On the Structural Basis of Conditional Ignorability

Elias Bareinboim and Drago Plečko

Abstract. Conditional ignorability is a cornerstone of the Neyman–Rubin potential outcomes framework, enabling identification of causal effects via covariate adjustment. However, we argue that such assumptions are more difficult to assess than commonly appreciated, due to the implicit judgments they require about relationships within the covariate set used for adjustment. Formally, we show that verifying ignorability over n covariates may implicitly require evaluating over $O(4^{n^2})$ structural configurations — posing a combinatorial challenge that is intractable for most data scientists to resolve by the naked eye, without structural or graphical support. To address this challenge, we develop a structural account of ignorability grounded in the semantics of structural causal models, and introduce a new class of graphical models — cluster causal diagrams over three distinct blocks (treatment, outcome, adjustment covariates), denoted $\mathrm{CG}(3)$ — that abstract away the internal structure within the set of covariates. We define the notion of *structural ignorability*, which can be evaluated using the back-door criterion on $\mathrm{CG}(3)$ diagrams, offering a transparent and practical method for assessing ignorability-type assumptions. Our proposal bridges potential outcomes and graphical frameworks, drawing on foundational ideas from statistics, genetics, econometrics, and computer science, while retaining the clarity of the structural approach and requiring fewer assumptions than full causal diagrams.

1. INTRODUCTION

Inference of causal effects is a fundamental task across the sciences. Although randomized experiments [11] are considered the gold standard for causal inference in many applied fields, much of the literature is concerned with inferring effects from non-experimental (observational) data. In such settings, however, inferences may remain impossible, even with unlimited data. As Cartwright famously put it, “no causes in, no causes out” [7]; causal conclusions require causal assumptions. This intuition is formalized by the *Causal Hierarchy Theorem* (CHT) [4, Thm. 1], which shows that, in an information-theoretic sense, causal inferences cannot be drawn in the absence of causal knowledge. Acknowledging this impossibility, a central question in causal inference becomes: when and under what conditions can causal effects be inferred from data? [32, 24]. Two major frameworks provide languages for articulating the necessary assumptions to connect data with causal claims — potential outcomes and structural causal models — each offering distinct perspectives on model elicitation, abstraction, and operational practice.

The first, *potential outcomes* (PO) framework, is associated with the work that started with Don Rubin and colleagues [32], which extended pioneering ideas by Jerzy Neyman on modeling counterfactual outcomes in randomized experiments [21]. This framework formalizes the notion of potential outcomes, enabling reasoning about their distributions (see the gray/left side of Fig. 1). Its assumptions are typically algebraic, expressed in terms of independence among factual and counterfactual variables. For inferring causal effects, a common route starts by assessing some form of unconfoundedness – usually encoded through the so-called conditional ignorability assumption [31, 29]. If a corresponding ignorability statement holds, the causal query can be reduced to a statistical functional of the observational data distribution (often in the form of adjustment), and attention shifts to estimating the query from a finite amount of data. While this focus on estimation has made the PO framework influential in applied fields, it offers little guidance for assessing whether key causal assumptions are justified. For the estimation step, a range of remarkably influential techniques have been developed, such as inverse-propensity weighting [29], matching [28], and doubly-robust methods [26, 27, 2], to name a few (see [14] for a review).

Causal Artificial Intelligence Lab, Columbia University.

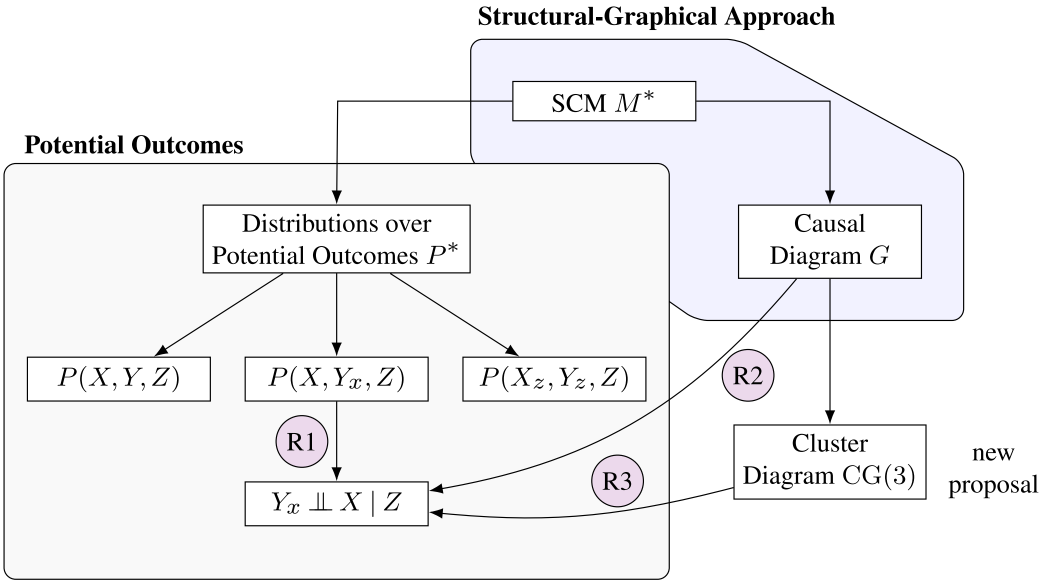


Figure 1: Inferential pipeline's overview, highlighting emphasis of the PO and graphical approaches. The PO approach route (R1) typically assesses ignorability directly from the joint distribution $P(X,Y_x,Z)$; the classical graphical approach assesses ignorability based on the causal diagram (R2). The new proposal in this paper elicits ignorability from a cluster diagram over three nodes (R3).

Effect estimation remains an active field of research, with several interesting recent developments [8, 17].

The second, alternative framework — rooted in early ideas from genetics [35] and in the Cowles Commission's econometric modeling of interdependent systems [12], and further developed in computer science — is known as the structural approach to causality, pioneered by Judea Pearl [24]. This framework begins from a different premise: mechanisms. It posits a data-generating process formalized as a structural causal model (SCM), which encodes the causal processes through which data arise. The SCM provides semantics for various causal objects, including both factual and counterfactual distributions, thereby supporting reasoning under hypothetical interventions, a key requirement for scientific understanding and policy analysis. Fundamentally, this approach aligns with the central aim of scientific discovery in that it describes phenomena through a mechanistic perspective. [1] A distinctive feature of the approach is its use of explicit, nonparametric graphical representations — where variables are nodes in a directed acyclic graph (DAG) — to express assumptions about the underlying causal mechanisms, making them an accessible abstraction for data scientists to reason about cause and effect. Within this framework, causal inference typically proceeds in two steps. The first, known as identification, determines whether a causal query can be uniquely computed from available data and structural assumptions encoded in the DAG. To this end, the structural approach offers powerful tools such as the interventional (do-) calculus [23] and the counterfactual calculus [10], which enable the derivation of arbitrary interventional and counterfactual quantities. A widely used identification criterion in this framework is the back-door criterion, which determines whether a set of variables that causally precede both the treatment variable X and the outcome Y suffices to block spurious associations. When this criterion is met, it identifies an admissible adjustment set for estimating the causal effect of X on Y using observational data. [2]

[1] As Marschak observed, "The aim of economic policy is to choose among actions, which requires prediction under hypothetical or counterfactual assumptions. Structural models serve this purpose because they represent the mechanism generating the data" [20]. Similarly, Haavelmo argued that "the reason why we are interested in structural relations is that they enable us to perform controlled experiments (at least conceptually), which provide the only way to answer the typical questions of economic policy" [12]. While the spirit of these foundational insights is preserved in our approach, namely, that data arise from underlying causal mechanisms and that policy analysis requires structural semantics, we depart from the Cowles-style practice of trying to learn or requiring fully specified models. That tradition, though conceptually rigorous, faced severe empirical difficulties, including unstable parameter estimates and poor predictive performance across policy regimes (e.g., see [19, 9]). This historical experience echoes the formal content of the CHT discussed earlier, which shows that learning the full SCM is generally impossible without strong assumptions. Our approach retains structural commitments where they are informative and tractable, but avoids the pitfalls of overcommitted, fully mechanistic modeling. We thank Guido Imbens for helpful discussion on this historical distinction.

