sub-models composed of $R_1, \ldots, R_\ell, T$ in M' and M produce the exact same distribution and since the set of parents and function for every other part of the model are exactly the same the overall distribution is identical. □

Finally, the following lemma allows to simplify the parametrization of an arbitrarily long chain of binary variables.

Lemma 7 (Collapsible Path Parametrization)**.** *Consider a causal diagram $\mathcal{G}$ and a probability distribution $P(\mathbf{v})$ induced by any SCM compatible with $\mathcal{G}$. If $\mathcal{G}$ contains a chain $W_0 \to W_1 \to \ldots \to W_k$, where each W_i represents a binary random variable, for every $1 \leqslant i \leqslant k$ the only incoming edge to W_i is from W_{i-1}, and every conditional distribution $P(w_i \mid w_{i-1}) = p$, $P(w_i \mid \overline{w_{i-1}}) = q$, for some $0 < p, q < 1$. Then, the conditional distribution $P(w_k \mid w_0) = \frac{q-(p-1)(p-q)^k}{q-p+1}$, $P(w_k \mid \overline{w_0}) = \frac{q-q(p-q)^k}{q-p+1}$.*

Proof. Since $W_0, \ldots, W_k$ is a chain, the value of W_k is a function of W_0 when all other $W_1, \ldots, W_{k-1}$ are marginalized. All W_i, $1 \leqslant i \leqslant k$ are independent of any other variable given W_0. Therefore, the distribution $P(w_k \mid w_0)$ is equal to $\sum_{i=1}^{k-1} \prod_{i=1}^{k} P(w_i \mid w_{i-1})$, because any other variable can be removed from any product in this expression and summed out. This distribution can be calculated as the product of 2x2 matrices corresponding to the conditional distributions $P(w_i \mid w_{i-1})$ when encoded as $W_M = \begin{bmatrix} p & q \\ 1-p & 1-q \end{bmatrix}$. The product of k of such matrices is readily available if W_M is decomposed using its eigenvalues $\{1, p-q\}$ and eigenvectors $\left\{ \left[\frac{q}{(1-p)}, -1 \right], [1, 1] \right\}$:

$$P(w_k \mid w_0) = \sum_{i=1}^{k-1} \prod_{i=1}^{k} P(w_i \mid w_{i-1}) = (W_M)^k$$
$$= \begin{bmatrix} \frac{q-(p-1)(p-q)^k}{q-p+1} & \frac{q-q(p-q)^k}{q-p+1} \\ 1 - \frac{q-(p-1)(p-q)^k}{q-p+1} & 1 - \frac{q-q(p-q)^k}{q-p+1} \end{bmatrix} \tag{56}$$

□

Appendix B. Proof of Theorem 2

Theorem 2. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two disjoint sets of variables and $\mathcal{G}$ a causal diagram over $\mathbf{V}$ and S. Let $\mathbf{D}{=}An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}$ and $D_1, \ldots, D_\ell$ be the c-components of $\mathcal{G}_{\mathbf{D}}$. Then, the effect $P_{\mathbf{x}}(\mathbf{y})$ is recoverable from $P(\mathbf{v}|S{=}1)$ if and only if each $D_i, i = 1, \ldots, \ell$ is recoverable by the function* RC.

Proof. (if) We have

$$P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{D}\setminus\mathbf{Y}} \prod_i Q[D_i] \tag{57}$$

Because RC is sound, if every D_i can be recovered by that procedure, $P_{\mathbf{x}}(\mathbf{y})$ is then recoverable by (57).

(only if) RC may fail in two ways: first, when IDENTIFY$(D_i, C_i, Q[C_i])$ fails, in which case the effect is not identifiable hence not recoverable. In the second case, $Q[D_i]$ cannot be recovered by RC if there exists some $\mathbf{C} \subset An(D_i \cup \{S\}), D_i \subseteq \mathbf{C}$ such that every c-component of $\mathcal{G}_{\mathbf{C}}$ contains an ancestor of S. We will show that there exists an active path p' from S to some $D' \in D_i$ in $\mathcal{G}_{\mathbf{C}}$. Let C_i be the c-component of $\mathcal{G}_{\mathbf{C}}$ such that $D_i \subseteq C_i$:

- If $C_i \cap An(D_i) \cap An(S) \neq \emptyset$, let C' be any element in that intersection. Then p' is the directed path from C' to S and from C' to some $D' \in D_i$ (possibly of length 0 if $C' \in D_i$). If the segment $C' \to \ldots \to S$ contains any variable in $X' \in \mathbf{X}$, then the outgoing edge of X' exists in $\mathcal{G}^{pbd}_{\mathbf{XY}}$ unless there is some $Y' \in \mathbf{Y}$ that is a descendant of X' through that edge
- Else $(C_i \cap An(S)) \setminus An(D_i)$ has to be non-empty, let C' be any element in that set connected with some $W \in C_i \cap An(D_i)$ with a bidirected arrow. Such C' and W must exist for C_i to be a c-component in $\mathcal{G}_{\mathbf{C}}$. Then p' is formed by the subpaths from $C' \to \cdots \to S$ (possibly of length zero if $C' = S$), $C' \leftarrow\!\!\dashrightarrow W$ and $W \to \cdots \to D'$, for some $D' \in D_i$.

