Conditional Adjustment in a Markov Equivalence Class

Sara LaPlante
University of Washington

Emilija Perković
University of Washington

Abstract

We consider the problem of identifying a conditional causal effect through covariate adjustment. We focus on the setting where the causal graph is known up to one of two types of graphs: a maximally oriented partially directed acyclic graph (MPDAG) or a partial ancestral graph (PAG). Both MPDAGs and PAGs represent equivalence classes of possible underlying causal models. After defining adjustment sets in this setting, we provide a necessary and sufficient graphical criterion – the *conditional adjustment criterion* – for finding these sets under conditioning on variables unaffected by treatment. We further provide explicit sets from the graph that satisfy the conditional adjustment criterion, and therefore, can be used as adjustment sets for conditional causal effect identification.

1 INTRODUCTION

Many scientific disciplines have an interest in identifying and estimating causal effects for specific subgroups of a population. For instance, researchers may want to know if a medical treatment is beneficial for people with heart disease or if the treatment will harm older patients (Brand and Xie, 2010; Health, 2010). Such causal effects are referred to as *conditional causal effects* or *heterogeneous causal effects*. The identification of these conditional causal effects from observational data is the subject of this work.

Much of the literature on estimating conditional causal effects from observational data focuses on the conditional average treatment effect (CATE; Athey and Imbens, 2016; Wager and Athey, 2018; Künzel et al., 2019; Nie and Wager, 2021; Kennedy et al., 2022). The CATE is represented as a contrast of means for a response Y under different do-interventions (see Section 2 for definition) of a treatment X when conditioning on a set of covariate values $\mathbf{z}$. These means take the form $\mathbb{E}[Y|do(X=x), \mathbf{Z}=\mathbf{z}]$.

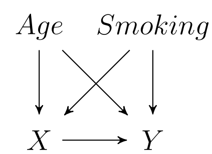

Figure 1: A causal DAG used in Section 1.

Some results on CATE estimation assume that the conditioning set $\mathbf{Z}$ is rich enough to capture all relevant common causes of X and Y – meaning that X and Y are *unconfounded* given $\mathbf{Z}$. This implies

$$\mathbb{E}[Y|do(X=x), \mathbf{Z}=\mathbf{z}] = \mathbb{E}[Y|X=x, \mathbf{Z}=\mathbf{z}], \quad (1)$$

which allows the CATE to be estimated as a difference of means from observational data.

However, this assumption does not hold in all applications. Consider, for example, the setting depicted in the causal directed acyclic graph (DAG) of Figure 1, where we want to compute a causal effect of X on Y given some set $\mathbf{Z}$. In this setting, age and smoking status are common causes of X and Y, and therefore, X and Y are confounded unless we condition on both age and smoking status ($\mathbf{Z} = \{Age, Smoking\}$). But we may want to know the causal effect of X on Y conditional on age alone ($\mathbf{Z} = \{Age\}$).

To allow for estimation of the CATE in such cases, various recent works (Abrevaya et al., 2015; Fan et al., 2022; Chernozhukov et al., 2023; Smucler et al., 2020) have proposed estimation methods that rely on knowing an additional set of covariates $\mathbf{S}$ that – together with $\mathbf{Z}$ – leads to X and Y being unconfounded. We refer to this set of variables as a *conditional adjustment set* (Definition 1). For such a set $\mathbf{S}$,

$$\begin{aligned} &\mathbb{E}[Y|do(X=x), \mathbf{Z}=\mathbf{z}] \quad (2)\\ &\quad = \mathbb{E}_{\mathbf{S}}\Big[\mathbb{E}[Y|X=x, \mathbf{Z}, \mathbf{S}] \;\Big|\; \mathbf{Z}=\mathbf{z}\Big]. \end{aligned}$$

Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

In the example above, if $\mathbf{Z} = \{Smoking\}$, then $\mathbf{S} = \{Age\}$.

Of course, not all conditional causal effect research focuses on estimation through the functional in Equation (2). Notably, other work has explored identifiability without limiting focus to a particular functional. For example, Shpitser and Pearl (2008) and Jaber et al. (2019, 2022) focus on the conditions under which the interventional distribution $f(\mathbf{y}|do(\mathbf{x}), \mathbf{z})$ is identifiable given a causal graph. Though these results broaden the options for identification, estimators based on these results would have to rely on functionals that may prove difficult to estimate, such as $\frac{f(\mathbf{y},\mathbf{z}|do(\mathbf{x}))}{f(\mathbf{z}|do(\mathbf{x}))}$ (Shpitser and Pearl, 2008; Jaber et al., 2019, 2022). Our work addresses this by focusing on identification of the same interventional distribution given a causal graph – but through the use of conditional adjustment sets, which may lead to more desirable estimators. To the best of our knowledge, this area of research is largely unexplored.

Our main contribution is the *conditional adjustment criterion* (Definitions 2 and 7), a graphical criterion that we show is necessary and sufficient for identifying a conditional adjustment set (Theorems 3 and 9). We additionally provide explicit sets that satisfy this criterion when any such set exists. We note, however, that these results are restricted to a setting where the conditioning set $\mathbf{Z}$ consists of variables known to be unaffected by treatment. While this restricted setting produces limitations (see the second example in the discussion, Section 5), our results are broadly applicable to a variety of research questions. For example, the restriction is met when the conditioning set includes exclusively pre-treatment variables.

In considering the problem of identifying a conditional adjustment set, we assume that the underlying causal system can be represented by a causal DAG. When we collect observational data on all variables in the system, we can attempt to learn this causal DAG by relying on the constraints present in the data (Spirtes et al., 1999; Chickering, 2002; Zhang, 2008b; Hauser and Bühlmann, 2012; Mooij et al., 2020; Squires and Uhler, 2022). However, this task is often impossible from observational data alone, regardless of the available sample size. And further, we cannot always observe every variable.

Thus, our work focuses on causal models that represent *Markov equivalence classes* of graphs that can be learned from observational data: a maximally oriented partially directed acyclic graph (MPDAG) (Meek, 1995) and a maximally oriented partial ancestral graph (PAG) (Richardson and Spirtes, 2002). An MPDAG represents a restriction of the Markov equivalence class of DAGs that can be learned from observational data and background knowledge when all variables are observed (Andersson et al., 1997; Meek, 1995; Chickering, 2002). A PAG represents a Markov equivalence class of maximal ancestral graphs (MAGs) (Richardson and Spirtes, 2002), which can be learned from observational data and which allows for unobserved variables (Spirtes et al., 2000; Zhang, 2008b; Ali et al., 2009). A MAG, in turn, can be seen as a marginalization of a DAG containing only the observed variables (Richardson and Spirtes, 2002). See Section 2 and Supp. A for further definitions.

The structure of this paper is as follows: Section 2 provides preliminary definitions, with the remaining definitions given in Supp. A. Section 3 contains all results for the MPDAG setting. In particular, we introduce our conditional adjustment criterion in Section 3.1; Section 3.2 illustrates applications of our criterion with examples; Section 3.3 provides several methods for constructing conditional adjustment sets; and Section 3.4 includes a discussion of the similarities of our conditional adjustment criterion with both the adjustment criterion of Perković et al. (2017) and the **Z**-dependent dynamic adjustment criterion of Smucler et al. (2020). We present some analogous results for PAGs in Section 4, and we discuss some limitations of our results and areas for future work in Section 5.

2 PRELIMINARIES

We use capital letters (e.g. X) to denote nodes in a graph as well as random variables that these nodes represent. Similarly, bold capital letters (e.g. $\mathbf{X}$) are used to denote node sets and random vectors.

Nodes, Edges, and Subgraphs. A graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ consists of a set of nodes (variables) $\mathbf{V} = \{V_1, \dots, V_p\}, p \geq 1$, and a set of edges $\mathbf{E}$. Edges can be *directed* ($\rightarrow$), *bi-directed* ($\leftrightarrow$), *undirected* ($\circ\!\!-\!\!\circ$ or $-$), or *partially directed* ($\circ\!\!\rightarrow$). We use $\bullet$ as a stand in for any of the allowed edge marks. An edge is *into* (*out of*) a node X if the edge has an arrowhead (tail) at X. An *induced subgraph* $\mathcal{G}_{\mathbf{V}'} = (\mathbf{V}', \mathbf{E}')$ of $\mathcal{G}$ consists of $\mathbf{V}' \subseteq \mathbf{V}$ and $\mathbf{E}' \subseteq \mathbf{E}$ where $\mathbf{E}'$ are all edges in $\mathbf{E}$ between nodes in $\mathbf{V}'$.

Directed and Partially Directed Graphs. A *directed graph* contains only directed edges ($\rightarrow$). A *partially directed graph* may contain undirected edges ($-$) and directed edges ($\rightarrow$).

Mixed and Partially Directed Mixed Graphs. A *mixed graph* may contain directed and bi-directed edges. The *partially directed mixed graphs* we consider can contain any of the following edge types: $\circ\!\!-\!\!\circ$, $\circ\!\!\rightarrow$, $\rightarrow$, and $\leftrightarrow$. Hence, an edge $\bullet\!\!\rightarrow$ in a partially directed graph can only refer to edge $\rightarrow$, whereas in a partially directed mixed graph, $\bullet\!\!\rightarrow$ can represent $\rightarrow$, $\leftrightarrow$, or $\circ\!\!\rightarrow$.

Paths and Cycles. For disjoint node sets $\mathbf{X}$ and $\mathbf{Y}$, a path from $\mathbf{X}$ to $\mathbf{Y}$ is a sequence of distinct nodes $\langle X, \dots, Y \rangle$ from some $X \in \mathbf{X}$ to some $Y \in \mathbf{Y}$ for which every pair of successive nodes is adjacent. A path consisting of undirected edges ($-$ or $\circ\!\!-\!\!\circ$) is an *undirected path*. A *directed path* from X to Y is a path of the form $X \rightarrow \cdots \rightarrow Y$. A directed path from X to Y and the edge $Y \rightarrow X$ form a *directed cycle*. A directed path from X to Y and the edge $X \rightarrow Y$ form an *almost directed cycle*. A path $\langle V_1, \dots, V_k \rangle$, $k > 1$, in a graph $\mathcal{G}$ is a *possibly directed path* if no edge $V_i \leftarrow\!\!\bullet V_j, 1 \leq i < j \leq k$, is in $\mathcal{G}$ (Perković et al., 2017, Zhang, 2008a).

A path from $\mathbf{X}$ to $\mathbf{Y}$ is *proper* (w.r.t. $\mathbf{X}$) if only its first node is in $\mathbf{X}$. A path from X to Y is a *back-door path* if does not begin with a visible edge out of X (see definition of *visible* below; Pearl, 2009, Maathuis and Colombo, 2015). For a path $p = \langle X_1, X_2, \dots, X_k \rangle$ and i, j, k such that $1 \leq i < j \leq k$, we define the *subpath* of p from X_i to X_j as the path $p(X_i, X_j) = \langle X_i, X_{i+1}, \dots, X_j \rangle$.

Colliders, Shields, and Definite Status Paths. If a path p contains $X_i \bullet\!\!\rightarrow X_j \leftarrow\!\!\bullet X_k$ as a subpath, then X_j is a *collider* on p. A path $\langle X_i, X_j, X_k \rangle$ is an *unshielded triple* if X_i and X_k are not adjacent. A path is *unshielded* if all successive triples on the path are unshielded. A node X_j is a *definite non-collider* on a path p if the edge $X_i \leftarrow X_j$ or $X_j \rightarrow X_k$ is on p, or if $\langle X_i, X_j, X_k \rangle$ is an undirected subpath of p and X_i is not adjacent to X_k. A node is of *definite status* on a path if it is a collider, a definite non-collider, or an endpoint on the path. A path p is of definite status if every node on p is of definite status.

Blocking, D-separation, and M-separation. Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a directed or partially directed graph $\mathcal{G}$. A definite-status path p from $\mathbf{X}$ to $\mathbf{Y}$ is *d-connecting* given $\mathbf{Z}$ if every definite non-collider on p is not in $\mathbf{Z}$ and every collider on p has a descendant in $\mathbf{Z}$. Otherwise, $\mathbf{Z}$ *blocks* p. If $\mathbf{Z}$ blocks all definite status paths between $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}$, then $\mathbf{X}$ is *d-separated* from $\mathbf{Y}$ given $\mathbf{Z}$ in $\mathcal{G}$ and we write $(\mathbf{X} \perp_d \mathbf{Y} | \mathbf{Z})_{\mathcal{G}}$ (Pearl, 2009).

If $\mathcal{G}$ is a mixed or partially directed mixed graph, the analogous terms to d-connection and d-separation are called *m-connection* and *m-separation* (Richardson and Spirtes, 2002). If a path is not m-connecting in such a graph $\mathcal{G}$ we will also call it blocked. We will also use the same notation $\perp_d$ to denote m-separation in a mixed or partially directed mixed graph $\mathcal{G}$.

Ancestral Relationships. If $X \rightarrow Y$, then X is a *parent* of Y. If $X - Y$, $X \circ\!\!-\!\!\circ Y$, $X \circ\!\!\rightarrow Y$, or $X \rightarrow Y$, then X is a *possible parent* of Y. If there is a directed path from X to Y, such as $X \rightarrow M_1 \rightarrow \cdots \rightarrow M_k$, $M_k = Y$, $k \geq 1$, then X is an *ancestor* of Y, Y is a *descendant* of X, and $M_1, \dots, M_k$ are *mediators* for X and Y. We use the convention that if Y is a descendant of X, then Y is also a mediator for X and Y. If there is a possibly directed path from X to Y, then X is a *possible ancestor* of Y, Y is a *possible descendant* of X, and any node on this path that is not X is a *possible mediator* of X and Y. We use the convention that if Y is a possible descendant of X, then Y is also a possible mediator for X and Y. We also use the convention that every node is an ancestor, descendant, possible ancestor, and possible descendant of itself. The sets of parents, possible parents, ancestors, descendants, possible ancestors, and possible descendants of X in $\mathcal{G}$ are denoted by $\mathrm{Pa}(X, \mathcal{G})$, $\mathrm{PossPa}(X, \mathcal{G})$, $\mathrm{An}(X, \mathcal{G})$, $\mathrm{De}(X, \mathcal{G})$, $\mathrm{PossAn}(X, \mathcal{G})$, and $\mathrm{PossDe}(X, \mathcal{G})$, respectively. Similarly, we denote the sets of mediators and possible mediators for X and Y in $\mathcal{G}$ by $\mathrm{Med}(X, Y, \mathcal{G})$ and $\mathrm{PossMed}(X, Y, \mathcal{G})$.

We let $\mathrm{An}(\mathbf{X}, \mathcal{G}) = \cup_{X \in \mathbf{X}} \mathrm{An}(X, \mathcal{G})$, with analogous definitions for $\mathrm{De}(\mathbf{X}, \mathcal{G})$, $\mathrm{PossAn}(\mathbf{X}, \mathcal{G})$, and $\mathrm{PossDe}(\mathbf{X}, \mathcal{G})$. For disjoint node sets $\mathbf{X}$ and $\mathbf{Y}$, we let $\mathrm{Med}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ be the union of all mediators of $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ that lie on a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$, with an analogous definition for $\mathrm{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$. Unconventionally, we define $\mathrm{Pa}(\mathbf{X}, \mathcal{G}) = (\cup_{X \in \mathbf{X}} \mathrm{Pa}(X, \mathcal{G})) \setminus \mathbf{X}$. We denote that X is adjacent to Y in $\mathcal{G}$ by $X \in \mathrm{Adj}(Y, \mathcal{G})$.

DAGs and PDAGs. A directed graph without directed cycles is a *directed acyclic graph* (DAG). A *partially directed acyclic graph* (PDAG) is a partially directed graph without directed cycles.

MAGs. A mixed graph without directed or almost directed cycles is called *ancestral*. Note that we do not consider ancestral graphs that represent selection bias (see Zhang, 2008a, for details). A *maximal ancestral graph* (MAG) is an ancestral graph $\mathcal{M} = (\mathbf{V}, \mathbf{E})$ where every pair of non-adjacent nodes X and Y in $\mathcal{M}$ can be m-separated by a set $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$. A DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ with unobserved variables $\mathbf{U} \subseteq \mathbf{V}$ can be uniquely *represented by* a MAG $\mathcal{M} = (\mathbf{V} \setminus \mathbf{U}, \mathbf{E}')$, which preserves the ancestry and m-separations among the observed variables (Richardson and Spirtes, 2002).

MPDAGs and Markov Equivalence. All DAGs over a node set $\mathbf{V}$ with the same adjacencies and unshielded colliders can be uniquely *represented by* a *completed PDAG* (CPDAG). These DAGs form a Markov equivalence class with the same set of d-separations. A *maximally oriented PDAG* (MPDAG) is formed by taking a CPDAG, adding background knowledge (by directing undirected edges), and completing Meek (1995)'s orientation rules. We say a DAG is *represented by* an MPDAG $\mathcal{G}$ if it has the same nodes, adjacencies, and directed edges as $\mathcal{G}$. The set of such DAGs – denoted by $[\mathcal{G}]$ – forms a restriction of the Markov equivalence class so that all DAGs in $[\mathcal{G}]$ have same set of d-separations. Note that if $\mathcal{G}$ has the edge $A - B$, then $[\mathcal{G}]$ contains at least one DAG with $A \to B$ and one DAG with $A \leftarrow B$ (Meek, 1995). Further, note that all DAGs and CPDAGs are MPDAGs.

PAGs and Markov Equivalence. All MAGs that encode the same set of m-separations form a Markov equivalence class, which can be uniquely *represented by* a *partial ancestral graph* (PAG; Richardson and Spirtes, 2002; Ali et al., 2009). $[\mathcal{G}]$ denotes all MAGs represented by a PAG $\mathcal{G}$. We say a DAG $\mathcal{D}$ is *represented by* a PAG $\mathcal{G}$ if there is a MAG $\mathcal{M} \in [\mathcal{G}]$ such that $\mathcal{D}$ is represented by $\mathcal{M}$.

We do not consider PAGs that represent selection bias (see Zhang, 2008b). Further, we only consider maximally informative PAGs (Zhang, 2008b). That is, if a PAG $\mathcal{G}$ has the edge $A \bullet\!\!-\!\!\circ B$, then $[\mathcal{G}]$ contains a MAG with $A \bullet\!\!\to B$ and a MAG with $A \leftarrow B$. (We preclude MAGs with $A - B$ by assuming no selection bias.) Any arrowhead or tail edge mark in a PAG $\mathcal{G}$ corresponds to that same arrowhead or tail edge mark in every MAG in $[\mathcal{G}]$. The edge orientations in every PAG we consider are completed with respect to orientation rules $R1 - R4$ and $R8 - R10$ of Zhang (2008b).

Visible and Invisible Edges. Given a MAG or PAG $\mathcal{G}$, a directed edge $X \to Y$ is *visible* in $\mathcal{G}$ if there is a node $V \notin \mathrm{Adj}(Y, \mathcal{G})$ such that $\mathcal{G}$ contains either $V \bullet\!\!\to X$ or $V \bullet\!\!\to V_1 \leftrightarrow \cdots \leftrightarrow V_k \leftrightarrow X$, where $k \geq 1$ and $V_1, \ldots, V_k \in \mathrm{Pa}(Y, \mathcal{G}) \setminus \{V, X, Y\}$ (Zhang, 2006). A directed edge that is not visible in a MAG or PAG is said to be *invisible.*

Markov Compatibility and Positivity. An *observational density* $f(\mathbf{v})$ is *Markov compatible* with a DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ if $f(\mathbf{v}) = \prod_{V_i \in \mathbf{V}} f(v_i | \operatorname{pa}(v_i, \mathcal{D}))$. If $f(\mathbf{v})$ is Markov compatible with a DAG $\mathcal{D}$, then it is Markov compatible with every DAG that is Markov equivalent to $\mathcal{D}$ (Pearl, 2009). Hence, we say that a density is *Markov compatible* with an MPDAG, MAG, or PAG $\mathcal{G}$ if it is Markov compatible with a DAG represented by $\mathcal{G}$. Throughout, we assume positivity. That is, we only consider distributions that satisfy $f(\mathbf{v}) > 0$ for all valid values of $\mathbf{V}$ (Kivva et al., 2023).

Probabilistic Implications of Graph Separation. Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a DAG, MPDAG, MAG, or PAG $\mathcal{G}$. If $\mathbf{X}$ and $\mathbf{Y}$ are d-separated or m-separated given $\mathbf{Z}$ in $\mathcal{G}$, then $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$ in any *observational density* that is Markov compatible with $\mathcal{G}$ (Lauritzen et al., 1990; Zhang, 2008a; Henckel et al., 2022).

Causal Graphs. Let $\mathcal{G}$ be a graph with nodes V_i and V_j. When $\mathcal{G}$ is an MPDAG, it is a *causal MPDAG* if every edge $V_i \to V_j$ represents a direct causal effect of V_i on V_j and if every edge $V_i - V_j$ represents a direct causal effect of unknown direction (either V_i affects V_j or V_j affects V_i). Note that all DAGs are MPDAGs.

When $\mathcal{G}$ is a MAG or PAG, it is a *causal MAG* or *causal PAG*, respectively, if every edge $V_i \to V_j$ represents the presence of a causal path from V_i to V_j; every edge $V_i \leftarrow\!\!\bullet V_j$ represents the absence of a causal path from V_i to V_j; and every edge $V_i \circ\!\!-\!\!\circ V_j$ represents the presence of a causal path of unknown direction or a common cause in the underlying causal DAG.

Causal and Non-causal Paths. Note that any directed or possibly directed path in a causal graph is *causal* or *possibly causal*, respectively. However, since we focus on causal graphs, we will use this causal terminology for paths in any of our graphs. We will say a path is *non-causal* if it is not possibly causal.

Consistency. Let $f(\mathbf{v})$ be an observational density over $\mathbf{V}$. The notation $do(\mathbf{X} = \mathbf{x})$, or $do(\mathbf{x})$ for short, represents an outside intervention that sets $\mathbf{X} \subseteq \mathbf{V}$ to fixed values $\mathbf{x}$. An *interventional density* $f(\mathbf{v}|do(\mathbf{x}))$ is a density resulting from such an intervention.

Let $\mathbf{F}^*$ denote the set of all interventional densities $f(\mathbf{v}|do(\mathbf{x}))$ such that $\mathbf{X} \subseteq \mathbf{V}$ (including $\mathbf{X} = \emptyset$). A causal DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ is a *causal Bayesian network compatible with* $\mathbf{F}^*$ if and only if for all $f(\mathbf{v}|do(\mathbf{x})) \in \mathbf{F}^*$, the following *truncated factorization* holds:

$$f(\mathbf{v}|do(\mathbf{x})) = \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}} f(v_i | \operatorname{pa}(v_i, \mathcal{D})) \mathbb{1}(\mathbf{X} = \mathbf{x}) \quad (3)$$

(Pearl, 2009; Bareinboim et al., 2012). We say an interventional density is *consistent* with a causal DAG $\mathcal{D}$ if it belongs to a set of interventional densities $\mathbf{F}^*$ such that $\mathcal{D}$ is compatible with $\mathbf{F}^*$. Note that any observational density that is Markov compatible with $\mathcal{D}$ is consistent with $\mathcal{D}$. We say an interventional density is *consistent* with a causal MPDAG, MAG, or PAG $\mathcal{G}$ if it is consistent with each DAG represented by $\mathcal{G}$ – were the DAG to be causal.

Identifiability. Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a causal MPDAG or PAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, and let $\mathbf{F_i^*} = \{f_i(\mathbf{v}|do(\mathbf{x'})) : \mathbf{X'} \subseteq \mathbf{V}\}$ be a set with which a DAG $\mathcal{D}_i$ represented by $\mathcal{G}$ is compatible – were $\mathcal{D}_i$ to be causal. We say the conditional causal effect of $\mathbf{X}$ on $\mathbf{Y}$ given $\mathbf{Z}$ is *identifiable* in $\mathcal{G}$ if for any $\mathbf{F_1^*}, \mathbf{F_2^*}$ where $f_1(\mathbf{v}) = f_2(\mathbf{v})$, we have $f_1(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) = f_2(\mathbf{y}|do(\mathbf{x}), \mathbf{z})$ (Pearl, 2009).

Forbidden Set. Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in an MPDAG or PAG $\mathcal{G}$. Then the *forbidden set* relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{G}$ is

$$\operatorname{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \begin{Bmatrix} \text{nodes in } \operatorname{PossDe}(W, \mathcal{G}), \text{ where} \\ W \in \operatorname{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \end{Bmatrix}.$$

3 RESULTS - MPDAGS

In this section, we present our results on identifying a conditional causal effect via our conditional adjustment criterion in the setting of an MPDAG (Definition 2). Examples of how to use our criterion and explicit conditional adjustment sets based on our criterion follow these results. We remark here that our criterion shares similarities with the adjustment criterion for total effect identification of Perković et al. (2017) and with the $\mathbf{Z}$-dependent dynamic adjustment criterion of Smucler et al. (2020), but we save these results and reflections for Section 3.4.

Note that the results of this section hold when a fully oriented DAG is known, since all DAGs are MPDAGs. Throughout, our goal is to identify the conditional causal effect of treatments $\mathbf{X}$ on responses $\mathbf{Y}$ conditional on covariates $\mathbf{Z}$ and given a known graph $\mathcal{G}$.