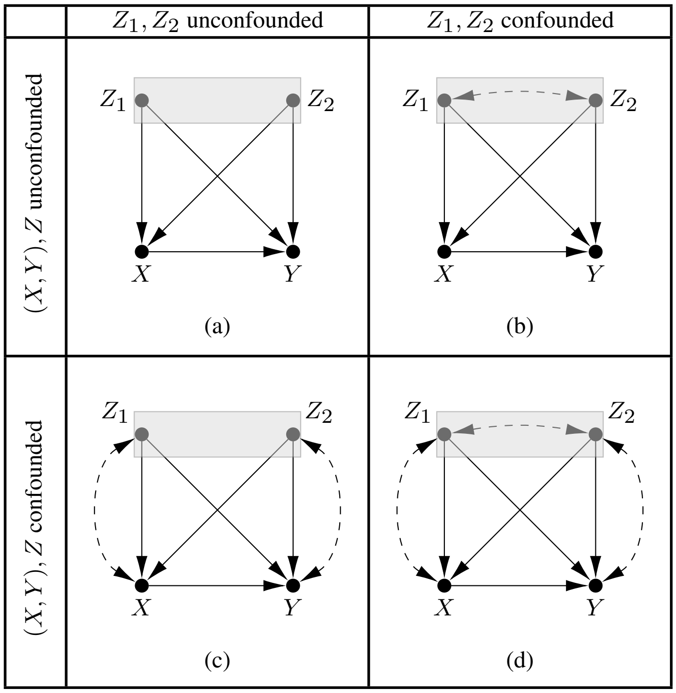
[2] We note that the identification machinery in the graphical approach does not end with the back-door criterion. Numerous examples

1.1 Limits of PO Assumptions and the Need for Structural Grounding

While the back-door criterion aligns closely with the conditional ignorability assumption used in the PO framework, their operationalization differs substantially. In the PO approach, ignorability must be assessed directly in algebraic form, typically through the independence statement $Y_x \perp\!\!\!\perp X \mid Z$ (within the gray region in Fig. 1). In contrast, the structural-graphical approach ties such assumptions to a causal diagram, allowing for systematic verification via graphical criteria.

Formally, the structural approach encompasses the PO framework, with each object in the analysis assigned proper semantics (as illustrated in Fig. 1, blue region). However, in practice, the latter's algebraic assumptions often lack semantic grounding. Analysts are asked to accept independence statements over counterfactual outcomes — without a mechanism to evaluate or interpret them transparently. This becomes especially problematic when deciding which covariates suffice to control for confounding.

To illustrate the challenge, consider the four diagrams in Fig. 2, each involving a treatment X, an outcome Y, and a set of covariates $\{Z_1, Z_2\}$. In all cases, the data scientist is expected to assess whether the conditional ignorability statement $Y_x \perp\!\!\!\perp X \mid Z$ holds, treating $Z = \{Z_1, Z_2\}$ as a single block. In the top row, this assumption is valid regardless of whether Z_1 and Z_2 are confounded, as in Fig. 2(b) (bidirected arrows). But in the bottom row, the same assumption fails to hold in the presence of additional confounding (as in Fig. 2(d)). The language of conditional ignorability is ill-suited to express such distinctions. As we will elaborate further, it treats Z monolithically, ignoring the internal structure that may be critical for deciding whether adjustment is valid. Consequently, it cannot distinguish between examples (a) and (b), or between (c) and (d) — despite these differences being decisive for identification. Analysts who rely on ignorability must make such distinctions implicitly, without formal tools to support them.



(a)
(b)
(c)
(d)

Figure 2: Motivating example, where X and Y represent the treatment and the outcome, and $\{Z_1, Z_2\}$ a set of confounders. The shaded-gray area illustrates what is considered a block, where the structure is often abstracted away by the data scientist.

This lack of transparency has been noted by some scholars. For instance, Judea Pearl argues that "it is almost impossible to articulate the ignorability statement in a language familiar to scientists" [25]. Marshall Joffe similarly observes that "such assumptions are usually made casually, largely because they justify the use of available statistical methods and not because they are truly believed" [16]. In defense, Guido Imbens notes that the assumption is "so common and well studied that merely referring to its label is probably sufficient" [15, Sec. 4.4]. Yet this current practice of invoking conditional ignorability without reference to an underlying structural model conceals a deeper concern: the algebraic statement $Y_x \perp\!\!\!\perp X \mid Z$, while concise, lacks meaningful semantics unless it is supported by explicit causal structure. A more principled foundation may require minimal but explicit causal commitments to assess these assumptions rigorously. To reconcile these mismatches while retaining operational clarity, we now introduce a structured yet parsimonious alternative.

1.2 Our Proposal: the $\mathrm{CG}(3)$ Model

The discussion so far has revealed a core tension: while the structural-graphical approach promotes transparency through diagrams, it often demands explicit assumptions; the PO framework, by contrast, is perceived as requiring fewer commitments, but its assumptions are algebraically opaque and combinatorially complex. To address this mismatch, we propose a middle-ground solution based on a graphical abstraction of the space of SCMs we call the $\mathrm{CG}(3)$ model. This construct retains the transparency and semantics of the graphical approach while offering a practical analogue to conditional ignorability. Instead of specifying a full causal diagram, we

show that where no back-door adjustment set exists, yet the identification of causal effects is still possible from observational data [23]; sound and complete algorithms for identification have also been developed [34, 33, 13, 5, 18]. Beyond the challenge of confounding bias, the discussion extends to a broader class of data fusion problems, including selection bias, external validity, and others, as surveyed in [6].

work with a coarser object defined over three conceptual blocks — X (treatment), Y (outcome), and Z (covariates) — that captures only the direct and bidirected relationships among these blocks. As shown in Fig. 3(a), the fully connected $\mathrm{CG}(3)$ includes six possible edges (three directed and three bidirected), and specific assumptions are represented by removing edges. This design allows analysts to articulate minimal causal commitments, avoiding the perceived rigidity of full DAGs, while still enabling principled graphical assessment of ignorability-like conditions.

DEFINITION 1 (Informal – Cluster Causal Diagram $\mathrm{CG}(3)$). *Let X be the treatment variable, Y the outcome, and Z the set of confounders. The class of cluster causal diagrams over three nodes is defined as the set of all subgraphs of the graph in Fig. 3(a), where the causal ($\rightarrow$) and confounding ($\dashleftarrow\!\dashrightarrow$) arcs between the pairs (X,Y), (X,Z), and (Y,Z) may be removed under additional assumptions.* □

This abstraction offers two major advantages. First, it enables a systematic graphical criterion for assessing conditional ignorability without requiring full specification of the internal structure of variables in Z. In particular, $\mathrm{CG}(3)$ supports a form of back-door analysis: whenever (i) there is no bidirected arc between X and Y, and (ii) at least one of the bidirected arcs between X and Z or Y and Z is missing, then the model implies the conditional independence $Y_x \perp\!\!\!\perp X \mid Z$. Second, the $\mathrm{CG}(3)$ framework preserves alignment with the PO perspective by focusing on the same key variables (X, Y, Z), but provides structural semantics that clarify which ignorability assumptions are justified. Rather than assuming away all potential confounding at once, this model allows for incremental exclusion of paths via explicit edge removals, making the assumption process more transparent and modular.

This modeling approach offers a practical alternative to directly assessing independence statements over counterfactuals — an exercise often more opaque and error-prone than practitioners realize. The example in Fig. 2 illustrates this point: diagrams (a) and (b) correspond to the $\mathrm{CG}(3)$ structure in Fig. 3(b), where conditional ignorability is implied via the back-door criterion. In contrast, diagrams (c) and (d) map to Fig. 3(c), where the assumptions required for ignorability are no longer entailed. This mapping shows how the $\mathrm{CG}(3)$ framework helps distinguish cases that would otherwise be conflated in the PO formalism.

At the same time, this abstraction reflects a deliberate compromise. The $\mathrm{CG}(3)$ framework explicitly acknowledges that eliciting the internal structure among covariates in Z may be infeasible. As a result, structurally distinct cases — like those in Fig. 2(c) and (d) — are intentionally coarsened to the same abstract object, reflecting a deliberate modeling choice. In contrast, the PO framework treats Z as a block but still asks the analyst to distinguish between these cases without providing the tools necessary to do so. This disconnect places an unrealistic burden on the data scientist's judgment. By striking a balance between abstraction and expressiveness, the $\mathrm{CG}(3)$ model offers a principled middle ground: it retains the operational simplicity of the PO framework while restoring transparency through graphical semantics.