Since $D_i \subseteq \mathbf{D} = An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}$, there exists a directed path from D' to some $Y' \in \mathbf{Y}$ (possibly of length zero if $D' \in \mathbf{Y}$) that contains no element in $\mathbf{X}$. As a consequence, we can construct an active path p in $\mathcal{G}^{pbd}_{\mathbf{XY}}$ formed by p' and the directed path from D' to Y'. This implies $(\mathbf{Y} \not\perp\!\!\!\perp S)_{\mathcal{G}^{pbd}_{\mathbf{XY}}}$ and by Thm. 1 $P_{\mathbf{x}}(\mathbf{y})$ is not recoverable from $P(\mathbf{v} \mid S{=}1)$. □

Appendix C. Proof of Theorem 3

In order to prove this statement we will leverage the properties of the c-component decomposition of Semi-Markovian causal models. Two of such properties are given by the following lemma from (Tian and Pearl 2002b).

Lemma 8. *Let $\mathbf{H} \subseteq \mathbf{V}$, and assume that $\mathbf{H}$ is partitioned into c-components $H_1, \ldots, H_l$ in the subgraph $\mathcal{G}_{\mathbf{H}}$. Then we have*

(i) $Q[H]$ decomposes as

$$Q[\mathbf{H}] = \prod_i Q[H_i]. \tag{58}$$

(ii) Let a topological order of the variables in $\mathbf{H}$ be $V_{h_1} < \cdots < V_{h_k}$ in $\mathcal{G}_{\mathbf{H}}$. Let $H^{\leq i} = \{V_{h_1}, \ldots, V_{h_i}\}$ be the set of variables in $\mathbf{H}$ ordered before V_{h_i} (including V_{h_i}), and $H^{>i} = \mathbf{H} \setminus H^{\leq i}$ for $i = 1, \ldots, k$, and $H^{\leq 0} = \emptyset$. Then each $Q[H_j]$, $j = 1, \ldots, l$, is computable from $Q[\mathbf{H}]$ and given by

$$Q[H_j] = \prod_{\{i \mid V_{h_i} \in H_j\}} \frac{Q[H^{\leq i}]}{Q[H^{\leq i-1}]}, \tag{59}$$

where each $Q[H^{\leq i}]$, $i = 0, 1, \ldots, k$, is given by

$$Q[H^{\leq i}] = \sum_{h^{>i}} Q[\mathbf{H}]. \tag{60}$$

Lemma 9. *Let $\mathbf{H} \subseteq \mathbf{V} \cup \{S\}$ and let $f(P(\mathbf{v}|S{=}1))$ be some recoverable quantity such that*

$$f(P(\mathbf{v} \mid S{=}1)) = \frac{Q[\mathbf{H}](\mathbf{v}, S{=}1)}{P(S{=}1)}. \tag{61}$$

Given $P(\mathbf{t})$, $\mathbf{T} \subset \mathbf{V}$, let $\mathbf{T}^0_{\mathbf{H}} = \mathbf{T} \setminus De(\mathbf{V} \backslash \mathbf{H})_{\mathcal{G}}$, $\mathbf{T}'$ be the set of all variables in $\mathbf{H}$ such that $(\mathbf{T}' \perp\!\!\!\perp S \mid \mathbf{T}^0_{\mathbf{H}})_{\mathcal{G}_{\mathbf{H}}}$, and let $\mathbf{T}_{\mathbf{H}} = \mathbf{T}^0_{\mathbf{H}} \cup \mathbf{T}'$, then $P_{\mathbf{v} \backslash \mathbf{h}}(\mathbf{t}_{\mathbf{H}})$ is recovered as

$$P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}_{\mathbf{H}}) = \frac{\sum_{(\mathbf{H}\backslash\{S\})\backslash\mathbf{T}_{\mathbf{H}}} f(P(\mathbf{v}|S{=}\,1))}{\sum_{(\mathbf{H}\backslash\{S\})\backslash\mathbf{T}^0_{\mathbf{H}}} f(P(\mathbf{v}|S{=}\,1))} P(\mathbf{t}^0_{\mathbf{H}}) \tag{62}$$

Proof.

$$P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}_{\mathbf{H}}) = P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}' \mid \mathbf{t}^0_{\mathbf{H}}) P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}^0_{\mathbf{H}}) \tag{63}$$