3.1 Conditional Adjustment Criterion

We include our definition of a conditional adjustment set below (Definition 1). Note that, while this section focuses on MPDAGs, we write Definition 1 broadly for further use in Section 4. Our goal in this section is to find an equivalent graphical characterization of a conditional adjustment set. Theorem 3 establishes that Definition 2 provides such a graphical characterization, which we call the conditional adjustment criterion, under the assumption that the conditioning set does not contain variables affected by treatment ($\mathbf{Z} \cap \operatorname{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$).

Definition 1 (**Conditional Adjustment Set for MPDAGs, PAGs**) *Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{S}$ be pairwise disjoint node sets in a causal MPDAG or PAG $\mathcal{G}$. Then $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$ if for any density f consistent with $\mathcal{G}$*

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) = \begin{cases} f(\mathbf{y}|\mathbf{x}, \mathbf{z}) & \mathbf{S} = \emptyset \\ \int f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z})\, \mathrm{d}\mathbf{s} & \mathbf{S} \neq \emptyset. \end{cases} \tag{4}$$

Definition 2 (**Conditional Adjustment Criterion for MPDAGs**) *Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{S}$ be pairwise disjoint node sets in an MPDAG $\mathcal{G}$, where $\mathbf{Z} \cap \operatorname{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$ and where every proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ starts with a directed edge. Then $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$ if*

(a) *$\mathbf{S} \cap \operatorname{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$, and*

(b) *$\mathbf{S} \cup \mathbf{Z}$ blocks all proper non-causal definite status paths from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$.*

Theorem 3 (**Completeness, Soundness of Conditional Adjustment Criterion for MPDAGs**) *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, and $\mathbf{S}$ be pairwise disjoint node sets in a causal MPDAG $\mathcal{G}$, where $\mathbf{Z} \cap \operatorname{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$. Then $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$ (Definition 1) if and only if $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$ (Definition 2).*

Proof of Theorem 3. First note the following facts.

(i) Every proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ starts with a directed edge.

(ii) $\mathbf{Z} \cap \operatorname{De}(\mathbf{X}, \mathcal{D}) = \emptyset$ in every DAG $\mathcal{D}$ in $[\mathcal{G}]$.

(iii) $\mathbf{Z} \cap \operatorname{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$.

We have that (i) holds in either direction – by definition ($\Leftarrow$) or by Proposition 36 (Supp. C) ($\Rightarrow$). Then Lemmas 20 and 26 (Supp. B) imply (ii) and (iii), respectively, given $\mathbf{Z} \cap \operatorname{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$ and (i).

Now consider the following statements.

(a) $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$.

(b) $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in each DAG in $[\mathcal{G}]$ – were the DAG to be causal.

(c) $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in each DAG in $[\mathcal{G}]$.

(d) $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$.

By definition, (a) $\Leftrightarrow$ (b). Then (b) $\Leftrightarrow$ (c) by Theorems 39 and 40 (Supp. D) and the fact that the conditional adjustment criterion does not require a causal DAG. Lastly, by the facts above and by applying Lemmas 21 and 22 (Supp. B) in turn, (c) $\Leftrightarrow$ (d). ■

3.2 Examples

To illustrate the usefulness of the results above, we provide examples below where we aim to find $f(\mathbf{y}|do(\mathbf{x}), \mathbf{z})$ when $\mathbf{Z} \cap \operatorname{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$. Theorem 3 allows us to use the conditional adjustment criterion to (a) check whether a set can be used for conditional adjustment (Examples 1-3) or (b) determine if no such set exists (Example 4).

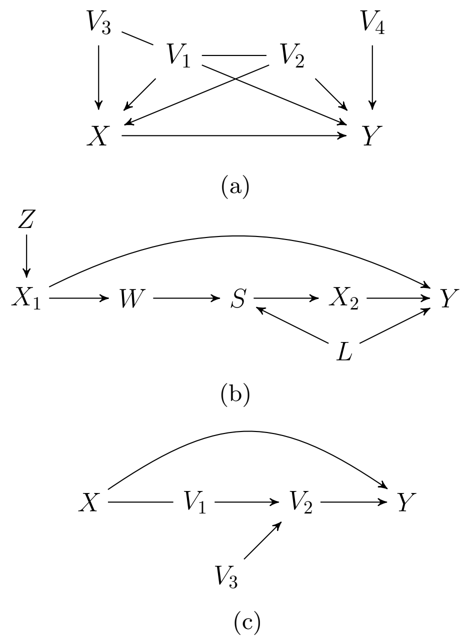

Figure 2: Causal MPDAGs used in Examples 1-4.

Example 1 (**Empty Conditional Adjustment Set.**) *Let $\mathcal{G}$ be the causal MPDAG in Figure 2a* [1]*, and let $\mathbf{X} = \{X\}$, $\mathbf{Y} = \{Y\}$, and $\mathbf{Z} = \{V_1, V_2\}$. Note that $\mathbf{Z} \cap \mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$ and that every possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ starts with a directed edge.*

Let $\mathbf{S} = \emptyset$. Note that $\mathbf{S} \cap (\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}) = \emptyset$, $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$, and $\mathbf{S} \cup \mathbf{Z}$ blocks all non-causal definite status paths from $\mathbf{X}$ to $\mathbf{Y}$. Thus, $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$, and by Theorem 3, $f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) = f(y|x, v_1, v_2)$.

Example 2 (**Only Nonempty Conditional Adjustment Sets.**) *Again let $\mathcal{G}$ be the causal MPDAG in Figure 2a, where $\mathbf{X} = \{X\}$ and $\mathbf{Y} = \{Y\}$. But now let $\mathbf{Z} = \{V_1\}$. We still have that $\mathbf{Z} \cap \mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$ and that every possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ starts with a directed edge.*

Note that if we let $\mathbf{S} = \emptyset$, $\mathbf{S} \cup \mathbf{Z}$ does not block the path $X \leftarrow V_2 \rightarrow Y$, which is a proper non-causal definite status path from $\mathbf{X}$ to $\mathbf{Y}$. Thus, the empty set is not a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$.

Consider, instead, the set $\mathbf{S} = \{V_2\}$. Note that $\mathbf{S} \cap (\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}) = \emptyset$, $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$, and $\mathbf{S} \cup \mathbf{Z}$ blocks all non-causal definite status paths from $\mathbf{X}$ to $\mathbf{Y}$. Thus, $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$, and by Theorem 3, $f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) = \int f(y|x, v_1, v_2) f(v_2|v_1)\,\mathrm{d}v_2$.

[1] Compare to Figure 5(a) of Perković (2020).

Example 3 (**Conditional Adjustment Set Contains Descendants of X.**) *Let $\mathcal{G}$ be the causal DAG (and therefore, MPDAG) in Figure 2b* [2]*, where we assume L is a variable that cannot be measured. Define $\mathbf{X} = \{X_1, X_2\}$, $\mathbf{Y} = \{Y\}$, and $\mathbf{Z} = \{Z\}$. Note that $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{G}) = \emptyset$.*

Consider the set $\mathbf{S} = \{S, W\}$. Note that $\mathbf{S} \cap (\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}) = \emptyset$, $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$, and $\mathbf{S}$ blocks all proper non-causal paths from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$. Hence, $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$, and by Theorem 3, $f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) = \int f(y|x_1, x_2, z, s, w) f(s, w|z)\,\mathrm{d}s\,\mathrm{d}w$.

Example 4 (**No Conditional Adjustment Set, Effect Non-identifiable.**) *Let $\mathcal{G}$ be the causal MPDAG in Figure 2c, and let $\mathbf{X} = \{X\}$, $\mathbf{Y} = \{Y\}$, and $\mathbf{Z} = \{V_3\}$. Note that $\mathbf{Z} \cap \mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$. However, $X - V_1 \rightarrow V_2 \rightarrow Y$ is a proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that starts with an undirected edge. Thus, by Theorem 3, there can be no conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$. In fact, by Proposition 36 (Supp. C), $f(\mathbf{y}|do(\mathbf{x}), \mathbf{z})$ is not identifiable in $\mathcal{G}$ using any method.*

3.3 Constructing Adjustment Sets

The conditional adjustment criterion provides a way to check if a set can be used for conditional adjustment given an MPDAG $\mathcal{G}$, but it does not provide a way to construct a conditional adjustment set – a task that may be difficult when $\mathcal{G}$ is large. The results in this section provide such a roadmap under certain assumptions. The proofs can be found in Supp. F.

Lemma 4 *Let $\mathbf{X} = \{X\}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a causal MPDAG $\mathcal{G}$, where $\mathbf{Z} \cap \mathrm{PossDe}(X, \mathcal{G}) = \emptyset$ and where every possibly causal path from X to $\mathbf{Y}$ in $\mathcal{G}$ starts with a directed edge. If $\mathbf{Y} \cap \mathrm{Pa}(X, \mathcal{G}) = \emptyset$, then the following is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$:*

$$\mathrm{Pa}(X, \mathcal{G}) \setminus \mathbf{Z}. \tag{5}$$

Theorem 5 *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a causal MPDAG $\mathcal{G}$, where $\mathbf{Z} \cap \mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$ and where every proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ starts with a directed edge.*

(a) If there is any conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$, then the following set is one:

$$\begin{aligned} &\mathrm{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \\ &\quad = \big[\mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y}, \mathcal{G}) \cup \mathrm{An}(\mathbf{Z}, \mathcal{G})\big] \\ &\qquad \setminus \big[\mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}\big]. \end{aligned} \tag{6}$$

[2] Compare to Figure 6(a) of Perković et al. (2018).

(b) Suppose $\mathbf{Y} \subseteq \mathrm{PossDe}(\mathbf{X}, \mathcal{G})$*. If there is any conditional adjustment set relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$*, then the following set is one:*

$$\mathrm{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \mathrm{Pa}\Big(\mathrm{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G}), \mathcal{G}\Big) \setminus \Big[\mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}\Big]. \quad (7)$$

Example 5 *Consider again the causal MPDAG* $\mathcal{G}$ *in Figure 2a, where* $\mathbf{X} = \{X\}$*,* $\mathbf{Y} = \{Y\}$*, and* $\mathbf{Z} = \{V_1\}$*. Note that the conditions of Lemma 4 and Theorem 5 are met, so we can construct three valid conditional adjustment sets using Equations* (5)*,* (6)*, and* (7)*.*

$$\begin{aligned}
\mathrm{Pa}(X, \mathcal{G}) \setminus \mathbf{Z} &= \{V_1, V_2, V_3\} \setminus \{V_1\} \\
&= \{V_2, V_3\}. \\
\mathrm{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) &= \{X, Y, V_1, V_2, V_3, V_4\} \setminus \{X, Y, V_1\} \\
&= \{V_2, V_3, V_4\}. \\
\mathrm{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) &= \{X, V_1, V_2, V_4\} \setminus \{X, Y, V_1\} \\
&= \{V_2, V_4\}.
\end{aligned}$$

3.4 Comparison of Contexts

In this section, we point out a bridge between our conditional adjustment results and prior literature on unconditional adjustment and adjustment under *dynamic treatment*. We begin by presenting Lemma 6, which provides an equivalence between our criterion and the criterion of Perković et al. (2017) used for unconditional adjustment given an MPDAG. Note that this lemma is used to prove Theorem 3 (see Figure 5 in Supp. D). See Supp. D for the lemma's proof.

Lemma 6 *Let* $\mathbf{X}$*,* $\mathbf{Y}$*,* $\mathbf{Z}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in an MPDAG* $\mathcal{G}$*, where* $\mathbf{Z} \cap \mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$*. Then we have the following.*

(a) ***Comparison of Adjustment Criteria:*** $\mathbf{S}$ *satisfies the conditional adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$ *(Definition 2) if and only if* $\mathbf{S} \cup \mathbf{Z}$ *satisfies the adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *(Definition 12, Supp. A).*

(b) ***Comparison of Adjustment Sets:*** $\mathbf{S}$ *is a conditional adjustment set relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$ *(Definition 1) if and only if* $\mathbf{S} \cup \mathbf{Z}$ *is an adjustment set relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *(Definition 11, Supp. A).*

Next we turn to the work of Smucler et al. (2020), where the authors consider causal effect estimation under a dynamic treatment. For this purpose, Smucler et al. (2020) define a dynamic adjustment set, which they then relate to the set used by Maathuis and Colombo (2015) for unconditional adjustment (Definition 11, Supp. A). Lemma 6 allows us to connect this dynamic adjustment to our work.

Before making this connection, we briefly describe the context of these authors' work. Unlike a do-intervention that sets $\mathbf{X}$ to fixed values $\mathbf{x}$, a dynamic intervention sets $\mathbf{X}$ to values $\mathbf{x}$ with probability $\pi(\mathbf{x}|\mathbf{Z} = \mathbf{z})$. However, a do-intervention can be seen as a special case of a dynamic intervention where $\pi(\mathbf{x}|\mathbf{Z} = \mathbf{z}) = \mathbb{1}(\mathbf{X} = \mathbf{x})$. Dynamic interventions are often of interest in personalized medicine (Robins, 1993; Murphy et al., 2001; Chakraborty and Moodie, 2013).

Smucler et al. (2020) refer to a causal effect under a dynamic intervention, whose assignment probability depends on $\mathbf{Z}$, as a ***Z****-dependent dynamic causal effect* (also called a *single stage dynamic treatment effect* in Chakraborty and Moodie (2013)). They consider these causal effects in the setting where $\mathbf{X}$ and $\mathbf{Y}$ are nodes, the given graph $\mathcal{G}$ is a DAG, and the following assumption holds: $\mathbf{Z} \cap \mathrm{De}(X, \mathcal{G}) = \emptyset$. They then define a ***Z****-dependent dynamic adjustment set* as a set $\mathbf{S}$ that satisfies

$$f(y|\pi(x|\mathbf{z})) = \begin{cases} \pi(x|\mathbf{z}) f(y|x, \mathbf{z}) & \mathbf{S} = \emptyset, \\ \pi(x|\mathbf{z}) \int f(y|x, \mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z})\, \mathrm{d}\mathbf{s} & \mathbf{S} \neq \emptyset. \end{cases}$$

To compare these sets to our conditional adjustment sets, we reference Proposition 1 of Smucler et al. (2020). This result states that, under their assumptions, $\mathbf{S} \cup \mathbf{Z}$ is a $\mathbf{Z}$-dependent dynamic adjustment set if and only if $\mathbf{S} \cup \mathbf{Z}$ is an adjustment set relative to (X, Y) in $\mathcal{G}$ (Definition 11, Supp. A). It follows from Lemma 6 that $\mathbf{S} \cup \mathbf{Z}$ is a $\mathbf{Z}$-dependent dynamic adjustment set if and only if $\mathbf{S}$ is a conditional adjustment set relative to $(X, Y, \mathbf{Z})$ in $\mathcal{G}$ – when $\mathcal{G}$ is a DAG such that $\mathbf{Z} \cap \mathrm{De}(X, \mathcal{G}) = \emptyset$. Thus, our results can be seen as generalizations of Smucler et al. (2020) for $|\mathbf{X}| > 1$ and, therefore, can be used for $\mathbf{Z}$-dependent dynamic causal effect identification.

4 RESULTS - PAGS

We now extend our results on conditional adjustment to the setting of a PAG.

4.1 Conditional Adjustment Criterion

We first introduce our conditional adjustment criterion for PAGs (Definition 7). Note that the difference between this criterion and the analogous criterion for MPDAGs is the use of a *visible* as opposed to a *directed* edge. Visibility is a stronger condition introduced by Zhang (2008a) (see Supp. A for definition).

Following this, Lemma 8 provides an equivalence between our criterion and the criterion of Perković et al. (2018) used for unconditional adjustment given a PAG. Theorem 9 is our main result in this section. It establishes that, under restrictions on $\mathbf{Z}$, the conditional adjustment criterion is an equivalent graphical characterization of a conditional adjustment set in causal PAGs. Proofs of these results are given in Supp. G.

Definition 7 (**Conditional Adjustment Criterion for PAGs**) *Let* $\mathbf{X}$*,* $\mathbf{Y}$*,* $\mathbf{Z}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a PAG* $\mathcal{G}$*, where* $\mathbf{Z} \cap \text{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$ *and where every proper possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *starts with a* ***visible edge*** *out of* $\mathbf{X}$*. Then* $\mathbf{S}$ *satisfies the conditional adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$ *if*

(a) $\mathbf{S} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$*, and*

(b) $\mathbf{S} \cup \mathbf{Z}$ *blocks all proper non-causal definite status paths from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$*.*

Lemma 8 *Let* $\mathbf{X}$*,* $\mathbf{Y}$*,* $\mathbf{Z}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a PAG* $\mathcal{G}$*, where* $\mathbf{Z} \cap \text{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$*. Then we have the following.*

(a) ***Comparison of Adjustment Criteria:*** $\mathbf{S}$ *satisfies the conditional adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$ *(Definition 7) if and only if* $\mathbf{S} \cup \mathbf{Z}$ *satisfies the adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *(Definition 12, Supp. A).*

(b) ***Comparison of Adjustment Sets:*** $\mathbf{S}$ *is a conditional adjustment set relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$ *(Definition 1) if and only if* $\mathbf{S} \cup \mathbf{Z}$ *is an adjustment set relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *(Definition 11, Supp. A).*

Proof of Lemma 8. *(a)* Follows from the fact that $\text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \subseteq \text{PossDe}(\mathbf{X}, \mathcal{G})$.

(b) We start by noting the following fact. Since $\mathbf{Z} \cap \text{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$, then $\mathbf{Z} \cap \text{De}(\mathbf{X}, \mathcal{D}) = \emptyset$ in every DAG represented by $\mathcal{G}$ (Lemma 49, Supp. G). Then consider the following statements.

(a) $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$.

(b) $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in each DAG represented by $\mathcal{G}$ – were the DAG to be causal.

(c) $\mathbf{S} \cup \mathbf{Z}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in each DAG represented by $\mathcal{G}$ – were the DAG to be causal.

(d) $\mathbf{S} \cup \mathbf{Z}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{G}$.

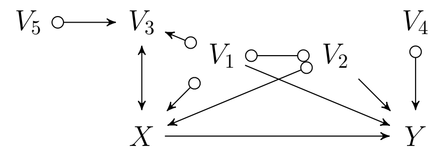

Figure 3: A causal PAG used in Example 6.

By definition, (a) $\Leftrightarrow$ (b). Then by Lemma 6*(b)* and the fact above, we have (b) $\Leftrightarrow$ (c). The statement (c) $\Leftrightarrow$ (d) follows again by definition. ■

Theorem 9 (**Completeness, Soundness of Conditional Adjustment Criterion for PAGs**) *Let* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a causal PAG* $\mathcal{G}$*, where* $\mathbf{Z} \cap \text{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$*. Then* $\mathbf{S}$ *is a conditional adjustment set relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$ *(Definition 1) if and only if* $\mathbf{S}$ *satisfies the conditional adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$ *(Definition 7).*

4.2 Constructing Adjustment Sets

We now provide a method for constructing conditional adjustment sets given a causal PAG (Theorem 10). We illustrate this result in Example 6. The proof of Theorem 10 can be found in Supp. H.

Theorem 10 *Let* $\mathbf{X}$*,* $\mathbf{Y}$*, and* $\mathbf{Z}$ *be pairwise disjoint node sets in a causal PAG* $\mathcal{G}$*, where* $\mathbf{Z} \cap \text{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$ *and where every proper possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *starts with a visible edge out of* $\mathbf{X}$*. If there is any conditional adjustment set relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$*, then the following set is one:*

$$\begin{aligned} &\text{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \\ &\quad = \left[\text{PossAn}(\mathbf{X} \cup \mathbf{Y}, \mathcal{G}) \cup \text{PossAn}(\mathbf{Z}, \mathcal{G})\right] \\ &\qquad \setminus \Big[\text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}\Big]. \end{aligned} \tag{8}$$

Example 6 *Let* $\mathcal{G}$ *be the causal PAG in Figure 3, and let* $\mathbf{X} = \{X\}$*,* $\mathbf{Y} = \{Y\}$*, and* $\mathbf{Z} = \{V_1\}$*. Note that* $\mathbf{Z} \cap \text{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$*. Furthermore, the only possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *is the edge* $X \to Y$*, which is visible due to the presence of* $V_3 \leftrightarrow X$*, where* $V_3 \notin \text{Adj}(Y, \mathcal{G})$*. If there is any conditional adjustment set relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$*, then the conditions of Theorem 10 are met. We consider the set from Equation* (8)*.*

$$\begin{aligned} \text{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) &= \{X, Y, V_1, V_2, V_4\} \setminus \{X, Y, V_1\} \\ &= \{V_2, V_4\}. \end{aligned}$$

To see that this is a conditional adjustment set relative to $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ *in* $\mathcal{G}$*, we note that it fulfills the requirements of Definition 7. That is,* $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cap \text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G}) = \emptyset$ *and* $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cup \mathbf{Z} = \{V_1, V_2, V_4\}$ *blocks all proper non-causal definite status paths from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$*.*

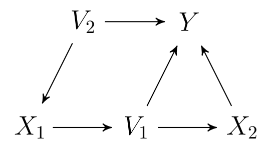

Figure 4: A causal DAG used in Section 5.

5 DISCUSSION

This paper defines a conditional adjustment set that can be used to identify a causal effect in a setting where a causal MPDAG or PAG is known (Definition 1). We give necessary and sufficient graphical conditions for identifying such a set when $\mathbf{Z} \cap \text{PossDe}(\mathbf{X},\mathcal{G}) = \emptyset$ (Theorems 3 and 9). Further, we provide multiple methods for constructing these sets (Sections 3.3 and 4.2). While our results can be used to identify a broad class of conditional causal effects, we discuss some limitations below.

One such limitation is that there are conditional causal effects that can be identified but cannot be identified using conditional adjustment sets. As an example, consider the causal DAG (and therefore, MPDAG) $\mathcal{G}$ in Figure 4, and let $\mathbf{X} = \{X_1, X_2\}$, $\mathbf{Y} = \{Y\}$, and $\mathbf{Z} = \{V_2\}$. Note that the conditional causal effect of $\mathbf{X}$ on $\mathbf{Y}$ given $\mathbf{Z}$ is identifiable using do calculus rules (Pearl, 2009, see Equations (12)-(14) in Supp. B):

$$\begin{aligned}
&f(\mathbf{y}|do(\mathbf{x}),\mathbf{z})\\
&\quad= \int_{v_1} f(y, v_1|do(\mathbf{x}), v_2)\,\mathrm{d}v_1\\
&\quad= \int_{v_1} f(y|do(\mathbf{x}), v_1, v_2) f(v_1|do(\mathbf{x}), v_2)\,\mathrm{d}v_1\\
&\quad= \int_{v_1} f(y|do(\mathbf{x}), v_1, v_2) f(v_1|do(\mathbf{x}))\,\mathrm{d}v_1 \qquad (9)\\
&\quad= \int_{v_1} f(y|do(x_2), v_1, v_2) f(v_1|do(x_1))\,\mathrm{d}v_1 \qquad (10)\\
&\quad= \int_{v_1} f(y|x_2, v_1, v_2) f(v_1|x_1)\,\mathrm{d}v_1. \qquad (11)
\end{aligned}$$

The first two equalities follow from basic probability rules. Equation (9) follows from Rule 1 of the do calculus, since $V_1 \perp_d V_2 | X_1, X_2$ in $\mathcal{G}_{\overline{\{X_1,X_2\}}}$. Equation (10) follows from Rule 3 of the do calculus, since $Y \perp_d X_1 | V_1, V_2, X_2$ in $\mathcal{G}_{\overline{X_2}}$ and $V_1 \perp_d X_2 | X_1$ in $\mathcal{G}_{\overline{\{X_1,X_2\}}}$. Equation (11) follows from Rule 2 of the do calculus, since $Y \perp_d X_2 | V_1, V_2$ in $\mathcal{G}_{\underline{X_2}}$ and $V_1 \perp_d X_1$ in $\mathcal{G}_{\underline{X_1}}$.

However, we can show that there is no conditional adjustment set relative to $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ in $\mathcal{G}$ that could have been used to identify the effect above. To see this, note that since $\mathbf{Z} \cap \text{PossDe}(\mathbf{X},\mathcal{G}) = \emptyset$, we can use Theorem 3 to state the following. A set $\mathbf{S}$ must satisfy the conditional adjustment criterion relative to $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ in $\mathcal{G}$ (Definition 2) in order to be a conditional adjustment set. Definition 2 requires that $\mathbf{S}$ block the path $X_2 \leftarrow V_1 \rightarrow Y$, since it is a proper non-causal definite status path from $\mathbf{X}$ to $\mathbf{Y}$. It follows that $\mathbf{S}$ must contain $V_1 \in \text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$, but this contradicts Definition 2's requirement that $\mathbf{S} \cap \text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G}) = \emptyset$.