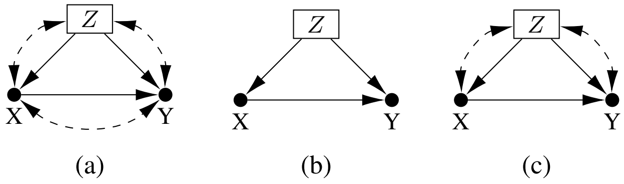


Figure 3: Causal diagrams in the $\mathrm{CG}(3)$ class.

The remainder of the paper develops the proposed framework in depth. Section 2 introduces the structural machinery and semantics of potential outcomes. Section 2.1 interprets causal diagrams as abstractions over the space of SCMs. Section 2.2 provides a structural reading of conditional ignorability. Section 3 explores the mismatch between graphical models and algebraic assumptions. Finally, Section 3.1 formalizes cluster diagrams and structural ignorability as a transparent middle ground.

2. PRELIMINARIES

We start by introducing a general class of generative models known as structural causal models [24], which will act as the basic semantical framework of our discussion. We will follow the presentation and results as developed in [4].

DEFINITION 2 (Structural Causal Model (SCM) [24, 4]). *A structural causal model (SCM) $\mathcal{M}$ is a 4-tuple $\langle V, U, \mathcal{F}, P(u)\rangle$, where*

(1) U is a set of exogenous variables, also called background variables, that are determined by factors outside the model;
(2) $V = \{V_1, ..., V_n\}$ is a set of endogenous (observed) variables, that are determined by variables in the model (i.e. by the variables in $U \cup V$);
(3) $\mathcal{F} = \{f_1, ..., f_n\}$ is the set of structural functions determining V, $v_i \leftarrow f_i(\mathrm{pa}(v_i), u_i)$, where $\mathrm{pa}(V_i) \subseteq V \setminus V_i$ and $U_i \subseteq U$ are the functional arguments of f_i;
(4) $P(u)$ is a distribution over the exogenous variables U.

We denote by Ω the space of all instantiations of all structural causal models (SCMs) over a fixed set of endogenous variables. □

The SCM represents the “ground truth" about the underlying causal phenomenon. In particular, the assignment mechanisms $\mathcal{F}$ determine how each of the observed variables V_i attains its value, based on other observed variables and the latent variables U. Together with the probability distribution $P(u)$ over the exogenous variables U, it specifies the behavior of the underlying phenomenon. We note that the SCM will almost never be fully observed and known to the data scientist. Still, it will provide a baseline from which we can derive formal semantics and essential properties of the system under study. For instance, the SCM specifies the *observational distribution* of the underlying phenomenon, defined as follows:

DEFINITION 3 (Observational Distribution [4]). *An SCM $\mathcal{M}$ that is a 4-tuple $\langle V, U, \mathcal{F}, P(u)\rangle$ induces a joint probability distribution $P(V)$ such that for each $Y \subseteq V$,*

$$P^{\mathcal{M}}(y) = \sum_u \mathbb{1}\Big(Y(u) = y\Big)P(u), \tag{1}$$

where $Y(u)$ is the solution for Y after evaluating $\mathcal{F}$ with $U = u$. □

An important related notion building on the concept of the SCM is that of a submodel, defined next:

DEFINITION 4 (Submodel [24]). *Let $\mathcal{M}$ be a structural causal model, X a set of variables in V, and x a particular value of X. A submodel $\mathcal{M}_x$ (of $\mathcal{M}$) is a 4-tuple:*

$$\mathcal{M}_x = \langle V, U, \mathcal{F}_x, P(u)\rangle \tag{2}$$

where

$$\mathcal{F}_x = \{f_i : V_i \notin X\} \cup \{X \leftarrow x\}, \tag{3}$$

and all other components are preserved from $\mathcal{M}$. □

In words, the SCM $\mathcal{M}_x$ is obtained from $\mathcal{M}$ by replacing the equations in $\mathcal{F}$ related to variables X by equations that set X to a specific value x. This corresponds to setting the value of $X = x$ in the model, which is written through the do-operator, do$(X = x)$. Building on submodels, the notion of a potential outcome (or potential response) follows naturally:

DEFINITION 5 (Potential Outcome [30, 24]). *Let X and Y be two sets of variables in V and $u \in \mathcal{U}$ be a unit. The potential outcome (or potential response) $Y_x(u)$ is defined as the solution for Y of the set of equations $\mathcal{F}_x$ evaluated with $U = u$. That is, $Y_x(u)$ denotes the solution of Y in the submodel $\mathcal{M}_x$ of $\mathcal{M}$.* □

In words, $Y_x(u)$ is the value variable Y would take if (possibly contrary to observed facts) X is set to x, for a specific unit $U = u$. Also related to the concept of a submodel, we next introduce the key concept of an interventional distribution:

DEFINITION 6 (Interventional Distribution). *Let $X, Y \subseteq V$ be disjoint sets of variables in an SCM $\mathcal{M}$. Then, the interventional distribution $P(Y \mid do(X = x))$ denotes the distribution of Y in the submodel $\mathcal{M}_x$.* □

Consider a simple two-variable causal diagram $X \to Y$. There is an immediate way of seeing how potential outcomes are defined based on SCMs, with the potential outcome Y_x being given by

$$Y_x \leftarrow f_Y(x, U_y), \tag{4}$$

and the distribution of $P(Y_x) = P(Y \mid do(X = x))$ is implied by this equation. In other words, the mechanisms in the SCM $\mathcal{M}$ span a set of potential outcomes computed from the corresponding submodels. We next define the notion of a joint counterfactual distribution:

DEFINITION 7 (Counterfactual Distributions [4]). *Let $\mathcal{M} = \langle V, U, \mathcal{F}, P(u)\rangle$ be an SCM, and let $Y_1, \dots, Y_k \subset V$, and $X_1, \dots, X_k \subset V$ be subsets of the observables, and let $x_1, \dots, x_k$ be specific values of $X_1, \dots, X_k$. Denote by $(Y_i)_{x_i}$ the potential response of variables Y_i when setting $X_i = x_i$. The SCM $\mathcal{M}$ induces a family of joint distributions over counterfactual events $(Y_1)_{x_1}, \dots, (Y_k)_{x_k}$, with $P^{\mathcal{M}}((y_1)_{x_1}, \dots, (y_k)_{x_k})$ defined via:*

$$\sum_u \mathbb{1}\Big(\bigwedge_{i=1}^{k} (Y_i)_{x_i}(u) = y_i\Big)P(u). \tag{5}$$

□

The distribution $P^{\mathcal{M}}((y_1)_{x_1}, \dots, (y_k)_{x_k})$ contains variables with different subscripts, which syntactically represent different potential outcomes (Def. 5), or counterfactual worlds. It may be instructive to apply this definition to the distribution $P(Z = z, X = x', Y_x = y)$, which is defined through

$$\sum \mathbb{1}\Big(Z(u) = z, X(u) = x', Y_x(u) = y\Big)P(u) \tag{6}$$

In words, in Eq. 6, for each unit $u \in \mathcal{U}$, we check:

(1) whether $Z(u) = z, X(u) = x'$ for the unit naturally, and
(2) whether $Y_x(u) = y$, that, in the modified SCM with the mechanism $\mathcal{F}_x$ for X replaced by a fixed value $X = x$, whether the solution for $Y(u)$ equals y,

and the probability mass $P(u)$ for each such unit is added. The counterfactual distribution $P(Z = z, X = x', Y_x = y)$, defined through Eq. 6, is the core building block from which the conditional ignorability statement needs to be judged, a point to which we return in Sec. 2.2.

As alluded to earlier, the mechanisms $\mathcal{F}$ and the distribution over the exogenous variables $P(u)$ are almost never observed in practice, despite their existence. However, to perform causal inference, we need a way to encode assumptions about the underlying SCM. A common method for representing such assumptions is the use of a causal diagram, which highlights one of the fundamental operational differences between the potential outcomes and the structural-graphical approaches to causality.