Since $\mathbf{T}^0_{\mathbf{H}}$ contains no descendants of $\mathbf{V}\backslash\mathbf{H}$ we have that $(\mathbf{T}^0_{\mathbf{H}} \perp\!\!\!\perp \mathbf{V}\backslash\mathbf{H})_{\mathcal{G}_{\overline{\mathbf{V}\backslash\mathbf{H}}}}$, then by rule 3 of do-calculus we have:

$$P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}_{\mathbf{H}}) = P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}' \mid \mathbf{t}^0_{\mathbf{H}}) P(\mathbf{t}^0_{\mathbf{H}}) \tag{64}$$

Furthermore, since $(S \perp\!\!\!\perp \mathbf{T}' \mid \mathbf{T}^0_{\mathbf{H}})_{\mathcal{G}_{\mathbf{H}}}$ we can add $S{=}1$ to the conditioning part of the expression,

$$P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}_{\mathbf{H}}) = P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}' \mid \mathbf{t}^0_{\mathbf{H}}, S{=}1) P(\mathbf{t}^0_{\mathbf{H}}) \tag{65}$$

$$= \frac{P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}_{\mathbf{H}}, S{=}1)}{P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}^0_{\mathbf{H}}, S{=}1)} P(\mathbf{t}^0_{\mathbf{H}}) \tag{66}$$

$$= \frac{\sum_{(\mathbf{H}\backslash\{S\})\backslash\mathbf{T}_{\mathbf{H}}} Q[\mathbf{H}](\mathbf{v}, S{=}1)}{\sum_{(\mathbf{H}\backslash\{S\})\backslash\mathbf{T}^0_{\mathbf{H}}} Q[\mathbf{H}](\mathbf{v}, S{=}1)} P(\mathbf{t}^0_{\mathbf{H}}) \tag{67}$$

$$= \frac{\sum_{(\mathbf{H}\backslash\{S\})\backslash\mathbf{T}_{\mathbf{H}}} \frac{Q[\mathbf{H}](\mathbf{v}, S=1)}{P(S=1)}}{\sum_{(\mathbf{H}\backslash\{S\})\backslash\mathbf{T}^0_{\mathbf{H}}} \frac{Q[\mathbf{H}](\mathbf{v}, S=1)}{P(S=1)}} P(\mathbf{t}^0_{\mathbf{H}}) \tag{68}$$

The operand in both sums is precisely the distribution $f(P(\mathbf{v}|S{=}\,1))$

$$P_{\mathbf{v}\backslash\mathbf{h}}(\mathbf{t}_{\mathbf{H}}) = \frac{\sum_{(\mathbf{H}\backslash\{S\})\backslash\mathbf{T}_{\mathbf{H}}} f(P(\mathbf{v}|S{=}\,1))}{\sum_{(\mathbf{H}\backslash\{S\})\backslash\mathbf{T}^0_{\mathbf{H}}} f(P(\mathbf{v}|S{=}\,1))} P(\mathbf{t}^0_{\mathbf{H}}) \tag{69}$$

□

Lemma 10. *Let $\mathbf{H} \subseteq \mathbf{V} \cup \{S\}$, and assume that $\mathbf{H}$ is partitioned into c-components $H_1, \ldots, H_l, H_s$, where $S \in H_s$, in the subgraph $\mathcal{G}_{\mathbf{H}}$. If*

$$f(P(\mathbf{v} \mid S{=}1)) = \frac{Q[H_s](\mathbf{v}, S{=}1)}{P(S{=}1)} \prod_i Q[H_i] \tag{70}$$

where $f(P(\mathbf{v} \mid S{=}1))$ is some recoverable quantity, then, for $j = 1, \ldots, l$, $Q[H_j]$ is recoverable if $H_j \cap An(H_s) = \emptyset$, that is, if H_j contains no ancestors of H_s.

Proof. Let $V_{h_1} < \cdots < V_{h_k}$ be a topological order in $\mathcal{G}_{\mathbf{H}}$ such that $An(H_s) < \mathbf{H} \setminus An(H_s)$. Let $H_{S'} = \mathbf{H} \setminus H_s$, then $Q[H_{S'}]$ is given by

$$Q[H_{S'}] = \frac{P(S{=}1)}{Q[H_s](\mathbf{v}, S{=}1)} f(P(\mathbf{v} \mid S{=}1)) \tag{71}$$