Adding to the limitation above, there are conditional causal effects that can be identified using conditional adjustment sets but where these conditional adjustment sets cannot be identified using our criterion. This can occur when $\mathbf{Z} \cap \text{PossDe}(\mathbf{X},\mathcal{G}) \neq \emptyset$, since our graphical criterion requires this restriction but our conditional adjustment set definition does not. As an example, consider again the causal DAG $\mathcal{G}$ given in Figure 2b, and let $\mathbf{X} = \{X_1, X_2\}$, $\mathbf{Y} = \{Y\}$, and $\mathbf{Z} = \{Z, W\}$. Since $\mathbf{Z} \cap \text{PossDe}(\mathbf{X},\mathcal{G}) \neq \emptyset$, no set satisfies the conditional adjustment criterion. However, using do calculus rules (Pearl, 2009), we can show that $\mathbf{S} = \{S\}$ is a conditional adjustment set relative to $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ in $\mathcal{G}$:

$$\begin{aligned}
f(\mathbf{y}|do(\mathbf{x}),\mathbf{z}) &= \int_{\mathbf{s}} f(\mathbf{y},\mathbf{s}|do(\mathbf{x}),\mathbf{z})\,\mathrm{d}\mathbf{s}\\
&= \int_{\mathbf{s}} f(\mathbf{y}|do(\mathbf{x}),\mathbf{z},\mathbf{s}) f(\mathbf{s}|do(\mathbf{x}),\mathbf{z})\,\mathrm{d}\mathbf{s}\\
&= \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x},\mathbf{z},\mathbf{s}) f(\mathbf{s}|\mathbf{z})\,\mathrm{d}\mathbf{s}.
\end{aligned}$$

The first and second equality follow from basic probability rules. The third follows by Rules 2 and 3 of the do calculus, since $\mathbf{Y} \perp_d \mathbf{X} \,|\, \mathbf{Z} \cup \mathbf{S}$ in $\mathcal{G}_{\underline{\mathbf{X}}}$ and $\mathbf{S} \perp_d \mathbf{X} \,|\, \mathbf{Z}$ in $\mathcal{G}_{\overline{\mathbf{X}(\mathbf{Z})}}$. Future work could address identification in this setting by expanding our graphical criterion to allow for arbitrary conditioning.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2210210.

References

J. Abrevaya, Y.-C. Hsu, and R. P. Lieli. Estimating conditional average treatment effects. *Journal of Business & Economic Statistics*, 33(4):485–505, 2015.

R. A. Ali, T. S. Richardson, and P. Spirtes. Markov equivalence for ancestral graphs. *Annals of Statistics*, 37:2808–2837, 2009.

S. A. Andersson, D. Madigan, and M. D. Perlman. A characterization of Markov equivalence classes for acyclic digraphs. *Annals of Statistics*, 25:505–541, 1997.

S. Athey and G. Imbens. Recursive partitioning for heterogeneous causal effects. In *Proceedings of the National Academy of Sciences*, volume 113, pages 7353–7360, 2016.

E. Bareinboim, C. Brito, and J. Pearl. Local characterizations of causal Bayesian networks. In *Graph Structures for Knowledge Representation and Reasoning: Second International Workshop, GKR 2011, Barcelona, Spain, July 16, 2011. Revised Selected Papers*, pages 1–17. Springer, 2012.

J. E. Brand and Y. Xie. Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, 75(2):273–302, 2010.

B. Chakraborty and E. E. Moodie. *Statistical Methods for Dynamic Treatment Regimes*. Springer, 2013.

V. Chernozhukov, W. K. Newey, and R. Singh. A simple and general debiased machine learning theorem with finite-sample guarantees. *Biometrika*, 110(1): 257–264, 2023.

D. M. Chickering. Learning equivalence classes of Bayesian-network structures. *Journal of Machine Learning Research*, 2:445–498, 2002.

Q. Fan, Y.-C. Hsu, R. P. Lieli, and Y. Zhang. Estimation of conditional average treatment effects with high-dimensional data. *Journal of Business & Economic Statistics*, 40(1):313–327, 2022.

A. Hauser and P. Bühlmann. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *Journal of Maching Learning Research*, 13:2409–2464, 2012.

W. H. O. R. Health. *Medical eligibility criteria for contraceptive use*. World Health Organization, 2010.

L. Henckel, E. Perković, and M. H. Maathuis. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *Journal of the Royal Statistical Society: Series B*, pages 579–599, 2022.

A. Jaber, J. Zhang, and E. Bareinboim. Identification of conditional causal effects under Markov equivalence. In *Proceedings of NeurIPS*, pages 11516–11524, 2019.

A. Jaber, A. Ribeiro, J. Zhang, and E. Bareinboim. Causal identification under Markov equivalence: Calculus, algorithm, and completeness. In *Proceedings of NeurIPS*, volume 35, pages 3679–3690, 2022.

M. Kalisch, M. Mächler, D. Colombo, M. H. Maathuis, and P. Bühlmann. Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software*, 47(11):1–26, 2012.

E. H. Kennedy, S. Balakrishnan, J. M. Robins, and L. Wasserman. Minimax rates for heterogeneous causal effect estimation. arXiv preprint arXiv:2203.00837, 2022.

Y. Kivva, J. Etesami, and N. Kiyavash. On identifiability of conditional causal effects. arXiv preprint arXiv:2306.11755, 2023.

S. R. Künzel, J. S. Sekhon, P. J. Bickel, and B. Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. In *Proceedings of the National Academy of Sciences*, volume 116, pages 4156–4165, 2019.

S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B*, pages 157–224, 1988.

S. L. Lauritzen, A. P. Dawid, B. N. Larsen, and H.-G. Leimer. Independence properties of directed Markov fields. *Networks*, 20(5):491–505, 1990.

M. H. Maathuis and D. Colombo. A generalized backdoor criterion. *Annals of Statistics*, 43:1060–1088, 2015.

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis (Probability and Mathematical Statistics)*. Academic Press London, 1980.

C. Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of UAI*, pages 403–410, 1995.

J. M. Mooij, S. Magliacane, and T. Claassen. Joint causal inference from multiple contexts. *The Journal of Machine Learning Research*, 21(1):3919–4026, 2020.

S. A. Murphy, M. J. van der Laan, J. M. Robins, and C. P. P. R. Group. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423, 2001.

X. Nie and S. Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2): 299–319, 2021.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

E. Perković. Identifying causal effects in maximally oriented partially directed acyclic graphs. In *Proceedings of UAI*, pages 530–539, 2020.

E. Perković, J. Textor, M. Kalisch, and M. H. Maathuis. A complete generalized adjustment criterion. In *Proceedings of UAI*, pages 682–691, 2015.

E. Perković, M. Kalisch, and M. H. Maathuis. Interpreting and using CPDAGs with background knowledge. In *Proceedings of UAI*, 2017.

E. Perković, J. Textor, M. Kalisch, and M. H. Maathuis. Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18, 2018.

T. S. Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Jouranl of Statistics*, 30:145–157, 2003.

T. S. Richardson and P. Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30:962–1030, 2002.

J. M. Robins. Analytic methods for estimating HIV-treatment and cofactor effects. *Methodological Issues in AIDS Behavioral Research*, pages 213–288, 1993.

D. Rothenhäusler, J. Ernest, and P. Bühlmann. Causal inference in partially linear structural equation models: identifiability and estimation. *Annals of Statistics*, 46:2904–2938, 2018.

I. Shpitser and J. Pearl. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, 9:1941–1979, 2008.

E. Smucler, F. Sapienza, and A. Rotnitzky. Efficient adjustment sets in causal graphical models with hidden variables. *Biometrika*, 2020.

P. Spirtes, C. Meek, and T. S. Richardson. *Computation, Causation and Discovery*, chapter An algorithm for causal inference in the presence of latent variables and selection bias, pages 211–252. MIT Press, 1999.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* MIT Press, second edition, 2000.

C. Squires and C. Uhler. Causal structure learning: A combinatorial perspective. *Foundations of Computational Mathematics*, pages 1–35, 2022.

J. Textor, B. Van der Zander, M. S. Gilthorpe, M. Liśkiewicz, and G. T. Ellison. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology*, 45 (6):1887–1894, 2016.

S. Wager and S. Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20(7):557–585, 1921.

J. Zhang. *Causal Inference and Reasoning in Causally Insufficient Systems.* PhD thesis, Carnegie Mellon University, 2006.

J. Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474, 2008a.

J. Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172:1873–1896, 2008b.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes/No/Not Applicable] **Not Applicable**
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes/No/Not Applicable] **Not Applicable**
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes/No/Not Applicable] **Not Applicable**

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes/No/Not Applicable] **Yes**
   (b) Complete proofs of all theoretical results. [Yes/No/Not Applicable] **Yes**
   (c) Clear explanations of any assumptions. [Yes/No/Not Applicable] **Yes**

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the Supplemental material or as a URL). [Yes/No/Not Applicable] **Not Applicable**
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes/No/Not Applicable] **Not Applicable**
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes/No/Not Applicable] **Not Applicable**

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes/No/Not Applicable] **Not Applicable**

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes/No/Not Applicable] **Not Applicable**
   (b) The license information of the assets, if applicable. [Yes/No/Not Applicable] **Not Applicable**
   (c) New assets either in the Supplemental material or as a URL, if applicable. [Yes/No/Not Applicable] **Not Applicable**
   (d) Information about consent from data providers/curators. [Yes/No/Not Applicable] **Not Applicable**
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Yes/No/Not Applicable] **Not Applicable**

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Yes/No/Not Applicable] **Not Applicable**
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Yes/No/Not Applicable] **Not Applicable**
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Yes/No/Not Applicable] **Not Applicable**

Supplement to:
Conditional Adjustment in a Markov Equivalence Class

A FURTHER PRELIMINARIES AND DEFINITIONS

A.1 Preliminaries

Path Construction. A *subsequence* of a path p is a path obtained by deleting non-endpoint nodes from p without changing the order of the remaining nodes. Let $p = \langle X_1, X_2, \dots, X_k \rangle$ and i, j, k such that $1 \le i < j \le k$. We denote the concatenation of paths by the symbol $\oplus$, so that $p = p(X_1, X_i) \oplus p(X_i, X_k)$. We use the notation $(-p)(X_j, X_i)$ to denote the path $\langle X_j, X_{j-1}, \dots, X_i \rangle$.

A.2 Definitions

Definition 11 (**Adjustment Set for MPDAGs (PAGs)**; Perković et al., 2017, 2018, 2015; cf. Maathuis and Colombo, 2015) *Let* $\mathbf{X}$*,* $\mathbf{Y}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a causal MPDAG (PAG)* $\mathcal{G}$*. Then* $\mathbf{S}$ *is an adjustment set relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *if for any density* f *consistent with* $\mathcal{G}$

$$f(\mathbf{y}|do(\mathbf{x})) = \begin{cases} f(\mathbf{y}|\mathbf{x}) & \mathbf{S} = \emptyset \\ \int f(\mathbf{y}|\mathbf{x}, \mathbf{s}) f(\mathbf{s}) \,\mathrm{d}\mathbf{s} & \mathbf{S} \neq \emptyset. \end{cases}$$

Definition 12 (**Adjustment Criterion for MPDAGs (PAGs)**; Perković et al., 2017, 2018) *Let* $\mathbf{X}$*,* $\mathbf{Y}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in an MPDAG (PAG)* $\mathcal{G}$*, where every proper possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *starts with a directed (visible) edge out of* $\mathbf{X}$*. Then* $\mathbf{S}$ *satisfies the adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *if*

(a) $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$*, and*

(b) $\mathbf{S}$ *blocks all proper non-causal definite status paths from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$*.*

Definition 13 (**Generalized Back-Door Criterion for DAGs**; cf. Maathuis and Colombo, 2015) *Let* $\mathbf{X}$*,* $\mathbf{Y}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a DAG* $\mathcal{D}$*. Then* $\mathbf{S}$ *satisfies the generalized back-door criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{D}$ *if*

(a) $\mathbf{S} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$*, and*

(b) $\mathbf{S} \cup \mathbf{X} \setminus \{X\}$ *blocks all back-door paths from* X *to* $\mathbf{Y}$ *in* $\mathcal{D}$*, for every* $X \in \mathbf{X}$*.*

Definition 14 (**Proper Back-Door Graph for DAGs**; cf. Perković et al., 2018) *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be disjoint node sets in a DAG* $\mathcal{D}$*. The proper back-door graph* $\mathcal{D}^{pbd}_{\mathbf{XY}}$ *is obtained from* $\mathcal{D}$ *by removing all edges out of* $\mathbf{X}$ *that are on proper causal paths from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{D}$*.*

Definition 15 (**Moral Graph for DAGs**; cf. Lauritzen and Spiegelhalter, 1988; cf. Perković et al., 2018) *Let* $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ *be a DAG. The moral graph* $\mathcal{D}^m$ *is formed by adding the edge* $A - B$ *to any structure of the form* $A \to C \leftarrow B$ *for any* $A, B, C \in \mathbf{V}$*, with* $A \notin \mathrm{Adj}(B, \mathcal{D})$ *(marrying unmarried parents) and subsequently making all edges in the resulting graph undirected.*

Definition 16 (**Distance to $\mathbf{Z}$**; Zhang, 2006; Perković et al., 2017) *Let $\mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ be pairwise disjoint node sets in an MPDAG or PAG $\mathcal{G}$. Let p be a path between $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{G}$ such that every collider C on p has a possibly directed path (possibly of length 0) to $\mathbf{Z}$. Define the distance to $\mathbf{Z}$ of C to be the length of a shortest possibly directed path (possibly of length 0) from C to $\mathbf{Z}$, and define the distance to $\mathbf{Z}$ of p to be the sum of the distances from $\mathbf{Z}$ of the colliders on p.*

B EXISTING RESULTS

Rules of the Do Calculus (Pearl, 2009). Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, and $\mathbf{W}$ be pairwise disjoint (possibly empty) node sets in a causal DAG $\mathcal{D}$. Let $\mathcal{D}_{\overline{\mathbf{X}}}$ denote the graph obtained by deleting all edges into $\mathbf{X}$ from $\mathcal{D}$. Similarly, let $\mathcal{D}_{\underline{\mathbf{X}}}$ denote the graph obtained by deleting all edges out of $\mathbf{X}$ in $\mathcal{D}$, and let $\mathcal{D}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ denote the graph obtained by deleting all edges into $\mathbf{X}$ and all edges out of $\mathbf{Z}$ in $\mathcal{D}$. The following rules hold for all densities consistent with $\mathcal{D}$.

Rule 1. If $(\mathbf{Y} \perp_d \mathbf{Z} \mid \mathbf{X} \cup \mathbf{W})_{\mathcal{D}_{\overline{\mathbf{X}}}}$, then

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}) = f(\mathbf{y}|do(\mathbf{x}), \mathbf{w}). \tag{12}$$

Rule 2. If $(\mathbf{Y} \perp_d \mathbf{X} \mid \mathbf{Z} \cup \mathbf{W})_{\mathcal{D}_{\underline{\mathbf{X}}\overline{\mathbf{W}}}}$, then

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, do(\mathbf{w})) = f(\mathbf{y}|\mathbf{x}, \mathbf{z}, do(\mathbf{w})). \tag{13}$$

Rule 3. If $(\mathbf{Y} \perp_d \mathbf{X} \mid \mathbf{Z} \cup \mathbf{W})_{\mathcal{D}_{\overline{\mathbf{X}(\mathbf{Z}) \cup \mathbf{W}}}}$, then

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, do(\mathbf{w})) = f(\mathbf{y}|\mathbf{z}, do(\mathbf{w})), \tag{14}$$

where $\mathbf{X}(\mathbf{Z}) = \mathbf{X} \setminus \mathrm{An}(\mathbf{Z}, \mathcal{D}_{\overline{\mathbf{W}}})$.

Lemma 17 (Wright's Rule of Wright, 1921) *Let $\mathbf{X} = \mathbf{A}\mathbf{X} + \epsilon$, where $\mathbf{Q} \in \mathbb{R}^{k \times k}$, $\mathbf{X} = (X_1, \ldots, X_k)^T$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_k)^T$ is a vector of mutually independent errors with means zero. Moreover, let $Var(\mathbf{X}) = \mathbf{I}$. Let $\mathcal{D} = (\mathbf{X}, \mathbf{E})$, be the corresponding DAG such that $X_i \to X_j$ is in $\mathcal{D}$ if and only if $A_{ji} \neq 0$. A non-zero entry A_{ji} is called the edge coefficient of $X_i \to X_j$. For two distinct nodes X_i, $X_j \in \mathbf{X}$, let $p_1, \ldots, p_r$ be all paths between X_i and X_j in $\mathcal{D}$ that do not contain a collider. Then $\mathrm{Cov}(X_i, X_j) = \sum_{s=1}^{r} \pi_s$, where π_s is the product of all edge coefficients along path p_s, $s \in \{1, \ldots, r\}$.*

Lemma 18 (Theorem 3.2.4 of Mardia et al., 1980) *Let $\mathbf{X} = (\mathbf{X_1}^T, \mathbf{X_2}^T)^T$ be a p-dimensional multivariate Gaussian random vector with mean vector $\mu = (\mu_\mathbf{1}^T, \mu_\mathbf{2}^T)^T$ and covariance matrix $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma_{11}} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma_{21}} & \mathbf{\Sigma}_{22} \end{bmatrix}$, so that $\mathbf{X_1}$ is a q-dimensional multivariate Gaussian random vector with mean vector $\mu_\mathbf{1}$ and covariance matrix $\mathbf{\Sigma_{11}}$ and $\mathbf{X_2}$ is a $(p-q)$-dimensional multivariate Gaussian random vector with mean vector $\mu_\mathbf{2}$ and covariance matrix $\mathbf{\Sigma_{22}}$. Then $E[\mathbf{X_2}|\mathbf{X_1} = \mathbf{x_1}] = \mu_\mathbf{2} + \mathbf{\Sigma_{21}}\mathbf{\Sigma_{11}}^{-1}(\mathbf{x_1} - \mu_\mathbf{1})$.*

Lemma 19 (cf. Theorem 1 and Proposition 3 of Lauritzen et al., 1990) *Let $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ be a DAG, and let f be an observational density over $\mathbf{V}$. Then f is Markov compatible with $\mathcal{D}$ if and only if*

$$V_i \perp\!\!\!\perp \Big[\mathbf{V} \setminus \big(\mathrm{De}(V_i, \mathcal{D}) \cup \mathrm{Pa}(V_i, \mathcal{D})\big)\Big] \,|\, \mathrm{Pa}(V_i, \mathcal{D})$$

for all $V_i \in \mathbf{V}$, where $\perp\!\!\!\perp$ indicates independence with respect to f.

Lemma 20 (cf. Lemma 3.2 of Perković et al., 2017) *Let $\mathbf{X}$ and $\mathbf{Z}$ be disjoint node sets in an MPDAG $\mathcal{G}$. If $\mathbf{Z} \cap \mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$, then $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$ in every DAG $\mathcal{D}$ in $[\mathcal{G}]$.*

Lemma 21 (Lemma C.2 of Perković et al., 2017, Lemma 9 of Perković et al., 2018) *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{S}$ be pairwise disjoint node sets in an MPDAG (PAG) $\mathcal{G}$, where every proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ starts with a directed (visible) edge out of $\mathbf{X}$. Then the following statements are equivalent.*

(i) $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$.

(ii) $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D}) = \emptyset$ *in every DAG (MAG)* $\mathcal{D}$ *in* $[\mathcal{G}]$.

Lemma 22 (cf. Lemma C.3 of Perković et al., 2017, Lemma 10 of Perković et al., 2018) *Let* $\mathbf{X}, \mathbf{Y}$ *and* $\mathbf{S}$ *be pairwise disjoint node sets in an MPDAG (PAG)* $\mathcal{G}$*, where every proper possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *starts with a directed (visible) edge out of* $\mathbf{X}$ *and where* $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$*. Then the following statements are equivalent.*

(i) $\mathbf{S}$ *blocks all proper non-causal definite status paths from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$.

(ii) $\mathbf{S}$ *blocks all proper non-causal definite status paths from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{D}$ *for every DAG (MAG)* $\mathcal{D}$ *in* $[\mathcal{G}]$.

Theorem 23 (cf. Proposition 3 of Lauritzen et al. (1990), cf. Corollary 2 of Richardson (2003)) *Let* $\mathbf{X}, \mathbf{Y}$*, and* $\mathbf{Z}$ *be pairwise disjoint node sets in a DAG* $\mathcal{D}$*. Further let* $(\mathcal{D}_{\mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{D})})^m$ *be the moral induced subgraph of* $\mathcal{D}$ *on nodes* $\mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{D})$ *(see Definition 15). Then* $\mathbf{Z}$ *d-separates* $\mathbf{X}$ *and* $\mathbf{Y}$ *in* $\mathcal{D}$ *if and only if all paths between* $\mathbf{X}$ *and* $\mathbf{Y}$ *in* $(\mathcal{D}_{\mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{D})})^m$ *contain at least one node in* $\mathbf{Z}$.

Theorem 24 (cf. Theorem 7 of Perković et al., 2018) *Consider the definition of the adjustment criterion for MPDAGs (Definition 12) in the specific setting of a DAG. In this setting, replacing condition* (b) *in Definition 12 with*

(b) $\mathbf{S}$ *d-separates* $\mathbf{X}$ *and* $\mathbf{Y}$ *in* $\mathcal{D}^{pbd}_{\mathbf{XY}}$ *(see Definition 14)*

results in a criterion that is equivalent to Definition 12 applied to a DAG.

Theorem 25 (cf. Theorem 3.1 of Maathuis and Colombo, 2015) *Let* $\mathbf{X}, \mathbf{Y}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a causal DAG* $\mathcal{D}$*. If* $\mathbf{S}$ *satisfies the generalized back-door criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{D}$ *(Definition 13), then* $\mathbf{S}$ *is an adjustment set relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{D}$ *(Definition 11).*

Lemma 26 (cf. Lemma E.6 of Henckel et al., 2022) *Let* $\mathbf{X}, \mathbf{Y}$ *be disjoint node sets in an MPDAG* $\mathcal{G}$*. If there is no proper possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *that starts with an undirected edge in* $\mathcal{G}$*, then* $\mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \subseteq \mathrm{De}(\mathbf{X}, \mathcal{G})$.

Lemma 27 (cf. Lemma 3.5 of Perković et al., 2017) *Let* $p = \langle V_1, \dots, V_k \rangle, k > 1$*, be a definite status path in MPDAG* $\mathcal{G}$*. Then* p *is a possibly causal path in* $\mathcal{G}$ *if and only if there is no edge* $V_i \leftarrow V_{i+1}$*,* $i \in \{1, \dots, k-1\}$ *in* $\mathcal{G}$.

Lemma 28 (cf. Lemma 3.3.1 of Zhang, 2006) *Let* X*,* Y*, and* Z *be distinct nodes in a PAG* $\mathcal{G}$*. If* $X \bullet\!\!\rightarrow Y \circ\!\!-\!\!\bullet Z$*, then there is an edge between* X *and* Z *with an arrowhead at* Z*. Furthermore, if the edge between* X *and* Y *is* $X \rightarrow Y$*, then the edge between* X *and* Z *is either* $X \circ\!\!\rightarrow Z$ *or* $X \rightarrow Z$ *(that is, not* $X \leftrightarrow Z$*).*

Lemma 29 (cf. Lemma 7.5 of Maathuis and Colombo, 2015) *Let* X *and* Y *be two distinct nodes in a MAG or PAG* $\mathcal{G}$*. Then* $\mathcal{G}$ *cannot have both an edge* $Y \bullet\!\!\rightarrow X$ *and a path* $\langle X = V_1, \dots, V_k = Y \rangle, k > 2$ *where each edge* $\langle V_i, V_{i+1} \rangle, i \in \{1, \dots, k-1\}$*, is of one of these forms:* $V_i \rightarrow V_{i+1}$ *or* $V_i \circ\!\!-\!\!\bullet V_{i+1}$.

Lemma 30 (cf. Lemma 17 of Perković et al., 2018) *Let* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ *and* $\mathbf{S}$ *be pairwise disjoint node sets in a MAG or PAG* $\mathcal{G}$*. Suppose that every proper possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *starts with a visible edge out of* $\mathbf{X}$ *and that* $[\mathbf{S} \cup \mathbf{Z}] \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$*. Suppose furthermore that there is a path* p *from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *such that*

(i) p *is a proper definite status non-causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$*,*

(ii) all colliders on p *are in* $\mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z} \cup \mathbf{S}, \mathcal{G}) \setminus [\mathbf{X} \cup \mathbf{Y} \cup \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})]$*, and*

(iii) no definite non-collider on p *is in* $\mathbf{S} \cup \mathbf{Z}$.

Then there is a proper definite status non-causal path from $\mathbf{X}$ *to* $\mathbf{Y}$ *that is m-connecting given* $\mathbf{S} \cup \mathbf{Z}$ *in* $\mathcal{G}$.