2.1 Causal Diagrams as a Modeling Tool for Coarsening the Space of SCMs Ω

Causal diagrams are formal constructs whose semantics are grounded in an underlying SCM as follows:

DEFINITION 8 (Causal Diagram [24, 4]). *Let an SCM $\mathcal{M}$ be a 4-tuple $\langle V, U, \mathcal{F}, P(u)\rangle$. A graph $\mathcal{G}$ is said to be a causal diagram (of $\mathcal{M}$) if:*

(1) there is a vertex for every endogenous variable $V_i \in V$,

(2) there is an edge $V_i \rightarrow V_j$ if V_i appears as an argument of $f_j \in \mathcal{F}$,

(3) there is a bidirected edge $V_i \dashleftarrow\dashrightarrow V_j$ if the corresponding $U_i, U_j \subset U$ are correlated or the corresponding functions f_i, f_j share some $U_{ij} \in U$ as an argument. □

The causal diagram can be understood as a specific type of coarsening operation over the space of structural causal models (Ω). It retains information about: (1) the functional arguments of the structural mechanisms $\mathcal{F}$, but not their functional forms; and (2) the independence relations between exogenous variables U, while it does not contain information about the specific distribution over the exogenous variables ($P(u)$). In particular, there is an edge from an endogenous variable V_i to V_j whenever V_j "listens to" V_i to determine its value. Similarly, a bidirected edge between V_i and V_j indicates shared, unobserved information affecting how both variables obtain their values.

While the SCM makes explicit both the functional mechanisms ($\mathcal{F}$) and the distribution over exogenous variables ($P(u)$), the causal diagram retains only qualitative features of each. That is, the diagram abstracts out the specifics of the functions $\mathcal{F}$ and distribution $P(u)$, while retaining information about their arguments and independence relations, respectively. Therefore, constructing a causal diagram can be understood as establishing a many-to-one mapping from the space of SCMs Ω to the space of causal diagrams, since many different SCMs map to the same diagram. In Fig. 4(a), this coarsening is visualized: multiple SCMs within Ω are grouped into a single causal diagram, represented by an oval shape.

Therefore, instead of encoding assumptions at the level of the SCM (which may be too challenging and require highly granular knowledge), assumptions are articulated in a coarser manner through a causal diagram. [3] The diagram can then be subsequently used to license the identification of a causal effect through different strategies. One prominent example of this is through the celebrated criterion known as the back-door:

DEFINITION 9 (Back-door Criterion [24]). *A set of variables Z satisfies the back-door criterion relative to an ordered pair of variables (X, Y) in a causal diagram $\mathcal{G}$ if:*

(i) no node in Z is a descendant of X,

(ii) Z blocks every path between X and Y that contains an arrow into X.

Here a path $X - \cdots - Y$ is said to be blocked by Z if either there exists $V_i \in Z$ along the path with at least one outgoing arrow, or there exists a V_j with both incoming arrows ($\ldots \rightarrow V_j \leftarrow \ldots$, usually known as collider) such that neither V_j nor any of its descendants are in Z. □

The back-door criterion, applied to the causal diagram, allows one to identify interventional distributions (and consequently causal effects):

PROPOSITION 1 (Back-door Identification). *If there exists a set of variables Z satisfying the back-door criterion relative to an ordered pair of variables (X, Y) in a causal diagram $\mathcal{G}$, then it follows that the distribution $P(Y \mid do(X = x))$ is identifiable and given by the formula*

$$P(y \mid do(X = x)) = \sum_z P(y \mid x, z) P(z), \tag{7}$$

that is, it can be computed uniquely from observational data and the causal diagram $\mathcal{G}$ through the adjustment formula. □

The process of identifying interventional distributions in the structural-graphical approach to causal inference can be understood as consisting of two steps. In the first step, assumptions are elicited in the form of a causal diagram. In the second step, graphical criteria (such as the back-door criterion) are applied to determine whether the query of interest is identifiable. This process of identification corresponds to the route R2 highlighted in Fig. 1. The strength of this approach lies in its ability to provide a powerful and systematic tool for eliciting assumptions — using the metaphor of the SCM to capture causal relations

[3] This is, in fact, a fundamental point about local versus global constraints and the original motivation for using graphical models to encode probabilistic knowledge—and later, to formalize modularity—a key insight from the early literature [22]. See also the discussions around Def. 16 and footnotes 47 and 48 in [4].

among variables based on domain expert knowledge, expressed through a causal diagram. The main challenge, however, is that constructing the causal diagram may still be demanding in practice, as it requires domain knowledge that, while less stringent than what is needed to fully specify an SCM, may nonetheless be highly granular and non-trivial to obtain.

2.2 Conditional Ignorability as a Tool for Coarsening the Space of SCMs Ω

We next contrast the structural-graphical approach to causal effect identification with the approach taken in the PO framework. To do so, we formally introduce the notion of conditional ignorability, which is commonly invoked in this literature:

DEFINITION 10 (Conditional Ignorability). *Let X be a treatment and Y the outcome. We say that the ordered pair (X,Y) satisfies the ignorability criterion conditional on a set Z if the following independence holds:*

$$Y_x \perp\!\!\!\perp X \mid Z, \tag{8}$$

where x is a fixed value of X. □

We provide a similar interpretation of conditional ignorability (C-Ign, for short), analogous to the coarsening described for causal diagrams. The idea is to introduce a common denominator — grounded in the space Ω — that allows these two approaches to be related in a unified framework. The notion of C-Ign can be understood as a mapping from Ω to the set $0,1$, where C-Ign$(M) = 1$ whenever $Y_x \perp\!\!\!\perp X \mid Z$ holds in the SCM M. In this way, C-Ign partitions Ω into two subsets: one in which the independence statement $Y_x \perp\!\!\!\perp X \mid Z$ is true, and one in which it is not. This partition is indicated by the red line in Fig. 4(a). C-Ign allows one to identify a causal effect through a well-known result:

PROPOSITION 2 (Conditional Ignorability Identification). *If the ordered pair (X,Y) satisfies the conditional ignorability criterion with respect to the set Z, then the effect of X on Y is identified by:*

$$P(Y_x) = \sum_z P(y \mid x,z)P(z). \tag{9}$$

□

Therefore, the purpose of invoking a C-Ign is to determine for which cases interventional distributions can be identified from observational data. An important connection of C-Ign to the back-door criterion is given by the following result:

PROPOSITION 3 (Back-door Criterion $\implies$ Conditional Ignorability). *If there exists a set of variables Z that satisfies the back-door criterion relative to an ordered pair of variables (X,Y) in a causal diagram $\mathcal{G}$, then it follows that the pair (X,Y) satisfies the ignorability criterion conditional on Z.* □

The back-door allows the data scientist to assess whether an effect is identifiable through the adjustment expression based on the set Z. As Prop. 3 also shows, the back-door criterion implies the ignorability statement. This result, together with Fig. 4(a), helps clarify the difference between the two approaches. The graphical approach begins by coarsening the space Ω, partitioning it into equivalence classes corresponding to different causal diagrams. Then each class can be evaluated for whether it supports identification through adjustment on Z (or by other means).

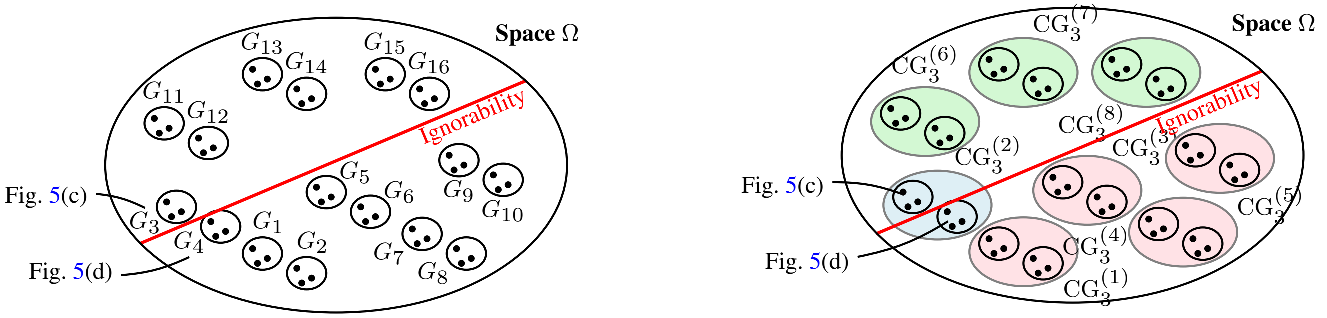
Alternatively, the conditional ignorability criterion skips the step of explicitly modeling causal connections and instead partitions Ω into two regions — depending on whether the independence statement $Y_x \perp\!\!\!\perp X \mid Z$ holds in the counterfactual distribution $P(X,Y_x,Z)$. The process of identification for conditional ignorability corresponds to the route R1 highlighted in Fig. 1. Although this approach may appear appealing, since it bypasses the need to elicit assumptions encoded in a causal diagram, this simplicity comes at the cost of transparency. The C-Ign approach does not offer a systematic procedure for assessing the assumption's validity; that is, it does not provide a sequence of steps by which a human analyst might judge whether the conditional ignorability assumption is justified.