$Q[H_s] = \prod_{\{i|V_{h_i} \in H_s\}} Q[H^{<i}]/Q[H^{<i-1}]$, where each factor is a function only of $Pa\big(H^{<i} \cap H_s\big)$, and thus $Q[H_s]$ is a function of $An(H_s)$. If H_j contains no ancestor of H_s, then all variables in H_j are ordered after the variables in $An(H_s)$, then for each $V_{h_i} \in H_j$, $h^{>i} \cup \{V_{h_i}\}$ contains no variables in $An(H_s)$. Therefore,

$$Q[H^{\leq i}] = \sum_{h^{>i}} \frac{P(s)}{Q[H_s](\mathbf{v}, S{=}1)} f(P(\mathbf{v} \mid S{=}1))$$

$$= \frac{P(s)}{Q[H_s](\mathbf{v}, S{=}1)} \sum_{h^{>i}} f(P(\mathbf{v} \mid S{=}1)) \tag{72}$$

$$Q[H^{\leq i-1}] = \frac{P(S{=}1)}{Q[H_s](\mathbf{v}, S{=}1)} \sum_{h^{>i}, V_{h_i}} f(P(\mathbf{v} \mid S{=}1)) \tag{73}$$

and finally

$$\frac{Q[H^{\leq i}]}{Q[H^{\leq i-1}]} = \frac{\sum_{h^{>i}} f(P(\mathbf{v} \mid S{=}1))}{\sum_{h^{>i}, V_{h_i}} f(P(\mathbf{v} \mid S{=}1))} \tag{74}$$

Since $Q[H_j]$ is given by

$$Q[H_j] = \prod_{\{i|V_{h_i} \in H_j\}} \frac{Q[H^{\leq i}]}{Q[H^{\leq i-1}]} \tag{75}$$

it is recoverable. □

Theorem 3. *Let $\mathbf{H} \subseteq \mathbf{V} \cup \{S\}$, such that $\mathbf{H}$ is partitioned into c-components $H_1, \ldots, H_l, H_s$ in the subgraph $\mathcal{G}_{\mathbf{H}}$, where $S \in H_s$. Assume*

$$f(P(\mathbf{v} \mid S{=}1)) = \frac{Q[H_s](\mathbf{v}, S{=}1)}{P(S{=}1)} \prod_i Q[H_i], \tag{5}$$

where $f(P(\mathbf{v}|S{=}1))$ is some recoverable quantity, and $P(\mathbf{t})$ is available. Let $\mathbf{T}^0_{\mathbf{H}} = \mathbf{T} \setminus De(\mathbf{V}\backslash\mathbf{H})_{\mathcal{G}}$ and $\mathbf{T}'$ be the set of all variables in $\mathbf{H}$ such that $(\mathbf{T}' \perp\!\!\!\perp S \mid \mathbf{T}^0_{\mathbf{H}})_{\mathcal{G}_{\mathbf{H}}}$. Also, let $\mathbf{T}_{\mathbf{H}} = \mathbf{T}^0_{\mathbf{H}} \cup \mathbf{T}'$ and let $\mathbf{R}_{\mathbf{H}} = \mathbf{H} \setminus \mathbf{T}_{\mathbf{H}}$. Then, for $j{=}1, \ldots, l$, $Q[H_j]$ is recoverable if H_j contains no variables that are both ancestors of H_s and belong to $\mathbf{R}_{\mathbf{H}}$ or its children (i.e. $H_j \cap An(H_s) \cap Ch(\mathbf{R}_{\mathbf{H}}) = \emptyset$) in $\mathcal{G}_{\mathbf{H}}$.

Proof. Let $V_{h_1} < \cdots < V_{h_{|\mathbf{H}|}}$ be a topological order in $\mathcal{G}_{\mathbf{H}}$ such that $An(H_s) < \mathbf{H} \setminus An(H_s)$. Let $H^{\leq i} = \{V_{h_1}, \ldots, V_{h_i}\}$ be the set of variables in $\mathbf{H}$ ordered before V_{h_i} (including V_{h_i}), and $H^{>i} = \mathbf{H} \setminus H^{\leq i}$ for $i = 1, \ldots, k$, and $H^{\leq 0} = \emptyset$. Recall from Lemma 8 that each $Q[H_j]$, $j = 1, \ldots, l$, is computable from $Q[\mathbf{H}]$ and given by

$$Q[H_j] = \prod_{\{i|V_{h_i} \in H_j\}} \frac{Q[H^{\leq i}]}{Q[H^{\leq i-1}]}, \tag{76}$$

where each $Q[H^{\leq i}]$, $i = 0, 1, \ldots, k$, is given by

$$Q[H^{\leq i}] = \sum_{h^{>i}} Q[\mathbf{H}]. \tag{77}$$

We will describe two ways to compute factors of the form $Q[H^{\leq i}]/Q[H^{\leq i-1}]$

(i) *For variables* $V_{h_i} \in H_j \setminus An(H_s)$: Every $V_{h_i} \in H_j \setminus An(H_s)$ is ordered after the variables in $An(H_s)$, then the same reasoning in the proof for Lemma 10 follows and $Q[H^{\leq i}]/Q[H^{\leq i-1}]$ is given by (74).