Theorem 31 (cf. Theorem 4.4 of Perković et al., 2017, Theorems 5 and 56 of Perković et al., 2018) *Let* $\mathbf{X}$*,* $\mathbf{Y}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a causal MPDAG (PAG)* $\mathcal{G}$*. Then* $\mathbf{S}$ *is an adjustment set relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *(Definition 11) if and only if* $\mathbf{S}$ *satisfies the adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{G}$ *(Definition 12).*

Lemma 32 (cf. Lemma F.1 of Rothenhäusler et al., 2018) *Let X and Y be nodes in an MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ such that $X - Y$ is in $\mathcal{G}$. Let $\mathcal{G}'$ be an MPDAG constructed from $\mathcal{G}$ by adding $X \rightarrow Y$ and completing the orientation rules R1 - R4 of Meek (1995). For any $Z, W \in \mathbf{V}$, if $Z - W$ is in $\mathcal{G}$ and $Z \rightarrow W$ is in $\mathcal{G}'$, then $W \in \mathrm{De}(Y, \mathcal{G}')$.*

Lemma 33 (cf. Lemma F.2 of Rothenhäusler et al., 2018) *Let X be a node in an MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, and let $\mathbf{S}$ be a set such that for all $S \in \mathbf{S}$, $X - S$ is in $\mathcal{G}$. Then there is an MPDAG $\mathcal{G}' = (\mathbf{V}, \mathbf{E}')$ that is formed by taking $\mathcal{G}$, orienting $X \rightarrow S$ for all $S \in \mathbf{S}$, and completing R1-R4 of Meek (1995).*

Lemma 34 (cf. Lemma 59 of Perković et al., 2018) *Let $\mathbf{X}, \mathbf{Y}$ and $\mathbf{S}$ be pairwise disjoint node sets in a DAG $\mathcal{D}$ such that $\mathbf{S}$ satisfies the adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$ (Definition 12). Let $\mathbf{J} \subseteq \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D}) \setminus (\mathrm{De}(\mathbf{X}, \mathcal{D}) \cup \mathbf{Y})$ and $\tilde{\mathbf{S}} = \mathbf{S} \cup \mathbf{J}$. Then the following statements hold:*

(i) $\tilde{\mathbf{S}}$ *satisfies the adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$, and*

(ii) $\int_{\mathbf{s}} f(\mathbf{y} \mid \mathbf{x}, \mathbf{s}) f(\mathbf{s}) d\mathbf{s} = \int_{\tilde{\mathbf{s}}} f(\mathbf{y} \mid \mathbf{x}, \tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}) d\tilde{\mathbf{s}}$, *for any density f consistent with $\mathcal{D}$.*

Lemma 35 (Lemma 60 of Perković et al., 2018) *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{S}$ be pairwise disjoint node sets in a causal DAG $\mathcal{D}$ such that $\mathbf{S}$ satisfies the adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$. Let $\mathbf{J} = \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D}) \setminus \big(\mathrm{De}(\mathbf{X}, \mathcal{D}) \cup \mathbf{Y}\big)$ and $\tilde{\mathbf{S}} = \mathbf{S} \cup \mathbf{J}$. Additionally, let $\tilde{\mathbf{S}}_{\mathbf{D}} = \tilde{\mathbf{S}} \cap \mathrm{De}(\mathbf{X}, \mathcal{D})$, $\tilde{\mathbf{S}}_{\mathbf{N}} = \tilde{\mathbf{S}} \setminus \mathrm{De}(\mathbf{X}, \mathcal{D})$, $\mathbf{Y}_{\mathbf{D}} = \mathbf{Y} \cap \mathrm{De}(\mathbf{X}, \mathcal{D})$ and $\mathbf{Y}_{\mathbf{N}} = \mathbf{Y} \setminus \mathrm{De}(\mathbf{X}, \mathcal{D})$. Then the following statements hold:*

(i) $(\mathbf{X} \cup \mathbf{Y}_{\mathbf{N}} \cup \tilde{\mathbf{S}}) \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D}) = \emptyset$,

(ii) if $p = \langle H, \ldots, Y_D \rangle$ is a non-causal path from $H \in \mathbf{X} \cup \mathbf{Y}_{\mathbf{N}} \cup \tilde{\mathbf{S}}$ to $Y_D \in \mathbf{Y}_{\mathbf{D}}$, then p is blocked by $(\mathbf{X} \cup \mathbf{Y}_{\mathbf{N}} \cup \tilde{\mathbf{S}}_{\mathbf{N}}) \setminus \{H\}$ in $\mathcal{D}$,

(iii) $\mathbf{Y}_{\mathbf{D}} \perp_d \tilde{\mathbf{S}}_{\mathbf{D}} \mid \mathbf{Y}_{\mathbf{N}} \cup \mathbf{X} \cup \tilde{\mathbf{S}}_{\mathbf{N}}$ *in $\mathcal{D}$, where $\mathbf{Y}_{\mathbf{N}} = \emptyset$ is allowed,*

(iv) if $\mathbf{Y}_{\mathbf{N}} = \emptyset$ then $\tilde{\mathbf{S}}_{\mathbf{N}}$ satisfies the generalized back-door criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$ (Definition 13),

(v) the empty set satisfies the generalized back-door criterion relative to $(\mathbf{X} \cup \mathbf{Y}_{\mathbf{N}} \cup \tilde{\mathbf{S}}_{\mathbf{N}}, \mathbf{Y}_{\mathbf{D}})$ in $\mathcal{D}$,

(vi) $\mathbf{Y}_{\mathbf{D}} \perp_d (\mathbf{Y}_{\mathbf{N}} \cup \tilde{\mathbf{S}}_{\mathbf{N}}) \mid \mathbf{X}$ *in* $\mathcal{D}_{\overline{\mathbf{X}}\underline{\mathbf{Y}_{\mathbf{N}} \cup \tilde{\mathbf{S}}_{\mathbf{N}}}}$, *and*

(vii) $\tilde{\mathbf{S}}_{\mathbf{N}} \perp_d \mathbf{X} \mid \mathbf{Y}_{\mathbf{N}}$ *in* $\mathcal{D}_{\overline{\mathbf{X}}}$.

C A NECESSARY CONDITION FOR IDENTIFIABILITY

This section includes the proof of Proposition 36, which provides a necessary condition for the identifiability of the conditional causal effect given an MPDAG. This result is needed twice – once for the proof of Theorem 3 in Section 3.1 and once for Example 4 in Section 3.2. Below we also provide two supporting results for the proof of Proposition 36 – namely, Lemmas 37 and 38.

C.1 Main Result

Proposition 36 *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a causal MPDAG $\mathcal{G}$. If there is a proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that starts with an undirected edge and does not contain any element of $\mathbf{Z}$, then the conditional causal effect of $\mathbf{X}$ on $\mathbf{Y}$ given $\mathbf{Z}$ is not identifiable in $\mathcal{G}$.*

Proof of Proposition 36. This lemma extends Proposition 3.2 of Perković (2020) and its proof follows similar logic to that of Perković (2020).

Suppose that there is a proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ that starts with an undirected edge and does not contain any element of $\mathbf{Z}$. Then by Lemma 37, there is one such path – call it $q = \langle X = V_0, \ldots, V_k = Y \rangle$, $X \in \mathbf{X}$, $Y \in \mathbf{Y}, k \geq 1$ – where the corresponding paths in two DAGs in $[\mathcal{G}]$ take the forms $X \rightarrow \cdots \rightarrow Y$ and $X \leftarrow V_1 \rightarrow \cdots \rightarrow Y$ ($X \leftarrow Y$ when $k = 1$). Call these DAGs $\mathcal{D}^1$ and $\mathcal{D}^2$ with paths q_1 and q_2, respectively.

To prove that the conditional causal effect of $\mathbf{X}$ on $\mathbf{Y}$ given $\mathbf{Z}$ is not identifiable in $\mathcal{G}$, it suffices to show that there are two families of interventional densities over $\mathbf{V}$ – call them $\mathbf{F_1^*}$ and $\mathbf{F_2^*}$, where for $i \in \{1,2\}$, we define $\mathbf{F_i^*} = \{f_i(\mathbf{v}|do(\mathbf{x}')) : \mathbf{X'} \subseteq \mathbf{V}\}$ – such that the following properties hold.

(i) $\mathcal{D}^1$ and $\mathcal{D}^2$ are compatible with $\mathbf{F_1^*}$ and $\mathbf{F_2^*}$, respectively. [1]

(ii) $f_1(\mathbf{v}) = f_2(\mathbf{v})$.

(iii) $f_1(\mathbf{y}|do(\mathbf{x}),\mathbf{z}) \neq f_2(\mathbf{y}|do(\mathbf{x}),\mathbf{z})$.

To define such families, we start by introducing an additional DAG and an observational density $f(\mathbf{v})$. That is, let $\mathcal{D}^{1'}$ be a DAG constructed by removing every edge from $\mathcal{D}^1$ except for the edges on q_1. Then let $f(\mathbf{v})$ be the multivariate normal distribution under the following linear structural equation model (SEM). Each random variable $A \in \mathbf{V}$ has mean zero and is a linear combination of its parents in $\mathcal{D}^{1'}$ and $\epsilon_A \sim N(0, \sigma_A^2)$, where $\{\epsilon_A : A \in \mathbf{V}\}$ are mutually independent. The coefficients in this linear combination are defined by the edge coefficients of $\mathcal{D}^{1'}$. We pick these edge coefficients in conjunction with $\{\sigma_A^2 : A \in \mathbf{V}\}$ in such a way that each coefficient is in $(0,1)$ and $\mathrm{Var}(A) = 1$ for all $A \in \mathbf{V}$.

From this, we define $\mathbf{F_1^*} = \{f_1(\mathbf{v}|do(\mathbf{x}')) : \mathbf{X'} \subseteq \mathbf{V}\}$ such that $\mathcal{D}^{1'}$ is compatible with $\mathbf{F_1^*}$ and such that $f_1(\mathbf{v}) = f(\mathbf{v})$. Note that $f(\mathbf{v})$ is Markov compatible with $\mathcal{D}^{1'}$ by construction, and we build the interventional densities in $\mathbf{F_1^*}$ by replacing the intervening random variables in the SEM with their interventional values (Pearl, 2009).

To construct the second family of interventional densities, we introduce the DAG $\mathcal{D}^{2'}$, which we form by removing every edge from $\mathcal{D}^2$ except for the edges on q_2. Then note that we could have defined $f(\mathbf{v})$ using a linear SEM based on the parents in $\mathcal{D}^{2'}$. In this case, the resulting observational density would again be a multivariate normal with mean vector zero and a covariance matrix with ones on the diagonal. The off-diagonal entries would be the covariances between the variables in $\mathcal{D}^{2'}$. But note that by Lemma 17, these values will equal the product of all edge coefficients between the relevant nodes in $\mathcal{D}^{2'}$. Since $\mathcal{D}^{1'}$ and $\mathcal{D}^{2'}$ contain no paths with colliders, the observational density $f(\mathbf{v})$ built using $\mathcal{D}^{2'}$ will be an identical distribution to that built under $\mathcal{D}^{1'}$. Thus, in an analogous way to $\mathbf{F_1^*}$, we define $\mathbf{F_2^*} = \{f_1(\mathbf{v}|do(\mathbf{x}')) : \mathbf{X'} \subseteq \mathbf{V}\}$ such that $\mathcal{D}^{2'}$ is compatible with $\mathbf{F_2^*}$ and such that $f_2(\mathbf{v}) = f(\mathbf{v})$.

Having defined $\mathbf{F_1^*}$ and $\mathbf{F_2^*}$, we check that their desired properties hold. Note that by construction, $\mathcal{D}^{1'}$ and $\mathcal{D}^{2'}$ are compatible with $\mathbf{F_1^*}$ and $\mathbf{F_2^*}$, respectively. Thus (i) holds by Lemma 38. Similarly by construction, (ii) holds. To show that (iii) holds, it suffices to show that $E[Y|do(\mathbf{X}=\mathbf{1}),\mathbf{Z}]$ is not the same under f_1 and f_2.

To calculate these expectations, we first want to apply Rules 1-3 of the do calculus (Equations (12)-(14)). Since $f_i(\mathbf{v}|do(\mathbf{x}))$, $i \in \{1,2\}$, is consistent with $\mathcal{D}^{i'}$, we apply these rules using graphical relationships in $\mathcal{D}^{i'}$. Because the path in $\mathcal{D}^{i'}$ corresponding to q_i, $i \in \{1,2\}$, does not contain nodes in $\mathbf{Z}$ or $\mathbf{X} \setminus \{X\}$, then $Y \perp_d \mathbf{Z}|\mathbf{X}$ and $Y \perp_d \mathbf{X} \setminus \{X\}|X$ in $\mathcal{D}^{i'}_{\overline{\mathbf{X}}}$. Further, $Y \perp_d X$ in $\mathcal{D}^{1'}_{\underline{X}}$ and $Y \perp_d X$ in $\mathcal{D}^{2'}_{\overline{X}}$. Thus by Rules 1-3 of the do calculus (Equations (12)-(14)), the following hold.

$$E_1[Y|do(\mathbf{X}=\mathbf{1}),\mathbf{Z}] = E_1[Y|do(X=1)] = E_1[Y|X=1] := a.$$
$$E_2[Y|do(\mathbf{X}=\mathbf{1}),\mathbf{Z}] = E_2[Y|do(X=1)] = E_2[Y] := b,$$

where $E_i, i \in \{1,2\}$ is the expectation under f_i. To calculate a and b, we rely on the observational density $f(\mathbf{v})$, which was constructed using $\mathcal{D}^{1'}$. By Lemma 18, a equals the covariance of X and Y under $f(\mathbf{v})$, and by Lemma 17, $\mathrm{Cov}(X,Y)$ equals the product of all edge coefficients in $\mathcal{D}^{1'}$, which were chosen to be in $(0,1)$. Therefore, $a \neq 0$. But by definition of $f(\mathbf{v})$, $b = 0$. ■

C.2 Supporting Result

Lemma 37 *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in an MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. Suppose that there is a proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that starts with an undirected edge and does not contain nodes in $\mathbf{Z}$. Then there is one such path $\langle X = V_0, \dots, V_k = Y\rangle$, $X \in \mathbf{X}$, $Y \in \mathbf{Y}$, $k \geq 1$, where the corresponding paths in two DAGs in $[\mathcal{G}]$ take the forms $X \to \cdots \to Y$ and $X \leftarrow V_1 \to \cdots \to Y$ ($X \leftarrow Y$ when $k = 1$), respectively.*

[1]For brevity, we say a DAG is "compatible with" a set of interventional densities and an interventional density is "consistent with" a DAG as shorthand for these claims holding only were the DAG to be causal.

Proof of Lemma 37. This lemma is similar to Lemma A.3 of Perković (2020) and its proof borrows from the proof strategy of Lemma C.1 of Perković et al. (2017).

Let q^* be an arbitrary proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that starts with an undirected edge and does not contain nodes in $\mathbf{Z}$. Then let $q = \langle X = V_0, \dots, V_k = Y \rangle$, $X \in \mathbf{X}$, $Y \in \mathbf{Y}$, $k \geq 1$, be a shortest subsequence of q^* in $\mathcal{G}$ that also starts with an undirected edge. Note that q is a proper possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that starts with an undirected edge and does not contain nodes in $\mathbf{Z}$.

Consider when q is of definite status. Since q is possibly causal, all non-endpoints of q are definite non-colliders. Let $\mathcal{D}^1$ be a DAG in $[\mathcal{G}]$ that contains $X \rightarrow V_1$. Then since V_1 is either Y or a definite non-collider on q, the path corresponding to q in $\mathcal{D}^1$ takes the form $X \rightarrow \dots \rightarrow Y$ by induction. Let $\mathcal{D}^2$ be a DAG in $[\mathcal{G}]$ with no additional edges into V_1 compared to $\mathcal{G}$ (Lemma 33). Since $\mathcal{G}$ contains $X - V_1$, $\mathcal{D}^2$ contains $X \leftarrow V_1$. When $k > 1$, $\mathcal{G}$ contains either $V_1 - V_2$ or $V_1 \rightarrow V_2$, and so $\mathcal{D}^2$ contains $X \leftarrow V_1 \rightarrow V_2$. Thus by the same inductive reasoning as above, the path corresponding to q in $\mathcal{D}^2$ takes the form $X \leftarrow V_1 \rightarrow \dots \rightarrow Y$ (or simply $X \leftarrow Y$ when $k = 1$).

Consider instead when q is not of definite status. Note that $k > 1$. To see that q contains $V_1 - V_2$, note that by the choice of q and the fact that q is possibly causal, $q(V_1, Y)$ is unshielded and possibly causal. Thus, $q(V_1, Y)$ is of definite status. However, q is not of definite status, so V_1 must not be of definite status on q, which implies that q cannot contain $V_1 \rightarrow V_2$. Since q is possibly causal, it also cannot contain $V_1 \leftarrow V_2$.

To find two DAGs in $[\mathcal{G}]$ with paths corresponding to q that fit our desired forms, we narrow our search to $[\mathcal{G}']$, where we let $\mathcal{G}'$ be an MPDAG constructed from $\mathcal{G}$ by adding $V_1 \rightarrow V_2$ and completing R1-R4 of Meek (1995). We show below that the path corresponding to q in $\mathcal{G}'$ takes the form $X - V_1 \rightarrow \dots \rightarrow Y$, and thus, there must be two DAGs in $[\mathcal{G}'] \subseteq [\mathcal{G}]$ with corresponding paths of the forms $X \rightarrow \dots \rightarrow Y$ and $X \leftarrow V_1 \rightarrow \dots \rightarrow Y$.

We first show that $\mathcal{G}'$ contains $X - V_1$ by the contraposition of Lemma 32. Note that we have already shown that $\mathcal{G}$ contains $V_1 - V_2$, that $\mathcal{G}'$ is formed by adding $V_1 \rightarrow V_2$ to $\mathcal{G}$, and that $\mathcal{G}$ contains $X - V_1$. It remains to show that $X, V_1 \notin \mathrm{De}(V_2, \mathcal{G}')$. To see this, note that $\mathcal{G}$ must contain an edge $\langle X, V_2 \rangle$, because V_1 is not of definite status on q. This edge must take the form $X \rightarrow V_2$ by the choice of q and the fact that q is possibly causal. Thus, $\mathcal{G}'$ contains $X \rightarrow V_2$ and $V_1 \rightarrow V_2$. Therefore, $X, V_1 \notin \mathrm{De}(V_2, \mathcal{G}')$. Finally, note that $\mathcal{G}'$ contains $V_1 \rightarrow \dots \rightarrow Y$ by R1 of Meek (1995), since we constructed $\mathcal{G}'$ be adding $V_1 \rightarrow V_2$ to a path $q(V_1, Y)$ that is unshielded and possibly causal. ■

Lemma 38 *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a causal DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$. Then let $\mathcal{D}^* = (\mathbf{V}, \mathbf{E}')$ be a causal DAG constructed by removing edges from $\mathcal{D}$, and let $f(\mathbf{v}|do(\mathbf{x}))$ be an interventional density over $\mathbf{V}$. If $f(\mathbf{v}|do(\mathbf{x}))$ is consistent with $\mathcal{D}^*$, then it is consistent with $\mathcal{D}$.*

Proof of Lemma 38. Suppose that $f(\mathbf{v}|do(\mathbf{x}))$ is consistent with $\mathcal{D}^*$. Then by definition, there exists a set of interventional densities $\mathbf{F}^*$ such that $\mathcal{D}^*$ is compatible with $\mathbf{F}^*$. Let $f(\mathbf{v})$ be the density in $\mathbf{F}^*$ under a null intervention. Note that by the truncated factorization in Equation (3), $f(\mathbf{v})$ is Markov compatible with $\mathcal{D}^*$. Thus by Lemma 19,

$$V_i \perp\!\!\!\perp \Big[\mathbf{V} \setminus \big(\mathrm{De}(V_i, \mathcal{D}^*) \cup \mathrm{Pa}(V_i, \mathcal{D}^*)\big)\Big] \,|\, \mathrm{Pa}(V_i, \mathcal{D}^*) \tag{15}$$

for all $V_i \in \mathbf{V}$, where $\perp\!\!\!\perp$ indicates independence with respect to $f(\mathbf{v})$. Further, since $\mathrm{De}(V_i, \mathcal{D}^*) \subseteq \mathrm{De}(V_i, \mathcal{D})$, then $\mathrm{De}(V_i, \mathcal{D}^*) \cap \mathrm{Pa}(V_i, \mathcal{D}) = \emptyset$ and thus $\mathrm{Pa}(V_i, \mathcal{D}) \subseteq \mathbf{V} \setminus \mathrm{De}(V_i, \mathcal{D}^*)$. Therefore it follows from (15) that

$$V_i \perp\!\!\!\perp \Big[\mathrm{Pa}(V_i, \mathcal{D}) \setminus \mathrm{Pa}(V_i, \mathcal{D}^*)\Big] \;\Big|\; \mathrm{Pa}(V_i, \mathcal{D}^*). \tag{16}$$

Let $f(\mathbf{v}|do(\mathbf{x}'))$, $\mathbf{X}' \subseteq \mathbf{V}$, be an arbitrary density in $\mathbf{F}^*$. Then by definition and (16)

$$\begin{aligned} f(\mathbf{v}|do(\mathbf{x}')) &= \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}'} f(v_i|\,\mathrm{pa}(v_i, \mathcal{D}^*)) \mathbb{1}(\mathbf{X}' = \mathbf{x}') \\ &= \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}'} f(v_i|\,\mathrm{pa}(v_i, \mathcal{D})) \mathbb{1}(\mathbf{X}' = \mathbf{x}'). \end{aligned}$$

Since $f(\mathbf{v}|do(\mathbf{x}'))$ was arbitrary, this holds for all densities in $\mathbf{F}^*$. Thus, $\mathcal{D}$ is compatible with $\mathbf{F}^*$. Since $f(\mathbf{v}|do(\mathbf{x})) \in \mathbf{F}^*$, then by definition, it is consistent with $\mathcal{D}$. ■

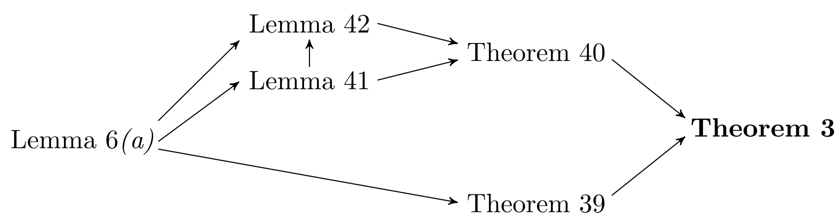

Figure 5: Proof structure of Theorem 3.

D PROOFS FOR SECTION 3.1: MPDAGS - CONDITIONAL ADJUSTMENT CRITERION

The following results show the completeness and soundness of the conditional adjustment criterion for identifying conditional adjustment sets in DAGs. We rely on these results to show the analogous results for MPDAGs in Theorem 3 of Section 3.1. Figure 5 shows how the results in this paper fit together to prove Theorem 3. Two supporting results needed for the proof of soundness in DAGs follow the main results below.

D.1 Main Results

Proof of Lemma 6. *(a)* Follows from Lemma 26. *(b)* Holds by Theorem 3, Lemma 6*(a)*, and Theorem 31. ■

Theorem 39 (**Completeness of the Conditional Adjustment Criterion for DAGs**) *Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{S}$ be pairwise disjoint node sets in a causal DAG $\mathcal{D}$, where $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$. If $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$ (Definition 1), then $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$ (Definition 2).*

Proof of Theorem 39. Let $\mathbf{S}$ be a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$, and let f be a density consistent with $\mathcal{D}$. We start by showing that $\mathbf{S} \cup \mathbf{Z}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$. To do this, we calculate the following. (Justification for the numbered equations is below.)

$$\begin{aligned}
f(\mathbf{y}|do(\mathbf{x})) &= \int_{\mathbf{z}} f(\mathbf{y}, \mathbf{z}|do(\mathbf{x}))\,\mathrm{d}\mathbf{z} \\
&= \int_{\mathbf{z}} f(\mathbf{z}|do(\mathbf{x}))f(\mathbf{y}|do(\mathbf{x}), \mathbf{z})\,\mathrm{d}\mathbf{z} \\
&= \int_{\mathbf{z}} f(\mathbf{z})f(\mathbf{y}|do(\mathbf{x}), \mathbf{z})\,\mathrm{d}\mathbf{z} && (17) \\
&= \int_{\mathbf{z}} f(\mathbf{z}) \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s})f(\mathbf{s}|\mathbf{z})\,\mathrm{d}\mathbf{s}\,\mathrm{d}\mathbf{z} && (18) \\
&= \int_{\mathbf{s},\mathbf{z}} f(\mathbf{y}|\mathbf{x}, \mathbf{s}, \mathbf{z})f(\mathbf{s}, \mathbf{z})\,\mathrm{d}\mathbf{s}\,\mathrm{d}\mathbf{z}.
\end{aligned}$$

Equation (17) follows from Rule 3 of the do calculus (Equation (14)). To show that this rule holds, let p be an arbitrary path from $\mathbf{X}$ to $\mathbf{Z}$ in $\mathcal{D}_{\overline{\mathbf{X}}}$. Note that p must begin with an edge out of $\mathbf{X}$. Since $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{G}) = \emptyset$, p cannot be causal and, therefore, must have colliders. Thus, p is blocked, and so $(\mathbf{Z} \perp_d \mathbf{X})_{\mathcal{D}_{\overline{\mathbf{X}}}}$. Equation (18) follows from the fact that $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$. This shows that $\mathbf{S} \cup \mathbf{Z}$ is an adjustment set relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$.