Furthermore, as we illustrate next, the boundary induced by C-Ign (red line in Fig. 4(a)) is more complex than commonly appreciated. This means that determining whether C-Ign holds in a given setting may require granular domain knowledge, which is unlikely to be accessible to a data scientist attempting to evaluate ignorability.

3. STRUCTURAL CONDITIONAL IGNORABILITY

We now take a closer look at how ignorability statements are assessed in practice, and illustrate through examples that the level of knowledge required to assess these statements can be very granular, contrary to popular belief. Therefore, in the absence of formal tools, we argue that applying ignorability in its current form may be difficult in practice. Ultimately, we propose an alternative approach that enables the assessment of ignorability through a graphical model, while avoiding the need to fully specify the causal diagram.

We begin by analyzing the separation of the ignorability statement into blocks, to illustrate the mental construct evoked by scientists when judging the plausibility



(a) Coarsening of ignorability & causal diagrams.

(b) Coarsening implied by structural ignorability.

Figure 4: Visualization of different coarsening approaches discussed in the paper. Each black point in the space of SCMs Ω represents an SCM. In (a), different SCMs are joined into groups if they have the same causal diagram, represented by black oval shapes. The notion of conditional ignorability splits the space Ω into two parts, based on whether the required conditional independence statement holds. In (b), the coarsening related to structural ignorability and $\mathrm{CG}(3)$ model (Def. 11) is visualized. Here, multiple causal diagrams are mapped to a single cluster diagram $\mathrm{CG}(3)$. Notably, examples discussed in Fig. 5(c, d) fall on opposite sides of the ignorability boundary, but belong to the same $\mathrm{CG}(3)$ model.

of such assumptions. Three distinct blocks are invoked in this analysis, reproduced below for visual convenience:

$$\underbrace{Y_x}_{\text{block } B_1} \perp\!\!\!\perp \underbrace{X}_{\text{block } B_2} \mid \underbrace{Z}_{\text{block } B_3} . \tag{10}$$

The block B_1 corresponds to the variable Y in the hypothetical regime $do(X = x)$, B_2 to the treatment variable X, and B_3 to the set of covariates Z. Elaborating on Fig. 2, we highlight a fundamental limitation of this mental construct through the analysis of the following examples:

EXAMPLE 1. *We consider the models in Fig. 5(a-b), where X is the treatment, Y is the outcome, and Z_1, Z_2 are the set of observed confounders. We analyze the pairwise relations between the blocks B_1 (Y_x), B_2 (X), and B_3 (Z) following Eq. 10:*

(i) B_1-B_2 relation: in both models, X may cause Y (but not the other way around), and there are no unobserved confounders between X and Y, but all confounders are observed and equal to B_3,

(ii) B_2-B_3 relation: in both models, Z_1, Z_2 affect X in the same fashion – Z may cause X (but not the other way around), and there are no unobserved confounders between these blocks of variables,

(iii) B_1-B_3 relation: again, in both models, Z_1, Z_2 affect Y in the same fashion – Z may cause Y (but not the other way around) and there are no unobserved confounders between these blocks of variables.

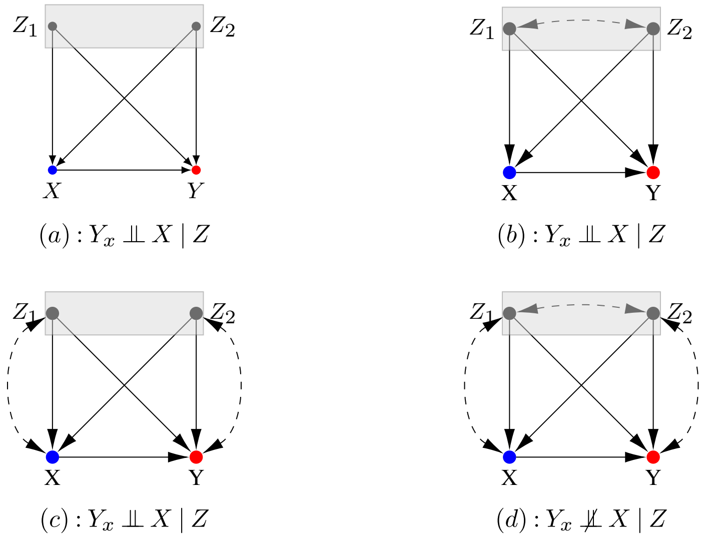
Therefore, upon closer inspection, we see that all the inter-block relationships are the same for models in Fig. 5(a) and (b). The relations between blocks are summarized in the table in the first row of Fig. 5 (right side).

Based on the above, one would be tempted to believe that ignorability holds (Eq. 10) since, in a pairwise manner, there is no unobserved confounder between any of the blocks. In fact, ignorability does hold in both models in Figs. 5(a) and (b), and this conclusion holds regardless of the relationship between the variables Z_1 and Z_2, which belong to block B_3. The shaded area that includes Z_1 and Z_2 is highlighted graphically to show that the modeler trying to assess ignorability is agnostic to the relationship among these variables, and simply treats them as a single block. □

In the above example, we see how the analyst can arrive at a correct conclusion, and at the same time abstract away the seemingly irrelevant causal structure within the ignorability blocks. Unfortunately, as the following example demonstrates, ignoring the structure within a block may also lead to invalid conclusions.

EXAMPLE 2. *Consider now the models in Fig. 5(c-d), which are similar to Fig. 5(a-b). In particular, the relations within the shaded area remain the same – in models (a) and (c), Z_1 and Z_2 are independent, while in models (b) and (d), they have an unobserved common cause (dashed arrow). Now we analyze the relations across blocks.*

(i) B_1-B_2: in both models, X may cause Y (but not the other way around), and there are no unobserved confounders between them, i.e., all exogenous variations go through B_3,



$(a) : Y_x \perp\!\!\!\perp X \mid Z$

$(b) : Y_x \perp\!\!\!\perp X \mid Z$

$(c) : Y_x \perp\!\!\!\perp X \mid Z$

$(d) : Y_x \not\!\perp\!\!\!\perp X \mid Z$

Cases (a,b)	Causal	Confounded
(Z, X)	✓	✗
(Z, Y)	✓	✗
(X, Y)	✓	✗ (thr. Z)

(a)-(b) block specification

Cases (c,d)	Causal	Confounded
(Z, X)	✓	✓
(Z, Y)	✓	✓
(X, Y)	✓	✗ (thr. Z)

(c)-(d) block specification

Figure 5: Causal diagrams and blocks specifications for Ex. 1 and Ex. 2.

(ii) B_2-B_3: in both models, Z_1 and Z_2 may cause X (but not the other way around), and there exists unobserved confounding between $\{Z_1, Z_2\}$ and X,

(iii) B_1-B_3: in both models, Z_1, Z_2 may cause Y (but not the other way around), and there exists unobserved confounding between $\{Z_1, Z_2\}$ and Y.

Again, the relations between the blocks are identical for models in Fig. 5(c) and (d), as summarized in the table in Fig. 5 (bottom-right). The difference between models in Fig. 5(c) and (d) is within-blocks, namely that in 5(c) Z_1, Z_2 are independent, whereas in 5(d) they are not.

When analyzing these examples, the analyst may be tempted to also abstract away the structure between the variables Z_1 and Z_2, i.e., within the block B_3, since this was a successful strategy in Ex. 1. Furthermore, the analyst may also be tempted to surmise that conditional ignorability holds as in Ex. 1, arguing that all confounding between X and Y pass through confounders Z_1, Z_2, which therefore could be removed by controlling for them.

Based on the specific causal diagrams, we now verify the back-door criterion model by model. Starting with the model in Fig. 5(c), the back-door criterion states that the set $\{Z_1, Z_2\}$ is back-door admissible since $(X \perp\!\!\!\perp Y \mid Z)_{G_{\underline{X}}}$. On the other hand, when considering Fig. 5(d), conditioning on the set $\{Z_1, Z_2\}$ opens the path $X \dashleftarrow\dashrightarrow Z_1 \leftarrow\!\!\rightarrow Z_2 \leftarrow\!\!\rightarrow Y$, meaning that X, Y are not d-separated by $\{Z_1, Z_2\}$.