(ii) *For variables* $V_{h_i} \in H_j \setminus Ch(\mathbf{R_H})$: Let V_{h^*} be the lastest variable in $H_j \setminus Ch(\mathbf{R_H})$ according to the topological order. From (76) $Q[H_j]$ can be separated in two parts:

$$Q[H_j] = Q[H_j \cap H^{\leq *}] \prod_{\{i|V_{h_i} \in H_j \setminus An(H_s)\}} \frac{Q[H^{\leq i}]}{Q[H^{\leq i-1}]} \quad (78)$$

The first factor in (78) is equal to

$$Q[H_j \cap H^{\leq *}] = \sum_{H_j \setminus H^{\leq *}} Q[H_j] \quad (79)$$

Consider the graph $PJ(\mathcal{G}_\mathbf{H}, H^{\leq *} \cap \mathbf{T_H})$ where variables in $\mathbf{R_H}$ and $H^{>*}$ are hidden, and let $Q_R^*[.]$ denote the c-factors for that graph.

$$Q_R^*[H^{\leq *} \cap \mathbf{T_H}] = \sum_{\mathbf{H} \setminus (H^{\leq *} \cap \mathbf{T_H})} Q[\mathbf{H}] \quad (80)$$

$$= \sum_{\mathbf{T_H} \setminus H^{\leq *}} \sum_{\mathbf{R_H}} Q[\mathbf{H}] \quad (81)$$

$$= \sum_{\mathbf{T_H} \setminus H^{\leq *}} P_{\mathbf{v} \setminus \mathbf{h}}(\mathbf{t_H}) \quad (82)$$

By Lemma 9 we can recover $P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})$ from $f(P(\mathbf{v} \mid S{=}1))$ and $P(\mathbf{t})$ and by Lemma 2 in (Tian and Pearl 2002b), each c-component $T_1, \ldots T_p$ of $PJ(G_\mathbf{H}, \mathbf{T_H} \cap H^{\leq *})$ is recoverable from $\sum_{\mathbf{T_H} \setminus H^{\leq *}} P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})$ as:

$$Q_R^*[T_k] = \prod_{\{i|V_{h_i} \in T_k\}} \frac{Q_R^*[H^{\leq i} \cap \mathbf{T_H}]}{Q_R^*[H^{\leq i-1} \cap \mathbf{T_H}]} \quad (83)$$

$$= \prod_{\{i|V_{h_i} \in T_k\}} \frac{\sum_{h^{>i} \cap h^{\leq *} \cap \mathbf{T_H}} Q_R^*[H^{\leq *} \cap \mathbf{T_H}]}{\sum_{h^{>i-1} \cap h^{\leq *} \cap \mathbf{T_H}} Q_R^*[H^{\leq *} \cap \mathbf{T_H}]} \quad (84)$$

$$= \prod_{\{i|V_{h_i} \in T_k\}} \frac{\sum_{h^{>i} \cap \mathbf{T_H}} P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})}{\sum_{h^{>i-1} \cap \mathbf{T_H}} P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})} \quad (85)$$

All variables in H_j that come before V_{h_*} must belong to $An(H_s) \setminus Ch(\mathbf{R_H})$. Therefore, $H_j \cap H^{\leq *} \cap Ch(\mathbf{R}) = \emptyset$, and $(H_j \cap H^{\leq *})$ is a c-component in $PJ(G_\mathbf{H}, \mathbf{T_H} \cap H^{\leq *})$ (no new bidirected arrow incoming to any node in $H_j \cap H^{\leq *}$). Then, there exists $T_k = H_j \cap H^{\leq *}$ that is recoverable.

Finally, $Q[H_j]$ given by

$$Q[H_j] = \prod_{\{i|V_{h_i} \in H_j \cap An(H_s)\}} \frac{\sum_{h^{>i} \cap \mathbf{T_H}} P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})}{\sum_{h^{>i-1} \cap \mathbf{T_H}} P_{\mathbf{v}\setminus\mathbf{h}}(\mathbf{t_H})} \times \prod_{\{i|V_{h_i} \in H_j \setminus An(H_s)\}} \frac{\sum_{h^{>i}} f(P(\mathbf{v}|S{=}1))}{\sum_{h^{>i}, V_{h_i}} f(P(\mathbf{v}|S{=}1))} \quad (86)$$

□

Appendix D. Proof of Theorem 4 and 5

Lemma 11. *The procedure* RCE *is sound.*

Proof. $P(\mathbf{v}|S{=}1)$ decomposes according to (5) which in turn is consistent with (70) and constitutes $\mathcal{P}$ in the initial call to RCE. If $\mathcal{G}$ contains variables that are not ancestors of S or $\mathbf{E} = D_i$, then they can be summed out without altering the result.