By Theorem 31, $\mathbf{S} \cup \mathbf{Z}$ satisfies the adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$. Then by Lemma 6*(a)*, $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$. ■

Theorem 40 (**Soundness of the Conditional Adjustment Criterion for DAGs**) *Let $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, and $\mathbf{S}$ be pairwise disjoint node sets in a causal DAG $\mathcal{D}$, where $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$. If $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$ (Definition 2), then $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$ (Definition 1).*

Proof of Theorem 40. This theorem is analogous to Theorem 58 of Perković et al. (2018) for the adjustment criterion. We use the same proof strategy and adapt the arguments to suit our needs.

Suppose that $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$ and let f be a density consistent with $\mathcal{D}$. Our goal is to prove that

$$f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) = \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z}) \, \mathrm{d}\mathbf{s}. \tag{19}$$

We consider three cases below. Before this, we prove an equality that holds in all cases. Let $\mathbf{Y_D} = \mathbf{Y} \cap \mathrm{De}(\mathbf{X}, \mathcal{D})$ and $\mathbf{Y_N} = \mathbf{Y} \setminus \mathrm{De}(\mathbf{X}, \mathcal{D})$. Then $\mathbf{Y_N} \perp_d \mathbf{X} \mid \mathbf{Z}$ in $\mathcal{D}_{\overline{\mathbf{X}}}$, since $\mathcal{D}_{\overline{\mathbf{X}}}$ does not contain edges into $\mathbf{X}$ and since all paths from $\mathbf{X}$ to $\mathbf{Y_N}$ that start with an edge out of $\mathbf{X}$ in $\mathcal{D}_{\overline{\mathbf{X}}}$ contain a collider – a collider that cannot be an element of $\mathrm{An}(\mathbf{Z}, \mathcal{D})$ since $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$. Rule 3 of the do calculus (Equation (14)) then implies

$$f(\mathbf{y_N}|do(\mathbf{x}), \mathbf{z}) = f(\mathbf{y_N}|\mathbf{z}). \tag{20}$$

Case 1: Assume that $\mathbf{Y_D} = \emptyset$ so that $\mathbf{Y} = \mathbf{Y_N}$. Then we have the following. (Justification for the numbered equations is below.)

$$\begin{aligned} f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) &= f(\mathbf{y}|\mathbf{z}) && (21) \\ &= \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z}) \, \mathrm{d}\mathbf{s} \\ &= \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z}) \, \mathrm{d}\mathbf{s}. && (22) \end{aligned}$$

Equation (21) follows from Equation (20) and $\mathbf{Y} = \mathbf{Y_N}$. Equation (22) follows from the following logic. Since $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$ and since $\mathbf{Y} = \mathbf{Y_N}$, it holds that $\mathbf{S} \cup \mathbf{Z}$ blocks all paths from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{D}$. Thus, $\mathbf{X} \perp_d \mathbf{Y} \mid \mathbf{S} \cup \mathbf{Z}$ in $\mathcal{D}$, which implies the analogous independence statement.

Case 2: Assume $\mathbf{Y_N} = \emptyset$ so that $\mathbf{Y} = \mathbf{Y_D}$. Define $\mathbf{H} = \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D}) \setminus (\mathrm{De}(\mathbf{X}, \mathcal{D}) \cup \mathbf{Y} \cup \mathbf{Z})$, $\tilde{\mathbf{S}} = \mathbf{S} \cup \mathbf{H}$, $\tilde{\mathbf{S}}_\mathbf{D} = \tilde{\mathbf{S}} \cap \mathrm{De}(\mathbf{X}, \mathcal{D})$, and $\tilde{\mathbf{S}}_\mathbf{N} = \tilde{\mathbf{S}} \setminus \mathrm{De}(\mathbf{X}, \mathcal{D})$. Then we have the following. (Justification for the numbered equations is below.)

$$\begin{aligned} f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) &= \int_{\tilde{\mathbf{s}}_\mathbf{N}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}_\mathbf{N}) f(\tilde{\mathbf{s}}_\mathbf{N}|\mathbf{z}) \, \mathrm{d}\tilde{\mathbf{s}}_\mathbf{N} && (23) \\ &= \int_{\tilde{\mathbf{s}}_\mathbf{N}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}_\mathbf{N}) \int_{\tilde{\mathbf{s}}_\mathbf{D}} f(\tilde{\mathbf{s}}_\mathbf{D}, \tilde{\mathbf{s}}_\mathbf{N}|\mathbf{z}) \, \mathrm{d}\tilde{\mathbf{s}}_\mathbf{D} \, \mathrm{d}\tilde{\mathbf{s}}_\mathbf{N} \\ &= \int_{\tilde{\mathbf{s}}_\mathbf{D}, \tilde{\mathbf{s}}_\mathbf{N}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}_\mathbf{N}) f(\tilde{\mathbf{s}}_\mathbf{D}, \tilde{\mathbf{s}}_\mathbf{N}|\mathbf{z}) \, \mathrm{d}\tilde{\mathbf{s}}_\mathbf{D} \, \mathrm{d}\tilde{\mathbf{s}}_\mathbf{N} && (24) \\ &= \int_{\tilde{\mathbf{s}}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}|\mathbf{z}) \, \mathrm{d}\tilde{\mathbf{s}} && (25) \\ &= \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z}) \, \mathrm{d}\mathbf{s}. && (26) \end{aligned}$$

Equation (23) holds since by Lemma 42(iv), $\tilde{\mathbf{S}}_\mathbf{N}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$. Equation (24) holds since $\tilde{\mathbf{S}}_\mathbf{D}$ is disjoint from $\mathbf{Y} \cup \mathbf{X} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z}$. Equation (25) holds since by Lemma 42(iii), we have $\mathbf{Y} \perp_d \tilde{\mathbf{S}}_\mathbf{D} \mid \mathbf{X} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z}$ in $\mathcal{D}$, where the analogous independence statement follows. Finally, Equation (26) results from applying Lemma 41(ii).

Case 3: Assume $\mathbf{Y_D} \neq \emptyset$ and $\mathbf{Y_N} \neq \emptyset$ and define $\mathbf{H}, \tilde{\mathbf{S}}, \tilde{\mathbf{S}}_\mathbf{D}$, and $\tilde{\mathbf{S}}_\mathbf{N}$ as in Case 2 above. We start by showing two equalities that rely on the do calculus. First note that by Lemma 42(vi), $\mathbf{Y_D} \perp_d \mathbf{Y_N} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z} \mid \mathbf{X}$ in $\mathcal{D}_{\overline{\mathbf{X}}\underline{\mathbf{Y_N} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z}}}$. Thus by Rule 2 of the do calculus (Equation (13)), we have that

$$f(\mathbf{y_D}|do(\mathbf{x}), \mathbf{y_N}, \mathbf{z}, \tilde{\mathbf{s}}_\mathbf{N}) = f(\mathbf{y_D}|do(\mathbf{x}, \mathbf{y_N}, \mathbf{z}, \tilde{\mathbf{s}}_\mathbf{N})). \tag{27}$$

Second, note by Lemma 42(vii), $\tilde{\mathbf{S}}_{\mathbf{N}} \perp_d \mathbf{X} \mid \mathbf{Y}_{\mathbf{N}} \cup \mathbf{Z}$ in $\mathcal{D}_{\overline{\mathbf{X}}}$. Thus by Rule 3 of the do calculus (Equation (14)), we have that

$$f(\tilde{\mathbf{s}}_{\mathbf{N}}|do(\mathbf{x}), \mathbf{y}_{\mathbf{N}}, \mathbf{z}) = f(\tilde{\mathbf{s}}_{\mathbf{N}}|\mathbf{y}_{\mathbf{N}}, \mathbf{z}). \tag{28}$$

Then we have the following. (Justification for the numbered equations is below.)

$$\begin{aligned}
f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) &= \int_{\tilde{\mathbf{s}}_{\mathbf{N}}} f(\mathbf{y}, \tilde{\mathbf{s}}_{\mathbf{N}}|do(\mathbf{x}), \mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{N}} \\
&= \int_{\tilde{\mathbf{s}}_{\mathbf{N}}} f(\mathbf{y}_{\mathbf{D}}|\tilde{\mathbf{s}}_{\mathbf{N}}, \mathbf{y}_{\mathbf{N}}, do(\mathbf{x}), \mathbf{z}) f(\tilde{\mathbf{s}}_{\mathbf{N}}|\mathbf{y}_{\mathbf{N}}, do(\mathbf{x}), \mathbf{z}) f(\mathbf{y}_{\mathbf{N}}|do(\mathbf{x}), \mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{N}} \\
&= \int_{\tilde{\mathbf{s}}_{\mathbf{N}}} f(\mathbf{y}_{\mathbf{D}}|do(\mathbf{x}, \mathbf{y}_{\mathbf{N}}, \mathbf{z}, \tilde{\mathbf{s}}_{\mathbf{N}})) f(\tilde{\mathbf{s}}_{\mathbf{N}}|\mathbf{y}_{\mathbf{N}}, \mathbf{z}) f(\mathbf{y}_{\mathbf{N}}|\mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{N}} \qquad (29) \\
&= \int_{\tilde{\mathbf{s}}_{\mathbf{N}}} f(\mathbf{y}_{\mathbf{D}}|do(\mathbf{x}, \mathbf{y}_{\mathbf{N}}, \mathbf{z}, \tilde{\mathbf{s}}_{\mathbf{N}})) \int_{\tilde{\mathbf{s}}_{\mathbf{D}}} f(\tilde{\mathbf{s}}_{\mathbf{N}}, \mathbf{y}_{\mathbf{N}}, \tilde{\mathbf{s}}_{\mathbf{D}}|\mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{D}}\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{N}} \\
&= \int_{\tilde{\mathbf{s}}_{\mathbf{N}}} f(\mathbf{y}_{\mathbf{D}}|do(\mathbf{x}, \mathbf{y}_{\mathbf{N}}, \mathbf{z}, \tilde{\mathbf{s}}_{\mathbf{N}})) \int_{\tilde{\mathbf{s}}_{\mathbf{D}}} f(\mathbf{y}_{\mathbf{N}}|\tilde{\mathbf{s}}, \mathbf{z}) f(\tilde{\mathbf{s}}|\mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{D}}\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{N}} \\
&= \int_{\tilde{\mathbf{s}}_{\mathbf{N}}} f(\mathbf{y}_{\mathbf{D}}|\mathbf{y}_{\mathbf{N}}, \mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}_{\mathbf{N}}) \int_{\tilde{\mathbf{s}}_{\mathbf{D}}} f(\mathbf{y}_{\mathbf{N}}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}|\mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{D}}\, \mathrm{d}\tilde{\mathbf{s}}_{\mathbf{N}} \qquad (30) \\
&= \int_{\tilde{\mathbf{s}}} f(\mathbf{y}_{\mathbf{D}}|\mathbf{y}_{\mathbf{N}}, \mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}) f(\mathbf{y}_{\mathbf{N}}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}|\mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}} \qquad (31) \\
&= \int_{\tilde{\mathbf{s}}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}|\mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}} \\
&= \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z})\, \mathrm{d}\mathbf{s}. \qquad (32)
\end{aligned}$$

Equation (29) holds by the applying Equations (27), (28), and (20). Equation (30) holds by the following logic. By Lemma 42(v), the empty set is an adjustment set relative to $(\mathbf{X} \cup \mathbf{Y}_{\mathbf{N}} \cup \tilde{\mathbf{S}}_{\mathbf{N}} \cup \mathbf{Z}, \mathbf{Y}_{\mathbf{D}})$ in $\mathcal{D}$. Then by Lemma 41(i), $\tilde{\mathbf{S}}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$, and so $\tilde{\mathbf{S}} \cup \mathbf{Z}$ blocks all paths from $\mathbf{X}$ to $\mathbf{Y}_{\mathbf{N}}$ in $\mathcal{D}$. Thus, $\mathbf{Y}_{\mathbf{N}} \perp_d \mathbf{X} \mid \tilde{\mathbf{S}} \cup \mathbf{Z}$ in $\mathcal{D}$, where the analogous independence statement follows.

Equation (31) holds since $\tilde{\mathbf{S}}_{\mathbf{D}}$ is disjoint from $\mathbf{Y} \cup \mathbf{X} \cup \tilde{\mathbf{S}}_{\mathbf{N}} \cup \mathbf{Z}$ and since by Lemma 42(iii), we have that $\mathbf{Y}_{\mathbf{D}} \perp_d \tilde{\mathbf{S}}_{\mathbf{D}} \mid \mathbf{Y}_{\mathbf{N}} \cup \mathbf{X} \cup \tilde{\mathbf{S}}_{\mathbf{N}} \cup \mathbf{Z}$ in $\mathcal{D}$, where the analogous independence statement follows. Finally, Equation (32) results from applying Lemma 41(ii). ■

D.2 Supporting Results

Lemma 41 *Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{S}$ be pairwise disjoint node sets in a causal DAG $\mathcal{D}$, where $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$ and where $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$ (Definition 2). Let $\mathbf{H} \subseteq \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D}) \setminus (\mathrm{De}(\mathbf{X}, \mathcal{D}) \cup \mathbf{Y} \cup \mathbf{Z})$ and $\tilde{\mathbf{S}} = \mathbf{S} \cup \mathbf{H}$. Then:*

(i) $\tilde{\mathbf{S}}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$, and

(ii) $\int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z})\, \mathrm{d}\mathbf{s} = \int_{\tilde{\mathbf{s}}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}|\mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}}$, for any density f consistent with $\mathcal{D}$.

Proof of Lemma 41. This lemma is analogous to Lemma 59 of Perković et al. (2018) (Lemma 34). We use the same proof strategy and adapt the arguments to suit our needs.

(i) By Lemma 6*(a)*, since $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$, then $\mathbf{S} \cup \mathbf{Z}$ satisfies the adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$. Then by Lemma 34, $\tilde{\mathbf{S}} \cup \mathbf{Z}$ satisfies the adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$. The statement follows by a second use of Lemma 6*(a)*.

(ii) Let f be an arbitrary density consistent with $\mathcal{D}$. We proceed with a proof by induction.

Base case: Suppose $\mathbf{H} = \{H\}$ so that $|\mathbf{H}| = 1$. When $H \in \mathbf{S}$, the claim clearly holds. Thus, we let $H \notin \mathbf{S}$. Note that the claim holds if either $\mathbf{Y} \perp_d H \mid \mathbf{X} \cup \mathbf{S} \cup \mathbf{Z}$ or $\mathbf{X} \perp_d H \mid \mathbf{S} \cup \mathbf{Z}$ in $\mathcal{D}$. To see this, we calculate the following.

(a) When $(\mathbf{Y} \perp_d H \mid \mathbf{X} \cup \mathbf{S} \cup \mathbf{Z})_\mathcal{D}$, then

$$\begin{aligned}\int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x},\mathbf{z},\mathbf{s}) f(\mathbf{s}|\mathbf{z})\,\mathrm{d}\mathbf{s} &= \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x},\mathbf{z},\mathbf{s}) \int_h f(\mathbf{s},h|\mathbf{z})\,\mathrm{d}h\,\mathrm{d}\mathbf{s} \\ &= \int_{\mathbf{s},h} f(\mathbf{y}|\mathbf{x},\mathbf{z},\mathbf{s}) f(\mathbf{s},h|\mathbf{z})\,\mathrm{d}\mathbf{s}\,\mathrm{d}h \\ &= \int_{\tilde{\mathbf{s}}} f(\mathbf{y}|\mathbf{x},\mathbf{z},\tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}|\mathbf{z})\,\mathrm{d}\tilde{\mathbf{s}},\end{aligned}$$

where the second equality holds since $H \notin \mathbf{Y} \cup \mathbf{X} \cup \mathbf{S} \cup \mathbf{Z}$.

(b) When $(\mathbf{X} \perp_d H \mid \mathbf{S} \cup \mathbf{Z})_\mathcal{D}$, then

$$\begin{aligned}\int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x},\mathbf{z},\mathbf{s}) f(\mathbf{s}|\mathbf{z})\,\mathrm{d}\mathbf{s} &= \int_{\mathbf{s}} f(\mathbf{s}|\mathbf{z}) \int_h f(\mathbf{y},h|\mathbf{x},\mathbf{z},\mathbf{s})\,\mathrm{d}h\,\mathrm{d}\mathbf{s} \\ &= \int_{\mathbf{s},h} f(\mathbf{y},h|\mathbf{x},\mathbf{z},\mathbf{s}) f(\mathbf{s}|\mathbf{z})\,\mathrm{d}\mathbf{s}\,\mathrm{d}h \\ &= \int_{\mathbf{s},h} f(\mathbf{y}|\mathbf{x},\mathbf{z},\mathbf{s},h) f(h|\mathbf{x},\mathbf{z},\mathbf{s}) f(\mathbf{s}|\mathbf{z})\,\mathrm{d}\mathbf{s}\,\mathrm{d}h \\ &= \int_{\mathbf{s},h} f(\mathbf{y}|\mathbf{x},\mathbf{z},\mathbf{s},h) f(h|\mathbf{z},\mathbf{s}) f(\mathbf{s}|\mathbf{z})\,\mathrm{d}\mathbf{s}\,\mathrm{d}h \\ &= \int_{\tilde{\mathbf{s}}} f(\mathbf{y}|\mathbf{x},\mathbf{z},\tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}|\mathbf{z})\,\mathrm{d}\tilde{\mathbf{s}},\end{aligned}$$

where the second equality holds since $H \notin \mathbf{S} \cup \mathbf{Z}$.

We use the remainder of the base case to show that *(a)* or *(b)* must hold. For sake of contradiction, suppose that neither hold. This implies that there are two paths in $\mathcal{D}$: one from $\mathbf{X}$ to H that is d-connecting given $\mathbf{S} \cup \mathbf{Z}$ and one from $\mathbf{Y}$ to H that is d-connecting given $\mathbf{X} \cup \mathbf{S} \cup \mathbf{Z}$. Let $p = \langle X, \dots, H\rangle$, $X \in \mathbf{X}$, and $q = \langle H, \dots, Y\rangle$, $Y \in \mathbf{Y}$, be such paths, respectively, where p is proper. In the arguments below, we use paths related to p and q – in the proper back-door graph $\mathcal{D}^{pbd}_{\mathbf{XY}}$ (see Definition 14) and in four of its moral induced subgraphs (see Definition 15) – before applying Theorems 23 and 24 to reach our final contradiction (that $\mathbf{S}$ cannot satisfy the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$).

First, we claim that both p and q are d-connecting given $\mathbf{S} \cup \mathbf{Z}$. This holds for p by definition. For sake of contradiction, suppose that q is blocked by $\mathbf{S} \cup \mathbf{Z}$. Since q is d-connecting given $\mathbf{X} \cup \mathbf{S} \cup \mathbf{Z}$, it must contain a collider in $\mathrm{An}(\mathbf{X}, \mathcal{D}) \setminus \mathrm{An}(\mathbf{S} \cup \mathbf{Z}, \mathcal{D})$. Let C be the closest collider to Y on q such that $C \in (\mathrm{An}(\mathbf{X}, \mathcal{D}) \setminus \mathrm{An}(\mathbf{S} \cup \mathbf{Z}, \mathcal{D})) \cup \mathbf{X}$, and let $r = \langle C, \dots, X'\rangle$, $X' \in \mathbf{X}$, be a shortest causal path in $\mathcal{D}$ from C to $\mathbf{X}$. Then let V be the node closest to X' on r that is also on $q(C, Y)$, and define the path $t = (-r)(X', V) \oplus q(V, Y)$. Note that t is non-causal since either $(-r)(X', V)$ is of non-zero length or $X' = V = C$, so that t is a path into X'. Further, by the definitions of q, C, and r, we have that t is proper non-causal path from $\mathbf{X}$ to $\mathbf{Y}$ that is d-connecting given $\mathbf{S} \cup \mathbf{Z}$. But this contradicts that $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$.

Next, we prove that the sequence of nodes in $\mathcal{D}^{pbd}_{\mathbf{XY}}$ corresponding to p forms a path. Note that since p is proper, we only need to show that p does not start with an edge $X \to W$, where W is a node that lies on a proper causal path in $\mathcal{D}$ from X to $\mathbf{Y}$. For sake of contradiction, suppose that p starts with $X \to W$ for such a $W \in \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D})$. Note that p cannot be causal from X to H, since $H \notin \mathrm{De}(\mathbf{X}, \mathcal{D})$ by the definition of $\mathbf{H}$. Thus, p is non-causal and there is a collider C' on p such that $C' \in \mathrm{De}(W, \mathcal{D})$. Since p is d-connecting given $\mathbf{S} \cup \mathbf{Z}$ and $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$, then $\mathbf{S} \cap \mathrm{De}(C', \mathcal{D}) \neq \emptyset$. Further, since $\mathrm{De}(C', \mathcal{D}) \subseteq \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D})$, this implies that $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D}) \neq \emptyset$. But this contradicts that $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$.

Similarly, we prove that the sequence of nodes in $\mathcal{D}^{pbd}_{\mathbf{XY}}$ corresponding to q also forms a path. For this, note that all nodes in $\mathbf{X}$ on q must be a colliders on q, since q is d-connecting given $\mathbf{X} \cup \mathbf{S} \cup \mathbf{Z}$. Thus, removing edges out of $\mathbf{X}$ from $\mathcal{D}$ in order to form $\mathcal{D}^{pbd}_{\mathbf{XY}}$ will not affect the edges on q.

Let $\tilde{p}$ and $\tilde{q}$ be the paths in $\mathcal{D}^{pbd}_{\mathbf{XY}}$ corresponding to p and q, respectively. Then for sake of contradiction, suppose either $\tilde{p}$ or $\tilde{q}$ is blocked given $\mathbf{S} \cup \mathbf{Z}$. Since p and q are d-connecting given $\mathbf{S} \cup \mathbf{Z}$, then there must be a node C

on p or q where C is a collider on p or q and every causal path in $\mathcal{D}$ from C to $\mathbf{S} \cup \mathbf{Z}$ contains the first edge of a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{D}$. Let d be an arbitrary such causal path in $\mathcal{D}$ from C to $\mathbf{S} \cup \mathbf{Z}$. Note that d is a path from C to $\mathbf{S}$, since d must contain a node in $\mathbf{X}$ and since $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$. But since d contains the first edge of a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{D}$, this implies that $\mathbf{S} \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D}) \neq \emptyset$, which contradicts that $\mathbf{S}$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$.

We continue the base case by reasoning with four moral induced subgraphs of $\mathcal{D}^{pbd}_{\mathbf{XY}}$ (see Definition 15). Start by defining the following.

$$\begin{aligned}
\mathbf{A}_{\mathbf{XHYSZ}} &= \mathrm{An}(\mathbf{X} \cup \mathbf{H} \cup \mathbf{Y} \cup \mathbf{S} \cup \mathbf{Z}, \mathcal{D}^{pbd}_{\mathbf{XY}}).\\
\mathbf{A}_{\mathbf{XYSZ}} &= \mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{S} \cup \mathbf{Z}, \mathcal{D}^{pbd}_{\mathbf{XY}}).\\
\mathbf{A}_{\mathbf{XHSZ}} &= \mathrm{An}(\mathbf{X} \cup \mathbf{H} \cup \mathbf{S} \cup \mathbf{Z}, \mathcal{D}^{pbd}_{\mathbf{XY}}).\\
\mathbf{A}_{\mathbf{HYSZ}} &= \mathrm{An}(\mathbf{H} \cup \mathbf{Y} \cup \mathbf{S} \cup \mathbf{Z}, \mathcal{D}^{pbd}_{\mathbf{XY}}).
\end{aligned}$$

Then define $\mathcal{D}_{\mathbf{XHYSZ}}$, $\mathcal{D}_{\mathbf{XYSZ}}$, $\mathcal{D}_{\mathbf{XHSZ}}$, and $\mathcal{D}_{\mathbf{HYSZ}}$ to be the moral induced subgraphs of $\mathcal{D}^{pbd}_{\mathbf{XY}}$ on nodes $\mathbf{A}_{\mathbf{XHYSZ}}$, $\mathbf{A}_{\mathbf{XYSZ}}$, $\mathbf{A}_{\mathbf{XHSZ}}$, and $\mathbf{A}_{\mathbf{HYSZ}}$, respectively. In order to use Theorem 24, we want to show that $\mathcal{D}_{\mathbf{XYSZ}}$ contains a path from $\mathbf{X}$ to $\mathbf{Y}$ that does not contain a node in $\mathbf{S} \cup \mathbf{Z}$.

Since $\tilde{p}$ and $\tilde{q}$ are d-connecting given $\mathbf{S} \cup \mathbf{Z}$, then by Theorem 23, the following two paths must exist in $\mathcal{D}_{\mathbf{XHSZ}}$: path a from X to H and path b from H to Y, where neither path contains a node in $\mathbf{S} \cup \mathbf{Z}$. Note that since $\mathbf{A}_{\mathbf{XHSZ}} \subseteq \mathbf{A}_{\mathbf{XHYSZ}}$ and $\mathbf{A}_{\mathbf{HYSZ}} \subseteq \mathbf{A}_{\mathbf{XHYSZ}}$, any path in $\mathcal{D}_{\mathbf{XHSZ}}$ or $\mathcal{D}_{\mathbf{HYSZ}}$ will also be in $\mathcal{D}_{\mathbf{XHYSZ}}$. Further, since $H \in \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D})$ by definition and since we form $\mathcal{D}^{pbd}_{\mathbf{XY}}$ by removing edges out of $\mathbf{X}$ from $\mathcal{D}$, then $H \in \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D}^{pbd}_{\mathbf{XY}})$. Therefore, $\mathbf{A}_{\mathbf{XHYSZ}} = \mathbf{A}_{\mathbf{XYSZ}}$ and $\mathcal{D}_{\mathbf{XHYSZ}} = \mathcal{D}_{\mathbf{XYSZ}}$. Thus, a and b are both paths in $\mathcal{D}_{\mathbf{XYSZ}}$.