The critical observation here is that the relationships between the blocks are the same in both (c) and (d), while ignorability holds in (c), but not in (d). The difference between the models is strictly within the block B_3, which ignorability does not take into account. □

In light of the above examples, we note that judgments about ignorability seem to provide conflicting answers since they systematically ignore fine-grained level of knowledge about the relationship inside the set of covariates (which is called a block in our terminology). At the same time, treating variables as blocks is one of the motivations for evoking ignorability statements in the first place. For some graphs, within block relationships can indeed be ignored, and the answer may not depend on these relations, such as in the models in Fig. 5(a, b). In other cases, there are graphs which share the same relationships across all the blocks but lead to different ignorability statements, depending on the relationships within a block, as in Fig. 5(c,d). Relating this back to the visualization in Fig. 4(a), we see that examples (c), (d) are on opposite sides of the ignorability line, and to be able to distinguish these cases, highly granular knowledge about the structure within the Z-block is required. Without a systematic way of eliciting assumptions, it is unlikely that the analyst can perform this task, and for this reason, we see that the boundary that delineates whether ignorability holds true or not is much more difficult to navigate than commonly believed. We next consider an empirical example that further grounds the above discussion and illustrates the possible implications of dismissing relations within blocks of variables.

EXAMPLE 3. *Consider the following SCMs $\mathcal{M}^{(i)}$ over endogenous variables $V = \{Z, X, Y\}$ and exogenous $U = \{U_z, U_x, U_y, U_{zx}, U_{zy}\}$. The structural mechanisms $\mathcal{F}$ are linear, and given by*

$$\mathcal{F} : \begin{cases} Z \leftarrow U_z + (U_{xz}, U_{zy})^T & (11) \\ X \leftarrow \alpha^T Z + U_x + U_{xz} & (12) \\ Y \leftarrow \beta^T Z + \gamma X + U_y + U_{zy}, & (13) \end{cases}$$

with Gaussian probability distributions over the exogenous variables, i.e.,

$$P(u):\begin{cases} U_x, U_y \sim N(0,1) & \text{(14)}\\ U_z \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho^{(i)}\\ \rho^{(i)} & 1\end{pmatrix}\right) & \text{(15)}\\ U_{xz}, U_{zy} \sim N(0,\lambda^{(i)}), & \text{(16)}\end{cases}$$

where the parameters $\rho^{(i)}, \lambda^{(i)}$ are given by:

$\mathcal{M}^{(1)}$: $\rho^{(i)} = 0,\ \lambda^{(i)} = 0$ *corresponding to Fig. 5(a),*
$\mathcal{M}^{(2)}$: $\rho^{(i)} = 0.5,\ \lambda^{(i)} = 0$ *corresponding to Fig. 5(b),*
$\mathcal{M}^{(3)}$: $\rho^{(i)} = 0,\ \lambda^{(i)} = 1$ *corresponding to Fig. 5(c),*
$\mathcal{M}^{(4)}$: $\rho^{(i)} = 0.5,\ \lambda^{(i)} = 1$ *corresponding to Fig. 5(d).*

The inferential challenge is, of course, that the data scientist does not have access to the true SCM, which includes the values of structural parameters $\rho^{(i)}, \lambda^{(i)}$. Firstly, we consider the cases (1) and (2) that correspond to causal diagrams in Fig. 5(a) and (b) discussed in Ex. 1. The goal is to estimate the effect of X on Y, represented by the structural coefficient γ. The scientist may estimate such an effect through linear regression (say using ordinary least squares, OLS). In Fig. 6 (top row), we plot the density of the OLS estimator $\hat{\gamma}$ over different repetitions of $n = 1{,}000$ samples generated from the SCM in Eq. 11-16 (that is, from the observational distribution), with the dashed-red line representing the true effect. Interestingly, regardless of dismissing the different relationships between Z_1 and Z_2 – $\rho^{(i)}$ is equal to 0 in the first model and different from 0 in the second – the structural coefficient γ is estimated consistently in both models. In other words, these differences could in fact be dismissed and assessing ignorability leads to correct adjustment – based on covariates Z.

*Secondly, we contrast that with cases (3) and (4) that correspond to causal diagrams Fig. 5(c) and (d) in Ex. 2, respectively. Using OLS again in this setting leads to the plots shown in Fig. 6 (bottom row), where the dashed-red line represents the true effect. The effect estimates are quite different in this case – one of the models has a positive effect while the other has a negative one. Thus, ignoring the structure within the set of variables **Z** may systematically lead to erroneous effects estimates.* □

The main takeaway from Exs. 1 and 2 is the following. Ignorability statements seemingly depend only on the blocks B_1, B_2, and B_3 and the relationships between them, which may lead the data scientist performing the analysis to believe that relationships within the set of covariates **Z** can be safely ignored. In Ex. 1, this strategy is, in fact, successful: valid conclusions about effect identification through ignorability are reached for both models (a) and (b). On the other hand, Ex. 2 reveals that the situation is more subtle. We introduced two models with the same structure between blocks, yet they differ with respect to their ignorability statements. Abstracting away the structure within blocks in Ex. 2 ignores the structural coefficient ρ, which plays a key role in determining whether ignorability holds. As a result, the estimated effect of X on Y using ordinary least squares (OLS), with Y regressed on X and Z, is correct in cases (1)–(3) of Fig. 5, but incorrect – and even of opposite sign – in case (4) (see Fig. 6). Clearly, this conclusion cannot be reached when assessing ignorability as in the potential outcomes approach, in the absence of a graphical model.

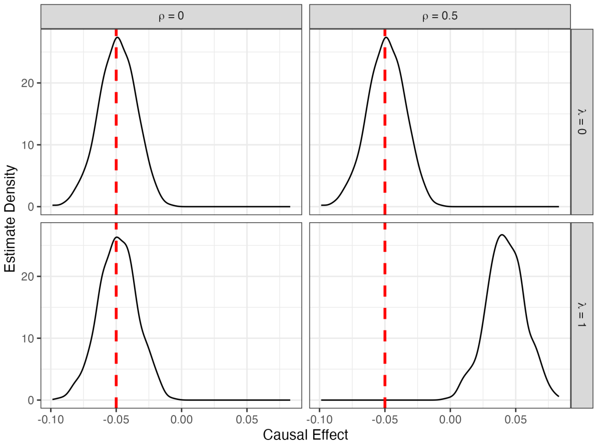


Figure 6: Density of the estimator of the causal effect of X on Y using OLS and adjusting for Z_1, Z_2. $n = 1000$ samples are generated from the system Eqs. (11)-(16) 500 times for $\rho \in \{0, 0.5\}, \lambda \in \{0, 1\}$. The dashed-red line indicates the true value of the causal effect.

Taken together, these examples illustrate that assessing ignorability may systematically lead to invalid conclusions when important structure within blocks is abstracted away. Still, one could argue that the judgment of what happens within a cluster **Z** may be carried out implicitly in the mind of the data scientist while inspecting Eq. 10. For instance, [15, Sec. 4.4] argues that, in the context of assessing ignorability, “adding a DAG is superfluous because researchers are familiar with the setting and its implications.” Consider a model shown in Fig. 7, which extends the previous discussion. Note that within the set of covariates **Z**, consisting of $|\mathbf{Z}| = n - 2$ variables, there are $\binom{n-2}{2}$ possible pairwise relationships that must be specified. Now, for any pair $Z_i, Z_j \in \mathbf{Z}$, there may be: (i) no relation between them; (ii) a causal relation; (iii) a confounding relation; or (iv) both a causal and a confounding relationship. This implies that the data scientist must implicitly evaluate an order of 4^{n^2} possible configurations. In fact, for a set of two covariates, there are only four possible relationships to consider, but for a set of five covariates, this number already grows to over a million configurations. The suggestion that such evaluation is superfluous—and can be carried out implicitly, without a clearer understanding of which relationships are or are not allowed and how they affect ignorability judgments—seems somewhat far-fetched.

As the preceding discussion suggests, the boundary delineated by ignorability (red line in Fig. 4(a)) is challenging to establish and characterize, and the level of knowledge needed to distinguish cases on either side can be rather strong, as illustrated by previous examples and the super-exponential number of assumptions that must be specified in the general case. To address this concern, we propose in the sequel a solution that strikes a balance between modeling flexibility and transparency in the encoding of assumptions, one that may prove appealing to proponents of both the PO and graphical frameworks.