We can further decompose $\mathcal{P} = f(P(\mathbf{v} \mid S{=}1))$ as

$$f(P(\mathbf{v} \mid S{=}1)) = \frac{Q[C_s]}{P(S{=}1)} \prod_i Q[B_i] \prod_i Q[C_i], \quad (87)$$

where C_s is the c-component of $\mathcal{G}$ to which S belongs, each C_i is recoverable by Thm. 3 and the product over B_is account for the remaining c-components of $\mathcal{G}$. If $\mathbf{E} \subset C_i$ for some C_i we know from (Huang and Valtorta 2006) that $Q[\mathbf{E}]$ is identifiable if and only if it is identifiable from $Q[C_i]$ by the algorithm IDENTIFY. If no C_i contains $\mathbf{E}$, we move the product over the recoverable C_i to the l.h.s and reduce the problem to recover $Q[\mathbf{E}]$ in $\mathcal{G}_{\mathbf{V}\setminus\mathbf{C}}$ from $f(P(\mathbf{v} \mid S{=}1)/\prod_i Q[C_i])$. If there is a component B_i that does not contain $\mathbf{E}$, we can sum out of $\mathcal{P}$, and remove from $\mathcal{G}$, the non-ancestors of $B_i \cup \{S\}$. Then try to recover $Q[B_i]$ with RCE, if successful; we can reduce the problem to that of recovering $Q[\mathbf{E}]$ in $\mathcal{G}_{\mathbf{V}\setminus B_i}$ from $f(P(\mathbf{v} \mid S{=}1)/Q[B_i])$. □

Theorem 4. *The procedure* IDSB *is sound.*

Proof. Line 1 of IDSB first compute the set $\mathbf{W}$ of variables that are relevant for the identification of $P_\mathbf{x}(\mathbf{y})$. Note that every variable in $\mathbf{V} \setminus \mathbf{W}$ can be summed out of P and removed from the graph $\mathcal{G}$ without compromising recoverability. Next, lines 2 and 3 recover a distribution over a set $\mathbf{T}$ from the given distribution $P(\mathbf{t^0})$ using lemma 1. Then, in line 4 it translates the problem of identifying $P_\mathbf{x}(\mathbf{y})$ to that of identifying the c-factors $Q_B[C_1], \ldots, Q_B[C_\ell]$ associated with the c-components of $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B})$ (see Eq. 10). In line 5, for each $Q_B[C_j]$ we determine whether $Q_B[C_j]$ is equal to the c-factor $Q_R[C_j]$ (in the context of $PJ(\mathcal{G}, \mathbf{T})$) that is possibly recoverable from $P(\mathbf{t})$, according to lemma 3. If $Q_B[C_j]$ is not equal to $Q_R[C_j]$ or $Q_R[C_j]$ is not identifiable from $P(\mathbf{t})$ alone, IDSB tries to recover the collection of factors $Q[D_i]$, that together give us $Q_B[C_j]$ as in Eq. 12, using RCE. □

Lemma 12. *If* IDSB *fails to recover* $P_\mathbf{x}(\mathbf{y})$ *from* $P(\mathbf{v}|S{=}1)$ *and* $P(\mathbf{t})$ *then at least one of the conditions of the Generalized Adjustment Criterion fail.*

Proof. First, let us determine the conditions under which IDSB will make use of RCE to recover some $Q[D_i]$, which is where the FAIL condition is first produced. Let C_j be a c-component of $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B})$. It follows that $C_j \subseteq \mathbf{D}$ and

$$Pa(C_j)_{\mathcal{G}_\mathbf{D}} = Pa(C_j)_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}} = Pa(C_j) \setminus \mathbf{X}. \quad (88)$$

By definition of $\mathbf{D} = An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}$ is an ancestral set in $\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}$, hence

$$\begin{aligned} Pa(C_j)_{\mathcal{G}_\mathbf{D}} &\subseteq \mathbf{D} \\ \Leftrightarrow Pa(C_j) \setminus \mathbf{X} &\subseteq \mathbf{D} \\ \Leftrightarrow (Pa(C_j) \setminus \mathbf{X}) \setminus \mathbf{Y} &\subseteq \mathbf{D} \setminus \mathbf{Y} \\ \Leftrightarrow Pa(C_j) \setminus (\mathbf{X} \cup \mathbf{Y}) &\subseteq \mathbf{D} \setminus \mathbf{Y} \\ \Rightarrow (\mathbf{R} \cap Pa(C_j)) \setminus (\mathbf{X} \cup \mathbf{Y}) &\subseteq \mathbf{D} \setminus \mathbf{Y} \\ \Rightarrow [(\mathbf{R} \cap Pa(C_j)) \setminus (\mathbf{X} \cup \mathbf{Y})] \setminus (\mathbf{D} \setminus \mathbf{Y}) &= \emptyset \end{aligned} \tag{89}$$

Lemma 3 states that the c-factors $Q_B[C_j]$ and $Q_R[C_j]$ are equal if $\mathbf{B} \cap Pa(C_j) = \mathbf{R} \cap Pa(C_j)$.

$$\begin{aligned} \mathbf{B} \cap Pa(C_j) &= \mathbf{R} \cap Pa(C_j) \\ \Leftrightarrow (\mathbf{D} \setminus \mathbf{Y}) \cap \mathbf{R} \cap Pa(C_j) &= \mathbf{R} \cap Pa(C_j) \\ \Leftrightarrow (\mathbf{R} \cap Pa(C_j)) \setminus (\mathbf{D} \setminus \mathbf{Y}) &= \emptyset \end{aligned} \tag{90}$$