We complete the base case by applying Theorems 23 and 24 to show our necessary contradiction. Since we can combine subpaths of a and b to form a path c in $\mathcal{D}_{\mathbf{XYSZ}}$ from $\mathbf{X}$ to $\mathbf{Y}$ that does not contain a node in $\mathbf{S} \cup \mathbf{Z}$, then by Theorem 23, $\mathbf{X}$ and $\mathbf{Y}$ are d-connecting given $\mathbf{S} \cup \mathbf{Z}$ in $\mathcal{D}^{pbd}_{\mathbf{XY}}$. By Theorem 24, this implies that $\mathbf{S} \cup \mathbf{Z}$ does not satisfy the adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$ (see Definition 12). Therefore, by the contraposition of Lemma 6*(a)*, $\mathbf{S}$ does not satisfy the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$, which is a contradiction.

Induction step: Assume that the result holds for $|\mathbf{H}| = k$, $k \in \mathbb{N}$, and let $|\mathbf{H}| = k + 1$. Take an arbitrary $H \in \mathbf{H}$, and define $\mathbf{S}' = \mathbf{S} \cup \{H\}$ and $\mathbf{H}' = \mathbf{H} \setminus \{H\}$. Since the base case holds and since $\{H\} \subseteq \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D}) \setminus (\mathrm{De}(\mathbf{X}, \mathcal{D}) \cup \mathbf{Y} \cup \mathbf{Z})$, then

$$\begin{aligned}
\int_{\mathbf{s}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}) f(\mathbf{s}|\mathbf{z})\, \mathrm{d}\mathbf{s} &= \int_{\mathbf{s},h} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}, h) f(\mathbf{s}, h|\mathbf{z})\, \mathrm{d}\mathbf{s}\, \mathrm{d}h\\
&= \int_{\mathbf{s}'} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}') f(\mathbf{s}'|\mathbf{z})\, \mathrm{d}\mathbf{s}'. \qquad (33)
\end{aligned}$$

Further, by part (i), $\mathbf{S}'$ satisfies the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$. Since $\mathbf{H}' \subseteq \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D}) \setminus (\mathrm{De}(\mathbf{X}, \mathcal{D}) \cup \mathbf{Y} \cup \mathbf{Z})$ and $|\mathbf{H}'| = k$, then by the induction assumption,

$$\begin{aligned}
\int_{\mathbf{s}'} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}') f(\mathbf{s}'|\mathbf{z})\, \mathrm{d}\mathbf{s}' &= \int_{\mathbf{s}',\mathbf{h}'} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \mathbf{s}', \mathbf{h}') f(\mathbf{s}', \mathbf{h}'|\mathbf{z})\, \mathrm{d}\mathbf{s}'\, \mathrm{d}\mathbf{h}'\\
&= \int_{\tilde{\mathbf{s}}} f(\mathbf{y}|\mathbf{x}, \mathbf{z}, \tilde{\mathbf{s}}) f(\tilde{\mathbf{s}}|\mathbf{z})\, \mathrm{d}\tilde{\mathbf{s}}. \qquad (34)
\end{aligned}$$

Combining (33) and (34) yields the desired result. ■

Lemma 42 *Let* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, *and* $\mathbf{S}$ *be pairwise disjoint node sets in a causal DAG* $\mathcal{D}$, *where* $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$ *and where* $\mathbf{S}$ *satisfies the conditional adjustment criterion relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{D}$ *(Definition 2). Let* $\mathbf{H} = \mathrm{An}(\mathbf{X} \cup \mathbf{Y}, \mathcal{D}) \setminus (\mathrm{De}(\mathbf{X}, \mathcal{D}) \cup \mathbf{Y} \cup \mathbf{Z})$ *and* $\tilde{\mathbf{S}} = \mathbf{S} \cup \mathbf{H}$. *Additionally, let* $\tilde{\mathbf{S}}_{\mathbf{D}} = \tilde{\mathbf{S}} \cap \mathrm{De}(\mathbf{X}, \mathcal{D})$, $\tilde{\mathbf{S}}_{\mathbf{N}} = \tilde{\mathbf{S}} \setminus \mathrm{De}(\mathbf{X}, \mathcal{D})$, $\mathbf{Y}_{\mathbf{D}} = \mathbf{Y} \cap \mathrm{De}(\mathbf{X}, \mathcal{D})$, *and* $\mathbf{Y}_{\mathbf{N}} = \mathbf{Y} \setminus \mathrm{De}(\mathbf{X}, \mathcal{D})$. *Then the following statements hold:*

(i) $(\mathbf{X} \cup \mathbf{Y}_{\mathbf{N}} \cup \tilde{\mathbf{S}} \cup \mathbf{Z}) \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D}) = \emptyset$,

(ii) *if* $p = \langle H, \dots, Y_D \rangle$ *is a non-causal path in* $\mathcal{D}$ *from* $H \in \mathbf{X} \cup \mathbf{Y_N} \cup \tilde{\mathbf{S}} \cup \mathbf{Z}$ *to a node* $Y_D \in \mathbf{Y_D}$*, then* p *is blocked by* $(\mathbf{X} \cup \mathbf{Y_N} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z}) \setminus \{H\}$*,*

(iii) $\mathbf{Y_D} \perp_d \tilde{\mathbf{S}}_\mathbf{D} \mid \mathbf{Y_N} \cup \mathbf{X} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z}$ *in* $\mathcal{D}$*,*

(iv) *if* $\mathbf{Y_N} = \emptyset$*, then* $\tilde{\mathbf{S}}_\mathbf{N}$ *is a conditional adjustment set relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{D}$ *(Definition 1),*

(v) *the empty set is an adjustment set relative to* $(\mathbf{X} \cup \mathbf{Y_N} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z}, \mathbf{Y_D})$ *in* $\mathcal{D}$ *(Definition 11),*

(vi) $\mathbf{Y_D} \perp_d (\mathbf{Y_N} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z}) \mid \mathbf{X}$ *in* $\mathcal{D}_{\overline{\mathbf{X}}\underline{\mathbf{Y_N} \cup \tilde{\mathbf{S}}_\mathbf{N} \cup \mathbf{Z}}}$*, and*

(vii) $\tilde{\mathbf{S}}_\mathbf{N} \perp_d \mathbf{X} \mid \mathbf{Y_N} \cup \mathbf{Z}$ *in* $\mathcal{D}_{\overline{\mathbf{X}}}$*.*

Proof of Lemma 42. This lemma is analogous to Lemma 60 of Perković et al. (2018) (Lemma 35), which is needed for adjustment in total effect identification. We rely on this result in the proof below.

Note that $\mathbf{X}, \mathbf{Y}$, and $\mathbf{S} \cup \mathbf{Z}$ are pairwise disjoint node sets in $\mathcal{D}$, where by Lemma 6*(a)*, $\mathbf{S} \cup \mathbf{Z}$ satisfies the adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$. Results (i)-(iii) and (vi) follow directly from Lemma 35. Result (v) follows additionally from Theorem 25. Result (iv) follows additionally from Theorem 25 and Lemma 6*(a)*.

(vii) Let p be an arbitrary path from $X \in \mathbf{X}$ to $\tilde{\mathbf{S}}_\mathbf{N}$ in $\mathcal{D}_{\overline{\mathbf{X}}}$. By definition of $\mathcal{D}_{\overline{\mathbf{X}}}$, p begins with an edge out of X. Since, by definition, $\tilde{\mathbf{S}}_\mathbf{N} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$, where $\mathrm{De}(\mathbf{X}, \mathcal{D}_{\overline{\mathbf{X}}}) \subseteq \mathrm{De}(\mathbf{X}, \mathcal{D})$, then p must contain at least one collider. Let $\mathbf{C}$ be the set containing the closest collider to X on p and its descendants in $\mathcal{D}_{\overline{\mathbf{X}}}$. Note that $\mathbf{C} \subseteq \mathrm{De}(\mathbf{X}, \mathcal{D}_{\overline{\mathbf{X}}}) \subseteq \mathrm{De}(\mathbf{X}, \mathcal{D})$. By definition of $\mathbf{Y_N}$ and by assumption, $(\mathbf{Y_N} \cup \mathbf{Z}) \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$, and thus, p is blocked by $\mathbf{Y_N} \cup \mathbf{Z}$. ∎

E CONDITIONAL BACK-DOOR CRITERION

This section extends Pearl's back-door criterion (2009) to the context of estimating a conditional causal effect in a DAG. Definition 43 provides the extended criterion, and Lemma 44 establishes that this criterion is sufficient for conditional adjustment. Lemma 45 makes a comparison between this criterion and the generalized back-door criterion of Maathuis and Colombo (2015) (Definition 13).

Definition 43 (**Conditional Back-door Criterion for DAGs**) *Let* $\mathbf{X}$*,* $\mathbf{Y}$*,* $\mathbf{Z}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a DAG* $\mathcal{D}$*, where* $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$*. Then* $\mathbf{S}$ *satisfies the conditional back-door criterion relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{D}$ *if*

(a) $\mathbf{S} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$*, and*

(b) $\mathbf{S} \cup \mathbf{Z}$ *blocks all proper back-door paths from* $\mathbf{X}$ *to* $\mathbf{Y}$*.*

Lemma 44 *Let* $\mathbf{X}$*,* $\mathbf{Y}$*,* $\mathbf{Z}$*, and* $\mathbf{S}$ *be pairwise disjoint node sets in a causal DAG* $\mathcal{D}$*, where* $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$*. If* $\mathbf{S}$ *satisfies the conditional back-door criterion relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{D}$ *(Definition 43), then* $\mathbf{S}$ *is a conditional adjustment set relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{D}$ *(Definition 1).*

Proof of Lemma 44. Let $\mathbf{S}$ be a set that satisfies the conditional back-door criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$, and let f be a density consistent with $\mathcal{D}$. Then

$$
\begin{aligned}
f(\mathbf{y}|do(\mathbf{x}), \mathbf{z}) &= \int_{\mathbf{s}} f(\mathbf{y}, \mathbf{s}|do(\mathbf{x}), \mathbf{z})\, \mathrm{d}\mathbf{s} \\
&= \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{s}, do(\mathbf{x}), \mathbf{z}) f(\mathbf{s}|do(\mathbf{x}), \mathbf{z})\, \mathrm{d}\mathbf{s} \\
&= \int_{\mathbf{s}} f(\mathbf{y}|\mathbf{s}, \mathbf{x}, \mathbf{z}) f(\mathbf{s}|\mathbf{z})\, \mathrm{d}\mathbf{s}
\end{aligned}
$$

The first two equalities follow from the law of total probability and the chain rule. The third equality follows from Rules 2 and 3 of the do calculus (Equations (13) and (14)) and the d-separations shown below.

In order to use Rule 2 to conclude that $f(\mathbf{y}|\mathbf{s}, do(\mathbf{x}), \mathbf{z}) = f(\mathbf{y}|\mathbf{s}, \mathbf{x}, \mathbf{z})$, we show that $(\mathbf{Y} \perp_d \mathbf{X} \mid \mathbf{S} \cup \mathbf{Z})_{\mathcal{D}_{\underline{\mathbf{X}}}}$. Note that $\mathcal{D}_{\underline{\mathbf{X}}}$ only contains back-door paths from $\mathbf{X}$ to $\mathbf{Y}$. So every path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{D}_{\underline{\mathbf{X}}}$ contains a proper back-door path from $\mathbf{X}$ to $\mathbf{Y}$ as a subpath. Since $\mathbf{S} \cup \mathbf{Z}$ blocks all proper back-door paths from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{D}$, the d-separation holds.

In order to use Rule 3 to conclude that $f(\mathbf{s}|do(\mathbf{x}), \mathbf{z}) = f(\mathbf{s}|\mathbf{z})$, we show that $(\mathbf{S} \perp_d \mathbf{X} \mid \mathbf{Z})_{\mathcal{D}_{\overline{\mathbf{X}(\mathbf{Z})}}}$. This follows from the assumptions that $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$ and $\mathbf{S} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$. ■

Lemma 45 (**Comparison of Back-door Criteria for DAGs**) *Let* $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, *and* $\mathbf{S}$ *be pairwise disjoint node sets in a DAG* $\mathcal{D}$, *where* $\mathbf{Z} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$. *Then* $\mathbf{S}$ *satisfies the conditional back-door criterion relative to* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{D}$ *(Definition 43) if and only if* $\mathbf{S} \cup \mathbf{Z}$ *satisfies the generalized back-door criterion relative to* $(\mathbf{X}, \mathbf{Y})$ *in* $\mathcal{D}$ *(Definition 13).*

Proof of Lemma 45. $\Leftarrow$: Follows immediately.

$\Rightarrow$: Since $\mathbf{S}$ satisfies the conditional back-door criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$, then $\mathbf{S} \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$. Combining this with our assumptions gives us that $(\mathbf{S} \cup \mathbf{Z}) \cap \mathrm{De}(\mathbf{X}, \mathcal{D}) = \emptyset$. In the remainder of the proof, we show that $\mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$ blocks all back-door paths from $\mathbf{X}$ to $\mathbf{Y}$. The result follows by Definition 13.

Let p_1 be an arbitrary back-door path from $X \in \mathbf{X}$ to $Y \in \mathbf{Y}$ in $\mathcal{D}$. For sake of contradiction, suppose that p_1 is d-connecting given $\mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$. Let X_C be the node in $\mathbf{X}$ closest to Y on p_1, and let $p_2 = p_1(X_C, Y)$. Note that p_2 is proper. When $X_C = X$, then $p_2 = p_1$ is a back-door path. When $X_C \neq X$, then because p_1 is d-connecting given $\mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$, we have that X_C is a collider on p_1, and therefore, p_2 is again a back-door path. Thus, p_2 is a proper back-door path from $\mathbf{X}$ to $\mathbf{Y}$ that, by assumption, must be blocked by $\mathbf{S} \cup \mathbf{Z}$.

Let A be the node on p_2 immediately following X_C. That is, p_2 contains $X_C \leftarrow A$. Note that since p_2 is blocked given $\mathbf{S} \cup \mathbf{Z}$, then $A \neq Y$. Thus, we consider the path $p_3 = p_2(A, Y)$. Since p_1 is d-connecting given $\mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$, where A is a non-collider on p_1, then $A \notin \mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$ and thus, p_3 is also d-connecting given $\mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$. Similarly, since p_2 is blocked given $\mathbf{S} \cup \mathbf{Z}$, where A is not a collider on p_2 and $A \notin \mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$, then p_3 is also blocked by $\mathbf{S} \cup \mathbf{Z}$.

Since p_3 is d-connecting given $\mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$ and blocked given $\mathbf{S} \cup \mathbf{Z}$, then p_3 must contain at least one collider in $\mathrm{An}(\mathbf{X} \setminus \{X\}, \mathcal{D}) \setminus \mathrm{An}(\mathbf{S} \cup \mathbf{Z}, \mathcal{D})$. Let C be the closest such collider to Y on p_3 and let $r = \langle C, \dots, X' \rangle$, $X' \in \mathbf{X}$, be a shortest causal path from C to $\mathbf{X}$ in $\mathcal{D}$. While there must be a causal path from C to $\mathbf{X} \setminus \{X\}$ in $\mathcal{D}$, note that r need not be one, and thus, we allow for the possibility that $X' = X$.

Let B be the node closest to X' on r that is also on $p_3(C, Y)$, and define the path $t = (-r)(X', B) \oplus p_3(B, Y)$. Note that since p_2 is proper, $(-r)(X', B)$ is at least of length one, and therefore, t is a back-door path. Further, since p_3 is d-connecting given $\mathbf{S} \cup \mathbf{Z} \cup \mathbf{X} \setminus \{X\}$ and by the definition of C and r, we have that t is proper back-door path from $\mathbf{X}$ to $\mathbf{Y}$ that is d-connecting given $\mathbf{S} \cup \mathbf{Z}$. But this contradicts that $\mathbf{S}$ satisfies the conditional back-door criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{D}$. ■

F PROOFS FOR SECTION 3.3: MPDAGS - CONSTRUCTING CONDITIONAL ADJUSTMENT SETS

This section includes the proofs of two results from Section 3.3: Lemma 4 and Theorem 5. We also provide three supporting results needed for these proofs.

F.1 Main Results

Proof of Lemma 4. By Lemma 26, $\mathrm{Pa}(X, \mathcal{G})$ must satisfy condition *(a)* of Definition 2, so it suffices to show that $\mathrm{Pa}(X, \mathcal{G}) \cup \mathbf{Z}$ blocks all non-causal definite status paths from X to $\mathbf{Y}$ in $\mathcal{G}$. Note that since $\mathbf{Y} \cap \mathrm{Pa}(X, \mathcal{G}) = \emptyset$, any definite status path from X to $\mathbf{Y}$ in $\mathcal{G}$ that starts with an edge into X is blocked by $\mathrm{Pa}(X, \mathcal{G}) \cup \mathbf{Z}$.

Further, any non-causal definite status path from X to $\mathbf{Y}$ in $\mathcal{G}$ that starts with an edge out of X or an undirected edge must contain a collider. Additionally, the closest collider to X on any such path and all of its descendants in $\mathcal{G}$ must be in $\mathrm{PossDe}(X, \mathcal{G})$ by Lemma 48. Then since $\big[\mathrm{Pa}(X, \mathcal{G}) \cup \mathbf{Z}\big] \cap \mathrm{PossDe}(X, \mathcal{G}) = \emptyset$, these paths are also blocked by $\mathrm{Pa}(X, \mathcal{G}) \cup \mathbf{Z}$. ■

Proof of Theorem 5. By Theorem 3, it suffices to show that $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})$ and $\text{O}(\mathbf{X},\mathbf{Y},\mathcal{G})$ separately satisfy the conditional adjustment criterion relative to $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ in $\mathcal{G}$ (Definition 2). We start by noting that $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})$ and $\text{O}(\mathbf{X},\mathbf{Y},\mathcal{G})$ are both disjoint from $\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})\cup\mathbf{X}\cup\mathbf{Y}\cup\mathbf{Z}$, so it suffices to prove that (a) $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z}$ and (b) $\text{O}(\mathbf{X},\mathbf{Y},\mathcal{G})\cup\mathbf{Z}$ block all proper non-causal definite status paths from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$. We prove (a) and (b) below. For these proofs, note that $\mathbf{Z}\cap\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})=\emptyset$ by the assumption that $\mathbf{Z}\cap\text{PossDe}(\mathbf{X},\mathcal{G})=\emptyset$ and by Lemma 26.

(a) $\mathbf{Adjust(X,Y,Z,\mathcal{G})\cup Z}$**:** Suppose for sake of contradiction that there is a proper non-causal definite status path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that is d-connecting given $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z}$. Let $p=\langle X,\dots,Y\rangle$ be a shortest such path.

Since p is proper, no non-endpoint on p is in $\mathbf{X}$. Suppose for sake of contradiction that there exists $Y'\in\mathbf{Y}$ that is a non-endpoint on p. By choice of p, this implies that $p(X,Y')$ is possibly causal. Then by Lemma 27, since p is non-causal, $p(Y',Y)$ must contain a collider on p. Let C be the closest such collider to Y' (possibly $C=Y'$). Note that by Lemma 27, $C\in\text{PossDe}(Y',\mathcal{G})$, so by Lemma 48, $\text{De}(C,\mathcal{G})\subseteq\text{PossDe}(Y',\mathcal{G})$, where $Y'\in\text{PossDe}(X,\mathcal{G})$. Thus, $\text{De}(C,\mathcal{G})\subseteq\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$. However, this contradicts that p is d-connecting given $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z}$. Therefore, no non-endpoint on p is in $\mathbf{X}\cup\mathbf{Y}$.

We now consider cases (1) and (2) below.

(1) Consider when there is no collider on p. Since p is d-connecting given $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z}$, no node on p is in $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z}$. Then by Equation (6), no node on p is in $\text{PossAn}(\mathbf{X}\cup\mathbf{Y},\mathcal{G})\setminus[\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})\cup\mathbf{X}\cup\mathbf{Y}\cup\mathbf{Z}]$. However, note that by Lemma 27, every non-endpoint on p is a possible ancestor of an endpoint on p and thus is in $\text{PossAn}(\mathbf{X}\cup\mathbf{Y},\mathcal{G})\setminus(\mathbf{X}\cup\mathbf{Y}\cup\mathbf{Z})$. Combining these, we have that all non-endpoints on p are in $\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$. But this implies that there is no set that is both disjoint from $\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$ and can block p. By Theorem 3, this contradicts our assumption that there is a conditional adjustment set relative to $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ in $\mathcal{G}$.

(2) Consider when there is at least one collider C on p. For sake of contradiction, suppose that there are more than three nodes on p. Then there is a non-collider $B\notin\mathbf{X}\cup\mathbf{Y}$ such that $C\leftarrow B$ or $B\rightarrow C$ is on p. Since p is d-connecting given $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z}$, then $B\notin\mathbf{Z}$ and $B\in\text{An}(\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z},\mathcal{G})$. By Equation (6) and Lemma 48, $B\in[\text{PossAn}(\mathbf{X}\cup\mathbf{Y},\mathcal{G})\cup\text{An}(\mathbf{Z},\mathcal{G})]\setminus(\mathbf{X}\cup\mathbf{Y}\cup\mathbf{Z})$. Additionally, since p is d-connecting given $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z}$, then $B\notin\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\equiv[\text{PossAn}(\mathbf{X}\cup\mathbf{Y},\mathcal{G})\cup\text{An}(\mathbf{Z},\mathcal{G})]\setminus(\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})\cup\mathbf{X}\cup\mathbf{Y}\cup\mathbf{Z})$. Combining these, we have that $B\in\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$. Since there is a causal path in $\mathcal{G}$ from B to every node in $\text{De}(C,\mathcal{G})$, by Lemma 48, $\text{De}(C,\mathcal{G})\subseteq\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$. However, this would contradict that p is d-connecting given $\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z}$.

   Hence, p must be of the form $X\rightarrow C\leftarrow Y$, where $C\in\text{An}(\text{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})\cup\mathbf{Z},\mathcal{G})$ and thus by Equation (6) and Lemma 48, $C\in\text{PossAn}(\mathbf{X}\cup\mathbf{Y},\mathcal{G})\cup\text{An}(\mathbf{Z},\mathcal{G})$. Note that $C\notin\text{An}(\mathbf{Z},\mathcal{G})$, since otherwise, $\mathbf{Z}\cap\text{PossDe}(\mathbf{X},\mathcal{G})\neq\emptyset$. Further, $C\notin\text{PossAn}(\mathbf{Y},\mathcal{G})$, because otherwise by Lemma 48, $C\in\text{PossMed}(\mathbf{X},\mathbf{Y},\mathcal{G})$, which would imply $\text{De}(C,\mathcal{G})\subseteq\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$ which we have shown is a contradiction. Therefore, $C\in\text{PossAn}(\mathbf{X},\mathcal{G})$.

   Let $q=\langle C=Q_1,\dots,Q_m=X'\rangle, m\geq 2$, be a shortest possibly causal path in $\mathcal{G}$ from C to $\mathbf{X}$. Further, define the node $Q_j, j\in\{1,\dots,m\}$, as follows. When q has no directed edges, let $Q_j=Q_m$. When q has at least one directed edge, let Q_j be the node on q closest to Q_1 such that $Q_j\rightarrow Q_{j+1}$ is on q. Note that by Lemma 46, q is unshielded. Thus by R1 of Meek (1995), q takes the form $Q_1-\dots-Q_j\rightarrow\dots\rightarrow Q_m$.

   Pause to consider the path $X\rightarrow Q_1\leftarrow Y$. Note that $X\leftarrow Y$ cannot be in $\mathcal{G}$, because no set can block this proper non-causal definite status path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$. By Theorem 3, this would contradict our assumption that there is a conditional adjustment set relative to $(\mathbf{X},\mathbf{Y},\mathbf{Z})$ in $\mathcal{G}$. Similarly, $X\rightarrow Y$ and $X-Y$ are not in $\mathcal{G}$, because this would imply $\text{De}(C,\mathcal{G})\subseteq\text{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$, which we have shown is a contradiction. Thus, $X\rightarrow Q_1\leftarrow Y$ is an unshielded collider in $\mathcal{G}$.

   We complete this case by showing that $\mathcal{G}$ contains $X\rightarrow Q_j\leftarrow Y$. If $j=1$, we are done. If instead $j>1$, then consider the node Q_2. Since $X\rightarrow Q_1-Q_2$ and $Y\rightarrow Q_1-Q_2$ are in $\mathcal{G}$, so is a path $\langle X,Q_2,Y\rangle$ by R1 of Meek (1995). The unshielded paths $X\rightarrow Q_2-Y$ and $X-Q_2\leftarrow Y$ contradict that R1 of Meek (1995) is completed in $\mathcal{G}$. Further, the path $Q_2\rightarrow Y\rightarrow Q_1-Q_2$ or $Q_2\rightarrow X\rightarrow Q_1-Q_2$ contradicts that R2 of Meek (1995) is completed in $\mathcal{G}$, and the path $X-Q_2-Y$ contradicts that R3 of Meek (1995) is completed in $\mathcal{G}$. This leaves only one option for $\langle X,Q_2,Y\rangle$, and that is $X\rightarrow Q_2\leftarrow Y$.