3.1 On the Structural Basis of Conditional Ignorability

In this section, we introduce the notion of structural ignorability and connect the block-based construction of ignorability with the structural semantics of causal diagrams. We begin by providing a formal definition of the corresponding graphical object, grounding Def. 1.

DEFINITION 11 (Cluster Diagram over Three Blocks). *Consider a causal diagram G over Z, X, and Y, and assume that no element of Z is a descendant of X or Y, and that X is not a descendant of Y. Construct a graph* $\mathrm{CG}(3)$ *over X, Y, and Z with the following set of edges:*

(i) ***Directed edges:*** *an edge $Z \to X$ exists if $Z_i \in \mathrm{pa}(X)$ for any $Z_i \in Z$; an edge $Z \to Y$ exists if $Z_i \in \mathrm{pa}(Y)$ for any $Z_i \in Z$; an edge $X \to Y$ exists if $X \in \mathrm{pa}(Y)$.*

(ii) ***Bidirected edges:*** *a dashed bidirected edge between blocks B_i and B_j exists whenever there exist variables $V_i \in B_i$ and $V_j \in B_j$ that share a common exogenous (unobserved) cause.*

Any such graph is referred to as a cluster diagram over three nodes X, Y, and Z, denoted $\mathrm{CG}(3)$. □

The above definition introduces a type of *cluster causal diagram*, in which the set of confounders Z is collapsed into a single group. [4] To ground this notion further, we provide examples of how a causal diagram relates to its $\mathrm{CG}(3)$ representation.

EXAMPLE 4 (Causal Diagram to $\mathrm{CG}(3)$). *Consider the causal diagram in Fig. 5(a). Since no bidirected arcs across blocks $\{X\}, \{Y\}, \{Z\}$ exist, this causal diagram is mapped to the* $\mathrm{CG}(3)$ *model in Fig. 8(h). Similarly, the causal diagram in Fig. 5(b) is also mapped to the* $\mathrm{CG}(3)$ *model in Fig. 8(h). Therefore, when moving from the space of causal diagrams to the* $\mathrm{CG}(3)$ *representation, the existence of a bidirected edge $Z_1 \dashleftarrow\dashrightarrow Z_2$ is not considered, and can be abstracted away.*

For the examples in Figs. 5(c,d), the existence of the arrow $X \dashleftarrow\dashrightarrow Z_1$ implies the existence of the arrow $X \dashleftarrow\dashrightarrow Z$ in $\mathrm{CG}(3)$. *Similarly, the existence of the arrow $Y \dashleftarrow\dashrightarrow Z_2$ implies the existence of the arrow $Y \dashleftarrow\dashrightarrow Z$ in* $\mathrm{CG}(3)$. *Therefore, both diagrams are mapped to the* $\mathrm{CG}(3)$ *model in Fig. 8(b), and the structure within the block Z is, once again, ignored.* □

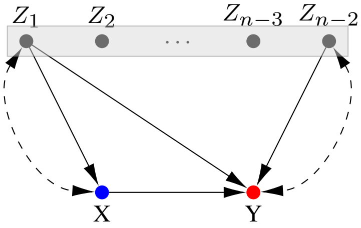


Figure 7: Extended graph entailing an exponential number of relations within the covariate set.

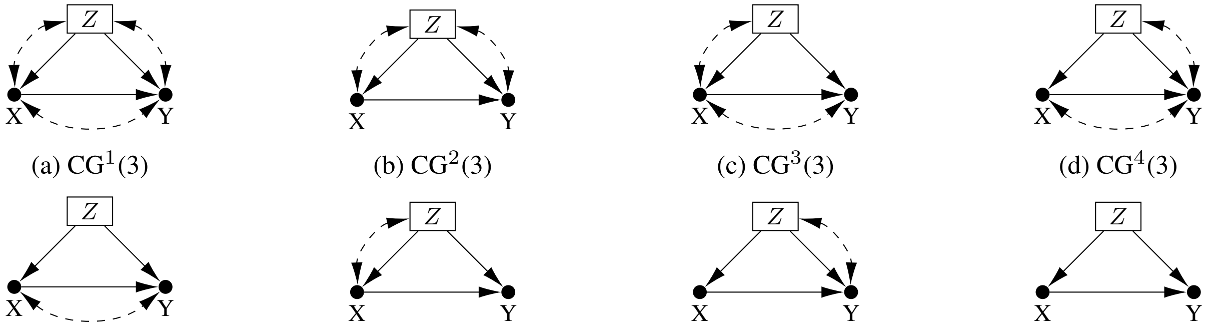
For simplicity, when considering a $\mathrm{CG}(3)$, we may assume that the directed edges are always present (i.e., $Z \to X$, $Z \to Y$, and $X \to Y$). Therefore, within the $\mathrm{CG}(3)$ class of models, we distinguish among the 8 possible graphs shown in Fig. 8. For instance, for $|Z| = 2$, the original space of causal diagrams contains 128 distinct elements (assuming a fully connected graph). As this suggests, the notion of a cluster causal diagram $\mathrm{CG}(3)$ represents a further coarsening of the space of SCMs Ω, compared to the coarsening induced by standard causal diagrams, as shown in Fig. 4(b). In particular, multiple causal diagrams may map to a single $\mathrm{CG}(3)$ representation, and in moving from a causal diagram to its $\mathrm{CG}(3)$ abstraction, some of the underlying structure of the original diagram is lost. In other words, a $\mathrm{CG}(3)$ diagram is strictly weaker than a standard causal diagram.

The key idea behind this new representation is that the process of eliciting assumptions is simplified—by eliciting the $\mathrm{CG}(3)$ diagram directly, rather than the full set of assumptions required for a complete causal diagram. Specifically, the data analyst needs to perform the following steps:

(1) Determine whether any unobserved common causes of X and Y exist; if not, the bidirected edge $X \dashleftarrow\dashrightarrow Y$ can be removed.

(2) Determine whether any unobserved common causes of X and Z_i exist, for any $Z_i \in Z$; if not, the bidirected edge $X \dashleftarrow\dashrightarrow Z$ can be removed.

[4] For simplicity, we focus on cluster graphs with three nodes, but this construction can be generalized for arbitrary graphs [1].



(a) CG^1(3)
(b) CG^2(3)
(c) CG^3(3)
(d) CG^4(3)
(e) CG^5(3)
(f) CG^6(3)
(g) CG^7(3)
(h) CG^8(3)

Figure 8: Collection of eight possible CG(3) cluster diagrams.

(3) Determine whether any unobserved common causes of Y and Z_i exist, for any $Z_i \in Z$; if not, the bidirected edge $Y \leftarrow\!\!-\!\!-\!\!\rightarrow Z$ can be removed.

The three steps above describe the core process of eliciting the assumptions required for ignorability-type identification. For example, consider the CG(3) model in Fig. 8(f), in which the existence of an exogenous variable U_{xy} affecting both X and Y has been ruled out, as well as U_{zy} affecting both Z and Y. However, an exogenous variable U_{xz} affecting X and Z may still exist.

In practice, this implies that the underlying structural causal model takes the form:

$$Z \leftarrow f_Z(U_z, U_{xz}) \tag{17}$$

$$X \leftarrow f_X(Z, U_x, U_{xz}) \tag{18}$$

$$Y \leftarrow f_Y(X, Z, U_y), \tag{19}$$

where all of the exogenous U_z, U_{xz}, U_x, U_y are independent. Based on this SCM, the potential outcome Y_x can be written as $f_Y(x, Z, U_y)$. This further allows us to say that

$$f_Y(x, z, U_y) \perp\!\!\!\perp f_X(z, U_x, U_{xz}) \mid Z = z, \tag{20}$$

which is implied by the independence of exogenous variables $U_y \perp\!\!\!\perp U_x, U_{xz}$, and the fact that $Z \leftarrow f_Z(U_z, U_{xz})$ is observed (i.e., conditioned on). Eq. 20 is an equivalent representation of the ignorability statement $Y_x \perp\!\!\!\perp X \mid Z$, but expressed at the level of the structural causal model.