Let $\mathbf{E} = Pa(C_j) \cap \mathbf{R} \cap (\mathbf{X} \cup \mathbf{Y})$, we want to show that condition (90) holds if and only if if $\mathbf{E} = \emptyset$. When this is the case, (89) will imply equality (90). If $\mathbf{E} \neq \emptyset$, then there exists $V_i \in Pa(C_j) \cap \mathbf{R} \cap (\mathbf{X} \cup \mathbf{Y})$. It is clear that $\mathbf{D} \setminus \mathbf{Y}$ does not contain any variable in $(\mathbf{X} \cup \mathbf{Y})$, therefore condition (90) will not be satisfied.

According to this we can conclude that IDSB will invoke RCE(D_i) with $D_i \in F_j$ when

(I) $Pa(C_j) \cap \mathbf{R} \cap (\mathbf{X} \cup \mathbf{Y}) = \emptyset$ and $Q_R[C_j]$ is not identifiable from $Q_R[T_k]$

(II) $Pa(C_j) \cap \mathbf{R} \cap (\mathbf{X} \cup \mathbf{Y}) \neq \emptyset$

Claim 1. *If $Q_R[C_j]$ is not identifiable from $Q_R[T_k]$ it implies that C_j contains an element in $De(\mathbf{X}) \cap An(\mathbf{Y}) \cap \mathbf{T}$*

Proof. Let $\mathcal{H}$ denote $PJ(\mathcal{G}, \mathbf{T})$. Since $\mathcal{H}_{C_j}$ and $\mathcal{H}_{T_k}$ have a single c-component, Thm. 4 in (Huang and Valtorta 2008) implies that $Q_R[C_j]$ is not identifiable from $Q_R[T_k]$ only if there exists a set $\mathbf{N}$ such that $C_j \subset \mathbf{N} \subseteq T_k$, $\mathcal{H}_\mathbf{N}$ has a single c-component and $\mathbf{N} \setminus C_j \subseteq An(C_j)_{\mathcal{H}_\mathbf{N}}$.

We claim that $(\mathbf{N} \setminus C_j) \cap \mathbf{X} \neq \emptyset$, to witness assume this is not the case. Then, $\mathbf{N} \setminus C_j \subseteq An(C_j)_{\mathcal{H}_{\mathbf{N}\setminus\mathbf{X}}} \subseteq An(C_j)_{\mathcal{H}_{\mathbf{T}\setminus\mathbf{X}}} \subseteq An(C_j)_{\mathcal{H}_{\mathbf{D}\cap\mathbf{T}}}$. Note that $\mathcal{H}_{\mathbf{D}\cap\mathbf{T}}$ has the same or less nodes that $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B}) = PJ(\mathcal{G}_\mathbf{D}, (\mathbf{D} \cap \mathbf{T}) \cup \mathbf{Y})$ and possibly less edges. Therefore, $An(C_j)_{\mathcal{H}_{\mathbf{D}\cap\mathbf{T}}}$ is a subset of $An(C_j)$ in $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B})$. This implies that all elements in $\mathbf{N} \setminus C_j$ are present in $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B})$ the same as C_j, hence $\mathbf{N} \neq C_j$ would be a c-component of $PJ(\mathcal{G}_\mathbf{D}, \mathbf{D} \setminus \mathbf{B})$ instead of C_j, a contradiction. We conclude that $\mathbf{N} \setminus C_j$ contains some $X' \in \mathbf{X}$ which is an ancestor of some variable in C_j. Since all variables in $\mathbf{N}$ are also in $\mathbf{T} \cap An(C_j) \subseteq \mathbf{T} \cap An(\mathbf{Y})$ the statement is proved. □

In summary when RCE(D_i) is invoked by IDSB, D_i belongs to the set F_j such that one of the following holds:

(a) C_j contains an element in $De(\mathbf{X}) \cap An(\mathbf{Y}) \cap \mathbf{T}$, or

(b) C_j contains an element in $Ch((\mathbf{X} \cup \mathbf{Y}) \cap \mathbf{R})$

In one hand, RCE may fail due to the call to IDENTIFY in line 3. We know that $Q[\mathbf{E}]$ is identifiable if and only if it is identifiable from $Q[C_i]$. Hence, if this fails the causal effect is not identifiable, much less recoverable. In particular no estimable adjustment expression will be equal to the effect either.