If $j = 2$, we are done. If instead $j > 2$, then we consider the node Q_3. By identical logic to that above, we can show that $\mathcal{G}$ contains $X \to Q_3 \leftarrow Y$. Continuing in this way, we have that $\mathcal{G}$ contains $X \to Q_j \leftarrow Y$.

With this shown, we derive our final contradictions. When $j = m$, then $\mathcal{G}$ contains $X' \leftarrow Y$. But this is a proper non-causal definite status path from $\mathbf{X}$ to $\mathbf{Y}$ that no set can block, which we have shown is a contradiction. When $j < m$, then $\mathcal{G}$ contains the following two paths: $X' \leftarrow \dots \leftarrow Q_j \leftarrow Y$ and $X \to Q_j \leftarrow Y$. These paths are proper non-causal definite status paths from $\mathbf{X}$ to $\mathbf{Y}$ that cannot both be blocked by the same set, which again is a contradiction.

(b) $\mathbf{O(X, Y, \mathcal{G}) \cup Z}$**:** Let p' be an arbitrary proper non-causal definite status path from $X \in \mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$, and let Y be the node in $\mathbf{Y}$ closest to X on p' such that $p'(X, Y)$ is still a proper non-causal definite status path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$. Then let $p = p'(X, Y)$, where $p = \langle X = V_1, \dots, V_k = Y \rangle, k \geq 2$. Additionally, note that by assumption, $Y \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$.

We now consider cases (1) and (2) below. In both cases, we show that p – and therefore p' – is blocked by $\text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$.

(1) Suppose that p ends with $V_{k-1} \leftarrow Y$ or $V_{k-1} - Y$. If p has no colliders, then by Lemma 27, $(-p)$ is a possibly causal path from Y to X. Since $Y \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, this implies that $V_2, \dots, V_{k-1} \in \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$. But then there is no set that is both disjoint from $\text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ and can block p. By Theorem 3, this contradicts our assumption that there is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$. Hence, there must be a collider on p.

    Let C be the closest collider to Y on p. By Lemma 27, $C \in \text{PossDe}(Y, \mathcal{G})$. Thus by Lemma 48, $\text{De}(C, \mathcal{G}) \subseteq \text{PossDe}(Y, \mathcal{G})$. By assumption, $Y \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, which implies that $\text{De}(C, \mathcal{G}) \subseteq \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$. Since $\big[\text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}\big] \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$, p is blocked by $\text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$.

(2) Suppose that p ends with $V_{k-1} \to Y$. Note that p is not a possibly causal path from X to Y, so by Lemma 27, there must be an edge $V_{i-1} \leftarrow V_i$, $i \in \{2, \dots, k-1\}$, on p. In particular, let V_i be the closest node to Y on p such that $V_{i-1} \leftarrow V_i$ is on p.

    In order to complete this proof, we want to show that either $\{V_i, \dots, V_{k-1}\} \cap [\text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}] \neq \emptyset$ or $\{V_i, \dots, V_{k-1}\} \subset \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$. In both cases, we will show that p is blocked by $\text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$. To do this, we briefly note that by the choice of V_i, the path $p(V_i, Y)$ is possibly causal and every node in $\{V_i, \dots, V_{k-1}\}$ is a non-collider on p. Further by the choice of p, no node in $\{V_i, \dots, V_{k-1}\}$ is in $\mathbf{X} \cup \mathbf{Y}$. We turn to consider each node in $\{V_i, \dots, V_{k-1}\}$, working backward through the set.

    Consider the node V_{k-1}. If $V_{k-1} \in \text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$, then since V_{k-1} is a non-collider on p, p is blocked by $\text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$, and we are done. Consider when $V_{k-1} \notin \text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$. Since $Y \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ and since $V_{k-1} \to Y$ is in $\mathcal{G}$, then either $V_{k-1} \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ or $V_{k-1} \in \text{Pa}(\text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G}), \mathcal{G})$. We show the latter is impossible. If $V_{k-1} \in \text{Pa}(\text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G}), \mathcal{G})$ and $V_{k-1} \notin \text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$, then by Equation (7), we have that $V_{k-1} \in \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$. But by Lemma 26, this implies that $V_{k-1} \in \text{De}(\mathbf{X}, \mathcal{G})$. Since $\mathcal{G}$ contains $V_{k-1} \to Y$, then $V_{k-1} \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$. But this contradicts that $V_{k-1} \in \text{Pa}(\text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G}), \mathcal{G})$ by the definition of a parent set. Therefore, either $V_{k-1} \in \text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$ and we are done, or $V_{k-1} \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$.

    In the latter case, we turn to consider V_{k-2} if such a node exists. If p contains $V_{k-2} \to V_{k-1}$, then since $V_{k-1} \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, we can use the same logic as above to show that either $V_{k-2} \in \text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$ and we are done, or $V_{k-2} \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$. If p contains $V_{k-2} - V_{k-1}$, then since $V_{k-1} \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, we have that $V_{k-2} \in \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \subseteq \text{De}(\mathbf{X}, \mathcal{G})$. Because $p(V_{k-2}, Y)$ is possibly causal, then by Lemma 48, $V_{k-2} \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$.

    Working backward in this way, either a node on $p(V_i, Y)$ is in $\text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$ and we are done, or $V_j \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ for all $j \in \{i, \dots, k-1\}$. In the latter case, we have that $V_i \in \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ and that every node in $\{V_i, \dots, V_{k-1}\} \subseteq \text{PossMed}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \subseteq \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ is a non-collider on p. We can now apply the same argument as in (1) above to show that $p(X, V_i)$ – and therefore p – is blocked given $\text{O}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \cup \mathbf{Z}$.

■

F.2 Supporting Results

Lemma 46 *Let X and Y be distinct nodes in an MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and let p be a possibly causal path from X to Y in $\mathcal{G}$. Then any shortest subsequence of p forms an unshielded, possibly causal path from X to Y.*

Proof of Lemma 46. This result is similar to Lemma 3.6 of Perković et al. (2017), but we derive a slightly more general statement.

Let k be the number of nodes on p. Pick an arbitrary shortest subsequence of p and call it p^*, where $p^* = \langle X = V_0, \ldots, V_\ell = Y \rangle$, $0 < \ell \leq k$. Note that there is no edge $V_i \leftarrow V_j, 0 \leq i < j \leq k$ in $\mathcal{G}$, since this would contradict that p is possibly causal. Thus, p^* is also possibly causal by definition. Further note that p^* is unshielded, since if any triple on the path is shielded, it either contradicts that p^* is possibly causal (i.e. $V_i \leftarrow V_{i+2}$ cannot be in p^*) or that p^* is a shortest subsequence of p (i.e. $V_i \rightarrow V_{i+2}$ and $V_i - V_{i+2}$ cannot be in p^*). ■

Lemma 47 *Let $p = \langle P_0, \ldots, P_k \rangle$ be a path in an MPDAG $\mathcal{G}$. Then p is possibly causal if and only if $\mathcal{G}$ does not contain any path $P_i \leftarrow \cdots \leftarrow P_j$, $0 \leq i < j \leq k$.*

Proof of Lemma 47. Suppose that $\mathcal{G}$ does not contain any path $P_i \leftarrow \cdots \leftarrow P_j$, $0 \leq i < j \leq k$. Then $\mathcal{G}$ does not contain any edge $P_i \leftarrow P_j$, $0 \leq i < j \leq k$. Therefore, by definition, p is possibly causal in $\mathcal{G}$.

Now suppose p is possibly causal in $\mathcal{G}$. For sake of contradiction, suppose $\mathcal{G}$ contains a path q from P_i to P_j, $0 \leq i < j \leq k$, of the form $P_i = Q_0 \leftarrow Q_1 \leftarrow \cdots \leftarrow Q_{\ell-1} \leftarrow Q_\ell = P_j$.

Consider the subpath of p from P_i to P_j. Note that this subpath is a possibly causal path. Let $r = \langle P_i = R_0, R_1, \ldots, R_m = P_j \rangle$ be a shortest subsequence of this subpath. By Lemma 46, r is an unshielded, possibly causal path.

Consider the edge $r(R_0, R_1)$. $R_0 \leftarrow R_1$ cannot be in r, since r is possibly causal. Neither is $R_0 \rightarrow R_1$ in r since r being unshielded would imply, by R1 of Meek (1995), that $\mathcal{G}$ contains the cycle $P_i = R_0 \rightarrow R_1 \rightarrow \cdots \rightarrow R_m = P_j = Q_\ell \rightarrow Q_{\ell-1} \rightarrow \cdots \rightarrow Q_0 = P_i$. Thus r contains $R_0 - R_1$.

However, note that no DAG in $[\mathcal{G}]$ can contain the edge $R_0 \rightarrow R_1$, since r being unshielded would imply, by R1 of Meek (1995), that the DAG contains the cycle $P_i = R_0 \rightarrow R_1 \rightarrow \cdots \rightarrow R_m = P_j = Q_\ell \rightarrow Q_{\ell-1} \rightarrow \cdots \rightarrow Q_0 = P_i$. This contradicts that r contains $R_0 - R_1$. Thus we conclude that $\mathcal{G}$ does not contain any path $P_i \leftarrow \cdots \leftarrow P_j$, $0 \leq i < j \leq k$. ■

Lemma 48 *Let X, Y, and Z be distinct nodes in an MPDAG $\mathcal{G}$.*

(i) If p is a possibly causal path from X to Y and q is a causal path from Y to Z, then $p \oplus q$ is a possibly causal path from X to Z.

(ii) If p is a causal path from X to Y and q is a possibly causal path from Y to Z, then $p \oplus q$ is a possibly causal path from X to Z.

Proof of Lemma 48. Let $p = \langle X = P_0, P_1, \ldots, P_k = Y \rangle$ and let $q = \langle Y = Q_0, Q_1, \ldots, Q_r = Z \rangle$. Before beginning the main arguments, we note that p and q cannot share any nodes other than Y, and thus, we can define a path $p \oplus q$. To see this, for sake of contradiction, suppose p and q share at least one node other than Y. Let $\mathbf{S}$ denote the collection of such nodes, and consider the node in $\mathbf{S}$ with the lowest index on q. That is, consider $Q_j \in \mathbf{S}$ such that $j \leq \ell$ for all $Q_\ell \in \mathbf{S}$. Let $Q_j = P_i$ for some $P_i \neq Y$ on p. Note that since q or p is causal, $\mathcal{G}$ contains either $P_k = Q_0 \rightarrow Q_1 \rightarrow \cdots \rightarrow Q_j = P_i$ or $Q_j = P_i \rightarrow P_{i+1} \rightarrow \cdots \rightarrow Y = Q_0$. By Lemma 47, the first option contradicts that p is possibly causal and the second contradicts that q is possibly causal. Thus we conclude that p and q cannot share any nodes other than Y.

For $p \oplus q$ to be possibly causal in $\mathcal{G}$ we only need to show that there is no backward edge between any two nodes on $p \oplus q$. Note that there is no edge $P_{i_1} \leftarrow P_{j_1}$ for $0 \leq i_1 < j_1 \leq k$, or $Q_{i_2} \leftarrow Q_{j_2}$ for $0 \leq i_2 < j_2 \leq r$ in $\mathcal{G}$, by choice of p and q.

(i) Assume for sake of contradiction that there exists an edge $P_i \leftarrow Q_j$ in $\mathcal{G}$ for $i \in \{0, \ldots, k-1\}$ and $j \in \{1, \ldots, r\}$. Note that P_i is on p and not q, and analogously, Q_j is on q and not p, since we have shown p and q cannot share nodes other than Y. Also note that since q is causal, it contains $Y \rightarrow Q_1 \rightarrow \cdots \rightarrow Q_j$.

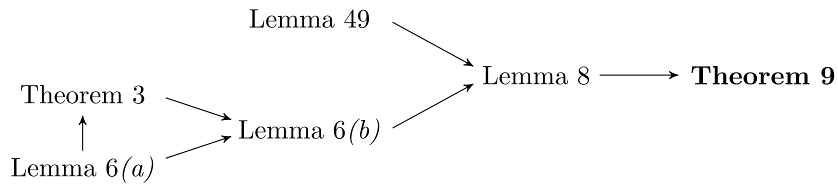

Figure 6: Proof structure of Theorem 9.

Consider the subpath $p(P_i, Y)$. Since p is possibly causal, so is this subpath. Pick an arbitrary shortest subsequence of $p(P_i, Y)$ and call it t, where $t = \langle P_i = T_0, \dots, T_m = Y \rangle$, $m \geq 1$. By Lemma 46, t forms an unshielded, possibly causal path from P_i to Y.

Consider the edge $t(P_i, T_1)$. Edge $P_i \leftarrow T_1$ cannot be on t, since t is possibly causal. Then $P_i \to T_1$ or $P_i - T_1$ must be in $\mathcal{G}$. However, note that no DAG in $[\mathcal{G}]$ can contain the edge $P_i \to T_1$, since t being unshielded would imply, by R1 of Meek (1995), that the DAG contains the cycle $P_i \to T_1 \to \dots \to Y \to \dots \to Q_j \to P_i$. This contradicts that t contains $P_i - T_1$ or $P_i \to T_1$. Thus, there does not exist an edge $P_i \leftarrow Q_j$ in $\mathcal{G}$.

(ii) Assume for sake of contradiction that there exists an edge $P_i \leftarrow Q_j$ in $\mathcal{G}$ for $i \in \{0, \dots, k-1\}$ and $j \in \{1, \dots, r\}$. Note that P_i is on p and not q, and analogously, Q_j is on q and not p, since we have shown p and q cannot share nodes other than Y. Also note that since p is causal, it contains $P_i \to P_{i+1} \to \dots \to Y$.

Consider the subpath $q(Y, Q_j)$. Since q is possibly causal, so is this subpath. Pick an arbitrary shortest subsequence of $q(Y, Q_j)$ and call it t, where $t = \langle Y = T_0, \dots, T_m = Q_j \rangle$, $m \geq 1$. By Lemma 46, t forms an unshielded, possibly causal path from Y to Q_j.

Consider the edge $t(Y, T_1)$. Edge $Y \leftarrow T_1$ cannot be on t, since t is possibly causal. Then $Y \to T_1$ or $Y - T_1$ must be in $\mathcal{G}$. However, note that no DAG in $[\mathcal{G}]$ can contain the edge $Y \to T_1$, since t being unshielded would imply, by R1 of Meek (1995), that the DAG contains the cycle $Y \to T_1 \to \dots \to Q_j \to P_i \to P_{i+1} \to \dots \to Y$. This contradicts that t contains $Y - T_1$ or $Y \to T_1$. Thus, there does not exist an edge $P_i \leftarrow Q_j$ in $\mathcal{G}$. ■

G PROOF FOR SECTION 4.1: PAGS - CONDITIONAL ADJUSTMENT CRITERION

This section includes the proof of Theorem 9 and one result (Lemma 49) needed for the proof of Lemma 8. The statements of Theorem 9 and Lemma 8 can be found in Section 4.1.

Figure 6 shows how the results in this paper fit together to prove Theorem 9. Note that Theorem 9 is an analogous result to Theorem 3 (Section 3.1), where the former applies to PAGs and the latter to MPDAGs. However, while the proof of Theorem 3 relies directly on completeness and soundness proofs for DAGs (see Figure 5 in Supplement D), the proof of Theorem 9 relies on them indirectly through Theorem 3.

Proof of Theorem 9. Follows from Lemma 8 and Theorem 31. ■

Lemma 49 *Let* $\mathbf{X}$ *and* $\mathbf{Z}$ *be disjoint node sets in a PAG* $\mathcal{G}$*. Then the following statements are equivalent.*

(i) $\mathbf{Z} \cap \text{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$.

(ii) $\mathbf{Z} \cap \text{De}(\mathbf{X}, \mathcal{D}) = \emptyset$ *in every DAG* $\mathcal{D}$ *represented by* $\mathcal{G}$*.*

Proof of Lemma 49. $\neg(i) \Rightarrow \neg(ii)$ Let p be a possibly causal path from $\mathbf{X}$ to $\mathbf{Z}$ in $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ and let $p^* = \langle X = V_0, \dots, V_k = Z \rangle$, $k \geq 1$, $X \in \mathbf{X}$, $Z \in \mathbf{Z}$, be an unshielded possibly causal subsequence of p in $\mathcal{G}$.

Since p^* contains $X \circ\!\!-\!\!\circ V_1$, $X \circ\!\!\rightarrow V_1$ or $X \to V_1$, there must be some MAG $\mathcal{M}$ in $[\mathcal{G}]$ with the edge $X \to V_1$. Let p^{**} be the path in $\mathcal{M}$ corresponding to p^* in $\mathcal{G}$. Then since p^* is unshielded, so is p^{**}, and so p^{**} takes the form $X \to V_1 \to \dots \to V_k$. Let $\mathcal{D}$ be a DAG created from $\mathcal{M}$, by retaining all the nodes in $\mathcal{M}$ and all the directed edges in $\mathcal{M}$ and by adding a node L_{AB} and edges $L_{AB} \to B$ and $L_{AB} \to A$ for each bidirected edge $A \leftrightarrow B$ in $\mathcal{M}$ (this DAG is titled the canonical DAG by Richardson and Spirtes, 2002). Now, DAG $\mathcal{D}$ contains a causal path from X to Z.

$\neg(ii) \Rightarrow \neg(i)$ If there is a DAG $\mathcal{D}$ represented by $\mathcal{G}$ with a causal path from $X \in \mathbf{X}$ to $Z \in \mathbf{Z}$, then any MAG $\mathcal{M}$ of $\mathcal{D}$ that contains X and Z will contain a causal path from X to Z. This is due to the fact that a MAG of a DAG will preserve ancestral relationships between observed variables. Then the path in $\mathcal{G}$ that corresponds to q in $\mathcal{M}$ cannot have any arrowheads pointing in the direction of X, and so it must be possibly causal. ■

H PROOFS FOR SECTION 4.2: PAGS - CONSTRUCTING CONDITIONAL ADJUSTMENT SETS

This section includes the proof of Theorem 10, which can be found in Section 4.2. We provide one supporting result needed for the proof of this theorem.

We make an important remark here on R software. Note that by Lemmas 6 and 8, any algorithms developed for checking the existence of an unconditional adjustment set (Definition 11) also apply to conditional adjustment sets – provided that $\mathbf{Z} \cap \mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$. First consider the R package `dagitty` (Textor et al., 2016). Suppose the condition on $\mathbf{Z}$ is satisfied and let $\mathbf{S}$ be a set such that $\mathbf{S} \cap (\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}) = \emptyset$. Then, one can apply the function `isAdjustmentSet` of the package `dagitty` to a PAG $\mathcal{G}$, set $\mathbf{S} \cup \mathbf{Z}$, exposure $\mathbf{X}$, and outcome $\mathbf{Y}$ to learn whether $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$. Next consider the R package `pcalg` (Kalisch et al., 2012). Suppose the condition on $\mathbf{Z}$ is satisfied and let $\mathbf{S}$ be a set such that $\mathbf{S} \cap (\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}) = \emptyset$. Then, one could apply the function `gac` of the package `pcalg` to the MPDAG or PAG $\mathcal{G}$ and to the node sets $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{S} \cup \mathbf{Z}$. These functions will return `TRUE` if and only if $\mathbf{S}$ is a conditional adjustment set relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$, and `FALSE` otherwise.

H.1 Main Result

Proof of Theorem 10. Suppose that $\mathrm{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G})$ does not satisfy the conditional adjustment criterion relative to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$. Since $\mathrm{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cap \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$ by construction, it must be that there is a proper definite status non-causal path from $\mathbf{X}$ to $\mathbf{Y}$ that is m-connecting given $\mathrm{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}$. By Lemma 50, there is then a proper definite status non-causal path p from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ such that all definite non-colliders on p are in $\mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ (case (ii) of Lemma 50) and all colliders on p are in $\mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G})$ (cases (iii) and (vi) of Lemma 50). Since $\mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \subseteq \mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z} \cup \mathbf{S}, \mathcal{G})$, for any set $\mathbf{S}$ that satisfies $[\mathbf{S} \cup \mathbf{Z}] \cap [\mathbf{X} \cup \mathbf{Y} \cup \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})] = \emptyset$, Lemma 30 implies that there is also a proper definite status non-causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that is open given $\mathbf{S}$. Since this is true for an arbitrary set $\mathbf{S}$ that satisfies condition *(a)* of Definition 7, it follows that there cannot be any set that satisfies the conditional adjustment criterion relative to to $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in $\mathcal{G}$. ■

H.2 Supporting Result

Lemma 50 *Let* $\mathbf{X}$, $\mathbf{Y}$, *and* $\mathbf{Z}$, *be pairwise disjoint node sets in a PAG* $\mathcal{G}$, *where* $\mathbf{Z} \cap \mathrm{PossDe}(\mathbf{X}, \mathcal{G}) = \emptyset$ *and where every proper possibly causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *starts with a visible edge out of* $\mathbf{X}$. *Suppose furthermore, that there exists a set* $\mathbf{S}$ *that satisfies the conditional adjustment criterion for* $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ *in* $\mathcal{G}$. *If there is a proper definite status non-causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *that is m-connecting given* $\mathrm{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}$ *(see definition in Theorem 10), then there is a path* p *from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$ *such that the following hold.*

(i) Path p *is a proper definite status non-causal path from* $\mathbf{X}$ *to* $\mathbf{Y}$ *in* $\mathcal{G}$.

(ii) All definite non-colliders on p *are in* $\mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$.

(iii) There is at least one collider on p, *and all colliders on* p *are in* $\mathbf{C_1} \cup \mathbf{C_2}$, *where* $\mathbf{C_1}$ *and* $\mathbf{C_2}$ *are disjoint sets such that*

$$\mathbf{C_1} \subseteq \mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \setminus \left[\mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \cup \mathbf{X} \cup \mathbf{Y} \cup \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})\right] \text{ and}$$
$$\mathbf{C_2} \subseteq \mathrm{An}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \setminus \left[\mathbf{X} \cup \mathbf{Y} \cup \mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})\right].$$

(iv) None of the colliders on p *can be possible descendants of a non-collider on* p.

(v) For any collider $C \in \mathbf{C_1}$ *on* p *there is an unshielded possibly directed path from* C *to* $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ *that does not start with* $\circ\!\!-\!\!\circ$.

(vi) $\mathbf{C_1} = \emptyset$*, that is for any collider* $C \in \mathbf{C_1}$ *on* p *there is an unshielded directed path from* C *to* $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$.

Proof of Lemma 50. Consider the sets of all proper definite status non-causal paths from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that are m-connecting given $\mathrm{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cup \mathbf{Z}$ and choose among them a shortest path with a shortest distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16). Let this path be called p, where $p = \langle X = V_1, V_2, \dots, V_k = Y \rangle$, $X \in \mathbf{X}, Y \in \mathbf{Y}$, $k \geq 2$. By choice of p, (i) is satisfied. We will now show that p also satisfies properties (ii)-(vi) above.

First, consider properties (ii) and (iii). Since p is m-connecting given $\mathrm{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cup \mathbf{Z}$, any collider on p is in $\mathrm{An}(\mathrm{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cup \mathbf{Z}, \mathcal{G})$. Furthermore, since $\mathrm{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cup \mathbf{Z} = \mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \setminus \big[\mathbf{X} \cup \mathbf{Y} \cup \mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})\big]$, and since in a PAG $\mathcal{G}$ for any set $\mathbf{W}$, $\mathrm{An}(\mathrm{PossAn}(\mathbf{W},\mathcal{G})) = \mathrm{PossAn}(\mathbf{W},\mathcal{G})$, we have that any collider on p is in $\mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G})$. Furthermore, since by definition, $\mathrm{De}(\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G}),\mathcal{G}) = \mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$, we have that no collider on p can be in $\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$. Hence, all colliders on p are in $\mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \setminus \mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$.

Also, since p is proper, a node in $\mathbf{X}$ cannot be a non-endpoint node on p. Now, since p is additionally chosen as a shortest proper non-causal definite status path from $\mathbf{X}$ to $\mathbf{Y}$ that is m-connecting given $\mathrm{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G})$, it holds that either a node in $\mathbf{Y}$ is not a non-endpoint node on p, or there is a node $Y' \in \mathbf{Y} \setminus \{Y\}$ on p such that $p(X, Y')$ is a possibly causal path from X to Y'. Moreover, in this case $p(X, Y')$ must be a causal path in $\mathcal{G}$ (because p must start with a visible edge and because $A \bullet\!\!\rightarrow B \circ\!\!-\!\!\bullet C$ cannot be a subpath of a definite status path). Since p itself is a non-causal path in $\mathcal{G}$, there is a collider on p that is a descendant of Y'. But since $Y' \in \mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$, this collider would then also have to be in $\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$, which we have ruled out as an option in the previous paragraph. Hence, a node on $\mathbf{Y}$ is also not a non-endpoint node on p.