On the other hand, RCE could fail because it is unable to recover $Q[\mathbf{E}]$, we will show that in this case, at least one of the conditions of GAC fails too. First note that $C_j, D_i \subset \mathbf{D}$, equivalently, $C_j, D_i \in An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}$. RCE failing to recover $Q[\mathbf{E}] = Q[D_i]$ implies that there exists $\mathbf{W} \subset An(\mathbf{D}_i \cup \{S\}), \mathbf{D}_i \subseteq \mathbf{W}$, such that every c-component $W_1, \ldots, W_p, W_S$ of $\mathcal{G}_{\mathbf{W}\cup\{S\}}$ contains a variable W' that is both an ancestor of W_S and a child of $\mathbf{R}$, where W_S is the c-component to which S belongs.

Claim 2. *For any $A \in W_s$ there exists an active path q from S to A given $\mathbf{T}$ ending with a bidirected edge; or the adjustment criterion fails.*

Proof. Let us prove it by induction in the length of q.

(Base case) When $|q| = 0$ the path is from S to S which is always active.

(Inductive Step) Suppose the path q_k with $|q_k| = k$ from S to A_k is active given $\mathbf{T}$. Add a variable A_{k+1} that is connected to A_k with a bidirected arrow. For A_k to be in $\mathbf{W}$ it must be an ancestor of $\mathbf{Y}$ or S. If it is an ancestor of $\mathbf{Y}$, for the third condition of the adjustment to hold, there must exists some variable in $De(A_k) \cap An(\mathbf{Y}) \cap \mathbf{T}$ that makes $\mathbf{Y}$ independent of S, and observing that variable makes A_k an active collider. Then A_{k+1} is connected to S with a path q_{k+1} active given $\mathbf{T}$ and ending with a bidirected into A_{k+1}. If A_k is in turn an ancestor of S, then either the directed path between A_k and S is open and so appending the bidirected to A_{k+1} completes a path q_{k+1}, or there exists $De(A_k) \cap \mathbf{T}$ that blocks the directed path but makes A_k an active collider and the same conclusion follows. □

Since $W' \in Ch(\mathbf{R})$, it is the case that there exists an active path q between S and W' when $\mathbf{T}$ is observed. Note that if the path happens to have some $T^* \in \mathbf{T}$ as a collider, not observing it opens a path from S to $\mathbf{Y}$ if $T^* \in An(\mathbf{Y})$, or that q remains open if $T^* \in An(S)$. If any descendant of T^* is observed to block any of those alternatives, then T^* becomes an active collider. Therefore, the path q cannot be blocked by observing or not variables in $\mathbf{T}$.

Consider the path in $\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}$ that witnesses that W' is an ancestor of $Y' \in \mathbf{Y}$, for the Adjustment Criterion to hold there must a variable in $De(W') \cap An(Y') \cap \mathbf{T}$ that should be used as a covariate to block the path from S to Y' passing through W'. However, by observing such variable W' becomes an active collider and q can be extended to some other variable in $\mathbf{W}$. Let W' be now that next variable, using the same reasoning we have that in order to avoid a path from S to $\mathbf{Y}$, the path q keeps extending to other variables in $\mathbf{W}$. At some point this reasoning extends to the variables in C_j, since at least some variable in $C_j \setminus D_i$ has to share a common ancestor in $\mathbf{R}$ with a variable in D_i which had been assigned to F_j. When the path q reaches a variable $W^* \in Ch((\mathbf{X} \cup \mathbf{Y}) \cap \mathbf{R}) \cup De(\mathbf{X}) \cap An(\mathbf{Y})$ (guaranteed by ((a)) and ((b))) that, as shown before, is also in $An(\mathbf{Y})_{\mathcal{G}_{\mathbf{V}\setminus\mathbf{X}}}$, we conclude that W^* needs to be a covariate in $\mathbf{Z}^{\mathbf{T}}$ for the third condition of the adjustment to hold, but this contradicts with the first condition that forbids variables in a proper causal path to be used as covariates. Therefore, there is no set of covariates satisfying the criterion and adjustment fails. □

Based on this result we can prove that instances that IDSB can recover is a superset of those recoverable by the generalized adjustment in (Correa, Tian, and Bareinboim 2018a), as the following result states.

Theorem 5. IDSB *is strictly more powerful than the Generalized Adjustment Criterion for the task of recovering a causal effect $P_{\mathbf{x}}(\mathbf{y})$ from a combination of biased distribution $P(\mathbf{v}|S{=}1)$ and unbiased distribution $P(\mathbf{t}^{\mathbf{0}})$ in $\mathcal{G}$.*

Proof. From Lemma 12 we have that whenever IDSB fails to recover the target effect, then GAC also fails to recover the effect from the same input. This implies that IDSB subsumes GAC. To show that IDSB is more general, we consider the problem of recovering $P_{\mathbf{x}}(\mathbf{y})$ from the model in Fig. 2(b) with external data over $\mathbf{T}^{\mathbf{0}} = \{W_2\}$. The conditions for the Generalized Adjustment Criterion are not satisfied, hence there is no adjustment expression that can be used to recover that effect. Nevertheless, IDSB recovers the effect with a different mapping (see main paper for details on this counter-example). □