Then all colliders on p are in $\mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \setminus [\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G}) \cup \mathbf{X} \cup \mathbf{Y}]$. Also, any definite non-collider on p is a possible ancestor of a collider on p or of an endpoint on p. Hence, every definite non-collider on p is in $\mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \setminus [\mathbf{X} \cup \mathbf{Y}]$. But, since p is m-connecting given $\mathrm{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cup \mathbf{Z}$, none of the definite non-colliders on p are in $\mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \setminus \big[\mathbf{X} \cup \mathbf{Y} \cup \mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})\big]$. Therefore, any definite non-collider on p is in $\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$. This proves property (ii).

Next, consider property (iii). We have already shown that any collider on p is in $\mathrm{PossAn}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}, \mathcal{G}) \setminus [\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G}) \cup \mathbf{X} \cup \mathbf{Y}]$. So it is only left to show that at least one collider is on p. Since we know that p must be blocked by $\mathbf{S} \cup \mathbf{Z}$ for some set $\mathbf{S}$, where $\mathbf{S} \cap \big[\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z} \cup \mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})\big] = \emptyset$, and since all definite non-colliders on p are in $\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$, there is at least one collider C on p.

Property (iv) follows almost directly now, since by (ii), all definite non-colliders on p are in $\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$ and by (iii), none of the colliders can be in $\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$. The claim then holds since by definition of the $\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$ in a PAG, $\mathrm{PossDe}(\mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G}),\mathcal{G}) = \mathrm{Forb}(\mathbf{X},\mathbf{Y},\mathcal{G})$.

Next, we show properties (v) and (vi). Let $C \in \mathbf{C_1}$ be a collider on p. Then $C \notin \big[\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}\big]$ and that there is an unshielded possibly directed path $r = \langle C, Q, \dots, V \rangle$ from C to a node $V \in \mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$.

(v) Suppose for a contradiction that edge $\langle C, Q \rangle$ on r is of type $C \circ\!\!-\!\!\circ Q$ (possibly $Q = V$). We derive a contradiction by constructing a proper definite status non-causal path from $\mathbf{X}$ to $\mathbf{Y}$ that is m-connecting given $\mathrm{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cup \mathbf{Z}$ and shorter than p, or of the same length as p but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16).

Let A and B be nodes on p such that $A \bullet\!\!\rightarrow C \leftarrow\!\!\bullet B$ is a subpath of p (possibly $A = X$, $B = Y$). Then paths $A \bullet\!\!\rightarrow C \circ\!\!-\!\!\circ Q$ and $B \bullet\!\!\rightarrow C \circ\!\!-\!\!\circ Q$ together with Lemma 28 imply that $A \bullet\!\!\rightarrow Q \leftarrow\!\!\bullet B$ is in $\mathcal{G}$.

Suppose first that $A \neq X$, and $B \neq Y$. Note that by property (iv) above, if $A \neq X$, then $A \leftrightarrow C$ is in $\mathcal{G}$. Moreover, if $A \leftrightarrow C$ is in $\mathcal{G}$, then $A \leftrightarrow Q$ is in $\mathcal{G}$, otherwise path $\langle A, Q, C \rangle$ and edge $A \leftrightarrow C$ contradict Lemma 29. Hence, if $A \neq X$, the collider/definite non-collider status of A is the same on p and on $p(X, A) \oplus \langle A, Q \rangle$. Analogous reasoning can be employed in the case when $B \neq Y$, to show that $B \leftrightarrow Q$, that is, the collider/definite non-collider status of B is the same on p and on $\langle Q, B \rangle \oplus p(B, Y)$.

Now, we return to the general case where we allow $A = X$ and $B = Y$. In each of the cases below we will derive the contradiction by finding a path s from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that is a proper non-causal definite status path in $\mathcal{G}$ and m-connecting given $\mathrm{Adjust}(\mathbf{X},\mathbf{Y},\mathbf{Z},\mathcal{G}) \cup \mathbf{Z}$. Additionally, the path s will either be shorter than p or of the same length as p, but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16) which implies a contradiction with our choice of p.

Suppose first that Q is not a node on p.

- If $Q \notin \mathbf{X} \cup \mathbf{Y}$, then
  - if $A \neq X$ and $B \neq Y$, then let $s = p(X, A) \oplus \langle A, Q, B\rangle \oplus p(B, Y)$. By the reasoning above, this path transformation amounts to replacing $A \leftrightarrow C \leftrightarrow B$ on p with $A \leftrightarrow Q \leftrightarrow B$ on s thereby creating a path with the same properties as p but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16).
  - If $A = X$, and $B \neq Y$, then let $s = \langle A, Q, B\rangle \oplus p(B, Y)$. This path transformation amounts to replacing $X \bullet\!\!\rightarrow C \leftrightarrow B$ on p, with $X \bullet\!\!\rightarrow Q \leftrightarrow B$ on s, thereby creating a path with the same properties as p and of the same length as p but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16).
  - If $A \neq X$, and $B = Y$, then let $s = p(X, A) \oplus \langle A, Q, B\rangle$. This path transformation amounts to replacing $A \leftrightarrow C \leftarrow\!\!\bullet Y$ on p, with $A \leftrightarrow Q \leftarrow\!\!\bullet Y$ on s, thereby creating a path with the same properties as p, that is of the same length, but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16).
  - If $A = X$, and $B = Y$, then let $s\langle A, Q, B\rangle$. Now s is of the form $X \bullet\!\!\rightarrow Q \leftarrow\!\!\bullet Y$ and clearly satisfies all the same properties as p while being of the same length, but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16).
- If $Q \equiv X'$, $X' \in \mathbf{X}$, then:
  - if $B \neq Y$, let $s = \langle Q, B\rangle \oplus p(B, Y)$. This path transformation amounts to replacing $X \ldots C \leftrightarrow B$ on p, with $X' \leftrightarrow B$ on s, thereby creating a shorter path with the same properties as p.
  - If $B = Y$, then let $s = \langle Q, B\rangle$. Due to the discussion above, s is of the form $X' \leftarrow\!\!\bullet Y$ in $\mathcal{G}$.
- Otherwise, $Q \equiv Y', Y' \in \mathbf{Y}$. If $Q \in \mathbf{Y} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, this would imply that $C \in \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, which contradicts (iii). So Q must be in $\mathbf{Y} \setminus \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$. Then:
  - if $A \neq X$, then let $s = p(X, A) \oplus \langle A, Q\rangle$. This path transformation amounts to replacing $A \leftrightarrow C \ldots Y$ on p, with $A \leftrightarrow Y'$ on s, thereby creating a shorter path with the same properties as p.
  - If $A = X$, then let $s = \langle A, Q\rangle$. Due to the discussion above, s is of the form $X \leftarrow\!\!\bullet Y'$ in $\mathcal{G}$.

Otherwise, Q is on p. Therefore, $Q \notin \mathbf{X} \cup \mathbf{Y}$. Also, Q is a collider on p, otherwise $Q \in \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$ and $C \in \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, because of $C \circ\!\!-\!\!\circ Q$.

- Suppose first that Q is on $p(C, Y)$. Then:
  - if $A \neq X$, then let $s = p(X, A) \oplus \langle A, Q\rangle \oplus p(Q, Y)$. This path transformation amounts to replacing $A \leftrightarrow C \leftrightarrow \cdots \leftrightarrow Q$ on p, with $A \leftrightarrow Q$ on s, thereby creating a shorter path with the same properties as p.
  - If $A = X$, then let $s = \langle A, Q\rangle \oplus p(Q, Y)$. This path transformation amounts to replacing $X \leftrightarrow C \leftrightarrow \cdots \leftrightarrow Q$ on p, with $X \leftrightarrow Q$ on s, thereby creating a shorter path with the same properties as p.
- Next, suppose that Q is on $p(X, C)$. Then depending on whether $B = Y$, we can choose one of the following paths as the path s:
  - if $B \neq Y$, then let $s = p(X, Q) \oplus \langle Q, B\rangle \oplus p(B, Y)$. This path transformation amounts to replacing $Q \leftrightarrow \cdots \leftrightarrow C \leftrightarrow B$ on p, with $Q \leftrightarrow B$ on s, thereby creating a shorter path with the same properties as p.
  - If $B = Y$, then let $s = p(X, Q) \oplus \langle Q, B\rangle$. Similarly to above, this path transformation amounts to replacing $Q \leftrightarrow \cdots \leftrightarrow C \leftrightarrow Y$ on p, with $Q \leftrightarrow Y$ on s, thereby creating a shorter path with the same properties as p.

(vi) Since we showed above that the starting edge $\langle C, Q\rangle$ on $r = \langle C, Q, \ldots, V\rangle$ is not of the form $C \circ\!\!-\!\!\circ Q$, and since r is an unshielded possibly directed path from C to $V \in \mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$, in order to prove property (vi) it is enough to show that $\langle C, Q\rangle$ is also not of the form $C \circ\!\!\rightarrow Q$ (since $P_1 \bullet\!\!\rightarrow P_2 \circ\!\!-\!\!\bullet P_3$ cannot be a subpath of any unshielded possibly directed path in $\mathcal{G}$, Zhang, 2008b). Suppose for a contradiction that $\langle C, Q\rangle$ is exactly of that form. Since $A \bullet\!\!\rightarrow C \leftarrow\!\!\bullet B$ and $C \circ\!\!\rightarrow Q$ are in $\mathcal{G}$, by Lemma 28, $A \bullet\!\!\rightarrow Q \leftarrow\!\!\bullet B$ is in $\mathcal{G}$.

Now, our goal is to identify a nodes A' and B' on p that satisfy the following. Node A' is on $p(X, A)$, and edge $A' \bullet\!\!\to Q$ is in $\mathcal{G}$. Additionally, $A' = X$ or A' is a non-endpoint node on p that has the same definite non-collider/collider status on p and on $p(X, A') \oplus \langle A', Q \rangle$. Similarly, B' is on $p(B, Y)$, and edge $B' \bullet\!\!\to Q$ is in $\mathcal{G}$. Additionally, $B' = Y$ or B' is a non-endpoint node on p that has the same definite non-collider/collider status on p and on $\langle Q, B' \rangle \oplus p(B', Y)$. We only show how to find node A' on $p(X, A)$, since the argument for finding B' on $p(B, Y)$ is exactly symmetric.

- Consider the path $p(X, C) = \langle X = V_1, V_2, \dots, V_{i-1} = A, V_i = C \rangle$. Note that by (iv) and the properties of unshielded paths, $p(X, C)$ is of the form $X \bullet\!\!\to V_2 \leftrightarrow \dots \leftrightarrow A \leftrightarrow C$ or $X \leftarrow V_2 \leftarrow \dots \leftarrow V_j \leftrightarrow \dots \leftrightarrow C$, for some V_j, $j \in \{2, \dots, i-1\}$.

  Hence, if there is any non-endpoint node W on $p(X, A)$ such that $W \leftrightarrow Q$, this node has the same definite collider / non-collider status on both p and on $p(X, W) \oplus \langle W, Q \rangle$. Then we choose $A' \equiv W$. Otherwise, if there is a non-endpoint node W on $p(X, A)$ such that $-p(W, X)$ is of the form $W \to \dots \to X$, and an edge $W \to Q$ or $W \circ\!\!\to Q$ is in $\mathcal{G}$, then W is a definite non-collider on both p and $p(X, W) \oplus \langle W, Q \rangle$ and we choose $A' \equiv W$.

  We will now show that if neither of the above choices for A' are possible in $\mathcal{G}$, then $p(X, C)$ is of the form $X \leftrightarrow V_2 \leftrightarrow \dots \leftrightarrow C$, and for every node V_j, $j \in \{1, \dots, i\}$ on $p(X, C)$, the edge $V_j \to Q$ or $V_j \circ\!\!\to Q$ is in $\mathcal{G}$. In this case, we choose $A' \equiv X$.

  Hence, consider first node $V_{i-1} = A$ on p. By above $V_{i-1} \bullet\!\!\to Q$ is in $\mathcal{G}$. Also, by our assumption $V_{i-1} \leftrightarrow Q$ is not in $\mathcal{G}$, so we must have either $V_{i-1} \to Q$ or $V_{i-1} \circ\!\!\to Q$ is in $\mathcal{G}$. Similarly, by the assumption above we now know that edge $\langle V_{i-2}, V_{i-1} \rangle$ is not of the form $V_{i-2} \leftarrow V_{i-1}$, so we can conclude that $V_{i-2} \leftrightarrow V_{i-1}$ is in $\mathcal{G}$.

  Now, $V_{i-2} \leftrightarrow V_{i-1} \leftrightarrow C \circ\!\!\to Q$ and either $V_{i-1} \to Q$ or $V_{i-1} \circ\!\!\to Q$ is in $\mathcal{G}$. If $V_{i-1} \to Q$ is in $\mathcal{G}$, then $R4$ of Zhang (2008b) would imply that $V_{i-2} \in \text{Adj}(Q, \mathcal{G})$. Moreover, since $V_{i-2} \leftrightarrow V_{i-1} \to Q$ is in $\mathcal{G}$, $R2$ of Zhang (2008b) would imply that $V_{i-2} \bullet\!\!\to Q$ is in $\mathcal{G}$, and our assumption further lets us conclude that $V_{i-2} \to Q$, or $V_{i-2} \circ\!\!\to Q$ is in $\mathcal{G}$.

  If $V_{i-1} \circ\!\!\to Q$ is in $\mathcal{G}$, then $V_{i-2} \leftrightarrow V_{i-1} \circ\!\!\to Q$ and Lemma 28 imply that, $V_{i-2} \bullet\!\!\to Q$ is in $\mathcal{G}$. Hence, as above either $V_{i-2} \to Q$, or $V_{i-2} \circ\!\!\to Q$ is in $\mathcal{G}$.

  If $V_{i-2} = X$ we are done. Otherwise, we can repeat the same argument as in the preceding three paragraphs to conclude that $V_{i-3} \leftrightarrow V_{i-2} \leftrightarrow V_{i-1} \leftrightarrow C$ is in $\mathcal{G}$, and either $V_{i-3} \to Q$ or $V_{i-3} \circ\!\!\to Q$ are in $\mathcal{G}$. If $X \neq V_{i-3}$, we can keep applying the same argument, until we reach X.

Now that we have chosen the appropriate A' and B' the remaining argument is very similar to case (v). In each of the cases below we will derive the contradiction by finding a path s from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ that is a proper non-causal definite status path in $\mathcal{G}$ and m-connecting given $\text{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}$. Additionally, the path s will either be shorter than p or of the same length as p, but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16) which implies a contradiction with our choice of p.

Suppose first that Q is not on p:

- If $Q \notin \mathbf{X} \cup \mathbf{Y}$ then:
  - if $A' \neq X$ and $B' \neq Y$, then let $s = p(X, A') \oplus \langle A', Q, B' \rangle \oplus p(B', Y)$. By the reasoning above, this path transformation amounts to replacing $p(A', B')$ on p with $\langle A', Q, B' \rangle$ on s such that the collider / definite non-collider status of A' and B' is the same on both paths. Therefore, s is a path with the same properties as p, but either shorter than p or of the same length but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16).
  - If $A' = X$, and $B' \neq Y$, then let $s = \langle A', Q, B' \rangle \oplus p(B', Y)$. By the reasoning above, this path transformation amounts to replacing $p(X, B')$ on p with $\langle X, Q, B' \rangle$ on s such that the collider / definite non-collider status of B' is the same on both paths, and s is a non-causal path because of $Q \leftarrow\!\!\bullet B'$ edge. Therefore, s is a path with the same properties as p but either shorter than p or of the same length but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16).
  - If $A' \neq X$, and $B' = Y$, then let $s = p(X, A') \oplus \langle A', Q, B' \rangle$. This path transformation amounts to replacing $p(A', Y)$ on p with $\langle A', Q, Y \rangle$ on s such that the collider / definite non-collider status of B' is

the same on both paths, and s is a non-causal path because of $Q \leftarrow\!\!\bullet Y$ edge. Therefore, s is a path with the same properties as p but either shorter than p or of the same length but with a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ (Definition 16).

  – If $A' = X$ and $B' = Y$, $\langle A', Q, B' \rangle$. Then s is of the form $X \bullet\!\!\rightarrow Q \leftarrow\!\!\bullet Y$ and $Q \in \text{An}(\text{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}, \mathcal{G})$ and Q has a shorter distance to $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ than C.

- If $Q \equiv X'$, $X' \in \mathbf{X}$, then:
  – if $B' \neq Y$, then let $s = \langle Q, B' \rangle \oplus p(B', Y)$. This path transformation amounts to replacing $p(X, B')$ on p with $\langle X', B' \rangle$ on s such that the collider / definite non-collider status of B' is the same on both paths, and s is a non-causal path because of $X' \leftarrow\!\!\bullet B'$ edge. Therefore, s is a path with the same properties as p shorter than p.
  – If $B' = Y$, then let $s = \langle Q, B' \rangle$, where based on the reasoning above, s is of the form $X' \leftarrow\!\!\bullet Y$.

- Otherwise, $Q \equiv Y'$, $Y' \in \mathbf{Y}$. Then
  – if $A' \neq X$, then $s = p(X, A') \oplus \langle A', Q \rangle$. Note that in this case s is of the form $X \leftrightarrow \cdots \leftrightarrow A' \leftrightarrow Y'$, or $X \leftarrow \cdots \leftarrow A' \circ\!\!\rightarrow Y'$, or $X \leftarrow \cdots \leftarrow A' \rightarrow Y'$. In all cases, s is a proper non-causal definite status path from $\mathbf{X}$ to $\mathbf{Y}$ that is m-connecting given $\text{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}$.
  – If $A' = X$, then let $s = \langle A', Q \rangle$. We now discuss why s is of the form $X \leftrightarrow Y'$ in $\mathcal{G}$.
  Note that $X \circ\!\!\rightarrow Y'$ cannot be in $\mathcal{G}$, since there exists a set $\mathbf{S}$ that can satisfy the conditional adjustment criterion relative to $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ in $\mathcal{G}$. If instead $X \rightarrow Y'$ is a visible edge in $\mathcal{G}'$, then there is either a node $D \notin \text{Adj}(Y', \mathcal{G})$ such that $D \bullet\!\!\rightarrow X$ is in $\mathcal{G}$ or there is a collection of nodes $D_1, \dots, D_k$, such that $D_1 \notin \text{Adj}(Y', \mathcal{G})$, $D_2, \dots, D_k \in \text{Pa}(Y', \mathcal{G})$, and $D_1 \bullet\!\!\rightarrow D_2 \leftrightarrow \cdots \leftrightarrow D_k \leftrightarrow X$ is in $\mathcal{G}$. Without loss of generality we will assume that we are in the fist case, that is $D \bullet\!\!\rightarrow X$ is in $\mathcal{G}$ and $D \notin \text{Adj}(Y', \mathcal{G})$, since the latter case has an analogous proof to what follows.
  By above, the only way way that $A' \equiv X$ is if $X \leftrightarrow V_2 \leftrightarrow \cdots \leftrightarrow C$ is in $\mathcal{G}$ and if for all nodes $V_j \in \{V_2, \dots, V_{i-2}, V_{i-1}, V_i\}, V_j \rightarrow Y'$, or $V_j \circ\!\!\rightarrow Y'$ is in $\mathcal{G}$. Now since, $D \bullet\!\!\rightarrow X \leftrightarrow V_2 \leftrightarrow \cdots \leftrightarrow V_{i-1} \leftrightarrow C$ is also in $\mathcal{G}$, and $D \notin \text{Adj}(Y', \mathcal{G})$, we can use $R4$ of Zhang (2008b) iteratively to conclude that $V_j \rightarrow Y'$ is in $\mathcal{G}$ for all $j \in \{1, \dots, i\}$. However, as $V_i \equiv C$, this contradicts our assumption that $C \circ\!\!\rightarrow Y'$ is in $\mathcal{G}$, for $Y' = Q$.

Otherwise, Q is on p. Therefore, $Q \notin \mathbf{X} \cup \mathbf{Y}$.

- Suppose first that Q is on $p(C, Y)$. By (iii), (iv), and the definition of $\text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, we have that $p(C, Y)$ is of one of the following forms:
  – $C \leftrightarrow \cdots \leftrightarrow Q \leftrightarrow \cdots \leftrightarrow V_k \leftarrow\!\!\bullet Y$ for $k > i$, or
  – $C \leftrightarrow \cdots \leftrightarrow Q \leftrightarrow \cdots \leftrightarrow T_1 \rightarrow \cdots \rightarrow Y$, for some T_1 on $p(C, Y)$, or
  – $C \leftrightarrow \cdots \leftrightarrow T_2 \rightarrow \cdots \rightarrow Q \rightarrow \cdots \rightarrow Y$, for some T_2 on $p(C, Y)$, or
  – $C \leftrightarrow \cdots \leftrightarrow Q \rightarrow \dots \cdots \rightarrow Y$.

  Then
  – If $A' \neq X$, then $s = p(X, A') \oplus \langle A', Q \rangle \oplus p(Q, Y)$. Note that by above forms of $p(C, Y)$ s is always a is a proper non-causal path from $\mathbf{X}$ to $\mathbf{Y}$. Additionally, by above listed options for $p(C, Y)$ we know that Q has the same collider / definite non-collider status on both p and s. Hence, s is also an m-connecting path given $\text{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}$. Since s is also shorter than p we obtain our contradiction.
  – If $A' \equiv X$, we let $s = p(X, Q) \oplus p(Q, Y)$. Path s is proper, since p itself is proper and $Q \notin \mathbf{X} \cup \mathbf{Y}$. Furthermore, by the above listed options for $p(C, Y)$ we know that Q has the same collider / definite non-collider status on both p and s and that s is a definite status path. Hence, s is also an m-connecting path given $\text{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}$. If s is a non-causal path in $\mathcal{G}$, we obtain a contradiction with the choice of p.
  Hence, suppose for a contradiction that s is a possibly causal path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$. By assumption, it must be that $X \rightarrow Q$ is a visible edge in $\mathcal{G}$. Now, similarly to the previous case, since $X \rightarrow Q$ is a visible edge in $\mathcal{G}$, there is either a node $D \notin \text{Adj}(Q, \mathcal{G})$ such that $D \bullet\!\!\rightarrow X$ is in $\mathcal{G}$ or there is a collection of nodes $D_1, \dots, D_k$ such that $D_1 \notin \text{Adj}(Q, \mathcal{G})$, $D_2, \dots, D_k \in \text{Pa}(Q, \mathcal{G})$, and $D_1 \bullet\!\!\rightarrow D_2 \leftrightarrow \cdots \leftrightarrow D_k \leftrightarrow X$

is in $\mathcal{G}$. We again assume without loss of generality that we are in the former case, that is $D \bullet\!\!\rightarrow X$ is in $\mathcal{G}$ and $D \notin \mathrm{Adj}(Q, \mathcal{G})$.
Since $A' \equiv X$, by the same reasoning as in the previous case above we know that $X \leftrightarrow V_2 \leftrightarrow \cdots \leftrightarrow C$ is in $\mathcal{G}$ and that for all nodes $V_j \in \{V_1, \dots, V_{i-1}, V_i\}$, $V_j \rightarrow Q$, or $V_j \circ\!\!\rightarrow Q$ is in $\mathcal{G}$. Now since, $D \bullet\!\!\rightarrow X \leftrightarrow V_2 \leftrightarrow \cdots \leftrightarrow C$ is in $\mathcal{G}$, and since $D \notin \mathrm{Adj}(Q, \mathcal{G})$, we can use $R4$ of Zhang (2008b) iteratively to conclude that $V_j \rightarrow Q$ is in $\mathcal{G}$ for all $j \in \{1, \dots, i\}$. However, as $V_i \equiv C$, this contradicts our assumption that $C \circ\!\!\rightarrow Q$ is in $\mathcal{G}$.

- Lastly, suppose that Q is on $p(X, C)$. Analogously to above, by (iii), (iv), and the definition of $\mathrm{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G})$, we have that $p(X, C)$ is of one of the following forms:
  - $X \bullet\!\!\rightarrow V_2 \leftrightarrow \cdots \leftrightarrow Q \leftrightarrow \cdots \leftrightarrow C$, or
  - $X \leftarrow \cdots \leftarrow T_1 \leftrightarrow \cdots \leftrightarrow Q \leftrightarrow \cdots \leftrightarrow C$, for some T_1 on $p(X, C)$, or
  - $X \leftarrow \cdots \leftarrow Q \leftarrow \cdots \leftarrow T_2 \leftrightarrow \cdots \leftrightarrow C$, for some T_2 on $p(X, C)$, or
  - $X \leftarrow \cdots \leftarrow Q \leftrightarrow \cdots \leftrightarrow C$.

  Then

  - If $B' \neq Y$, we have that $s = p(X, Q) \oplus \langle Q, B' \rangle \oplus p(B, Y)$ is a proper non-causal path from $\mathbf{X}$ to $\mathbf{Y}$ that is shorter than p. Additionally, Q is of the same collider / definite non-collider status on both p and s and therefore, s is not only of definite status, but also m-connecting given $\mathrm{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}$ in $\mathcal{G}$ which leads to a contradiction.
  - If $B' \equiv Y$, then $s = p(X, Q) \oplus \langle Q, B' \rangle$ is a proper definite status non-causal path that is m-connecting given $\mathrm{Adjust}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathcal{G}) \cup \mathbf{Z}$ in $\mathcal{G}$ and shorter than p.

■