This offers an alternative perspective: the language of structural causality provides a basis for assessing ignorability statements, where the necessary assumptions are elicited through the notion of a cluster diagram over three nodes, CG(3). Motivated by these observations, we can finally define the notion of structural ignorability:

Definition 12 (Structural Ignorability). *Consider an SCM M over observables Z, X, and Y, and assume that no element of Z is a descendant of X or Y, and that X is not a descendant of Y. Let U denote the set of exogenous variables. We say that structural ignorability (Str-Ign, for short) holds if the following conditions hold:*

(i) There exists no $U_{xy} \in U$ such that U_{xy} is an argument of both the f_X and f_Y mechanisms.
(ii) There exists no pair $U_1, U_2 \in U$ such that U_1 is an argument of both f_X and f_{Z_i} for some $Z_i \in Z$, and U_2 is an argument of both f_Y and f_{Z_j} for some $Z_j \in Z$. □

This is a fundamental notion as it provides proper semantics for assessing ignorability-type statements in the language of structural causality. [5]

Since the SCM M is not generally observable, structural ignorability is not directly testable. In practice, however, this definition can be assessed empirically by graphical means — specifically, by evaluating the back-door criterion in the corresponding CG(3), as discussed next.

Proposition 4 (Back-door in CG(3) $\implies$ Structural Ignorability). *If the set Z satisfies the back-door criterion relative to an ordered pair (X, Y) in a CG(3) diagram, then (X, Y) satisfies the structural ignorability criterion conditional on Z. Specifically, if the following conditions hold:*

(a) There is no bidirected edge $X \leftarrow\!\!-\!\!-\!\!\rightarrow Y$ in CG(3), and
(b) There is either no bidirected edge $Z \leftarrow\!\!-\!\!-\!\!\rightarrow X$ or no bidirected edge $Z \leftarrow\!\!-\!\!-\!\!\rightarrow Y$ in CG(3),

then structural ignorability holds in the underlying structural causal model. □

[5] For simplicity, we did not explicitly list the case in which correlation among exogenous variables in U may exist even when no common cause relates them. For instance, the first condition could be replaced with the following alternative (similar to (ii)):

(i') There exists no pair $U_x, U_y \in U$ such that U_x is an argument of f_X, U_y is an argument of f_Y, and U_x and U_y are not independent.

Therefore, the $\mathrm{CG}(3)$ class of models can be used to infer whether structural ignorability holds based on the back-door criterion. In this way, we obtain a straightforward operational method for assessing structural ignorability. The process of evaluating whether structural ignorability holds then corresponds to the route R3 highlighted in Fig. 1. This proposal leverages the $\mathrm{CG}(3)$ representation of the causal diagram instead of the full causal diagram itself, since the former may be easier to obtain in practice.

3.1.1 Relationship to Coarsening of Conditional Ignorability Finally, a crucial connection can be drawn between the notions of structural ignorability and conditional ignorability, as shown below (all proofs are provided in Supplement A).

PROPOSITION 5 (Str-Ign $\implies$ C-Ign). *Let M be an SCM over observables Z, X, and Y, and assume that no element of Z is a descendant of X or Y, and that X is not a descendant of Y. If structural ignorability holds in M, written Str-Ign$(M) = 1$, then conditional ignorability also holds; that is,*

$$\text{Str-Ign}(M) = 1 \implies Y_x \perp\!\!\!\perp X \mid Z. \tag{21}$$

□

The reverse implication does not hold, however. That is, structural ignorability is a more restrictive notion than conditional ignorability. To illustrate this, we recall the discussion of cases (c) and (d) from Fig. 5. In particular, both of these cases map to the same $\mathrm{CG}(3)$ model, which contains bidirected arcs $X \leftarrow\!\!-\!\!-\!\!\rightarrow Z$ and $Y \leftarrow\!\!-\!\!-\!\!\rightarrow Z$, and corresponds to the shaded blue area in Fig. 4(b). For this $\mathrm{CG}(3)$ model, the set Z does not satisfy the back-door criterion for (X, Y), and hence structural ignorability is not implied, according to Prop. 4.

As discussed earlier in Sec. 3, however, conditional ignorability does hold in case (c), but not in case (d). This is visually depicted in Fig. 4(b), where the ignorability boundary cuts through the blue region corresponding to the $\mathrm{CG}(3)$ model in Fig. 8(b), placing the causal diagrams (c) and (d) onto opposite sides of the boundary. This highlights that, in general, assessing conditional ignorability requires a highly granular understanding of the underlying system. In the particular example, to evaluate whether ignorability holds, one must know whether variables Z_1 and Z_2 are confounded or in some other causal relation – something that may be challenging to determine in practice.

Following these observations — in particular, the fact that the conditional ignorability boundary cuts through a single $\mathrm{CG}(3)$ model (Fig. 4(b)) — it becomes clear that no function defined solely over $\mathrm{CG}(3)$ diagrams can reliably determine whether conditional ignorability holds. We state this result formally below:

PROPOSITION 6 (No Function from $\mathrm{CG}(3)$ to Conditional Ignorability). *Let Ω be the space of structural causal models over $n \geq 4$ endogenous variables $V = \{Z_1, \dots, Z_{n-2}, X, Y\}$, with X and Y labeling the last two variables in topological order. Let $\mathcal{A} : \Omega \to \{0, 1\}$ be the adjustment validity operator, indicating whether the conditional ignorability statement $Y_x \perp\!\!\!\perp X \mid Z$ holds in the SCM $M \in \Omega$. Let $\gamma : \Omega \to \mathrm{CG}(3)$ be the mapping from an SCM M to its corresponding $\mathrm{CG}(3)$ cluster diagram. Then, there exists no function $\mathcal{A}^{\mathrm{coarse}}$ such that*

$$\mathcal{A}^{\mathrm{coarse}}(\gamma(M)) = \mathcal{A}(M) \quad \textit{for all } M \in \Omega. \tag{22}$$

□

In words, for any function defined over $\mathrm{CG}(3)$ diagrams, there exists a model M such that, once its details are abstracted away (i.e., when only the $\mathrm{CG}(3)$ representation $\gamma(M)$ is used), it becomes impossible to recover the correct conditional ignorability assessment $\mathcal{A}(M)$ using any function $\mathcal{A}^{\mathrm{coarse}}$. This means that, after abstraction, no such function can correctly assess conditional ignorability. Therefore, the notion of conditional ignorability is not compatible with the level of abstraction at which one can realistically operate, such as the $\mathrm{CG}(3)$ representation. In contrast, the notion of structural ignorability is explicitly constructed to be compatible with the level of abstraction used in the $\mathrm{CG}(3)$ representation, and it can be viewed as a natural construct in light of the following result:

PROPOSITION 7. *Let Ω be the space of SCMs, μ a strictly positive measure over Ω, $\mathcal{A} : \Omega \to \{0, 1\}$ the conditional ignorability indicator, $\gamma : \Omega \to \mathrm{CG}(3)$ the mapping from an SCM to its corresponding $\mathrm{CG}(3)$ cluster diagram, and $\underline{\mathcal{A}} : \Omega \to \{0, 1\}$ a function. If $\underline{\mathcal{A}}$ satisfies:*

(i) $\underline{\mathcal{A}}(M) \leq \mathcal{A}(M)$ for all M,
(ii) $\int_{\Omega} |\underline{\mathcal{A}} - \mathcal{A}| \, d\mu$ is minimal,
(iii) There exists a function $\mathcal{A}^{\mathrm{coarse}} : \mathrm{CG}(3) \to \{0, 1\}$ such that $\mathcal{A}^{\mathrm{coarse}}(\gamma(M)) = \underline{\mathcal{A}}(M)$ for all M,

then $\underline{\mathcal{A}}$ corresponds to structural ignorability. □

Requirement (i) stipulates that the function $\underline{\mathcal{A}}$ is *conservative*, meaning that it is always smaller than $\mathcal{A}$. In other words, $\underline{\mathcal{A}}$ never provides a positive answer for ignorability's assessment if conditional ignorability does not hold (that is, $\underline{\mathcal{A}}$ gives no false positives with respect to $\mathcal{A}$). The requirement (ii) stipulates that $\underline{\mathcal{A}}$ needs to be *tight*, that is, as close as possible to $\mathcal{A}$, with the equality $\underline{\mathcal{A}}(M) = \mathcal{A}(M)$ holding whenever possible. Finally, requirement (iii) stipulates that $\underline{\mathcal{A}}$ must be $\mathrm{CG}(3)$-preserving, so that a function $\mathcal{A}^{\mathrm{coarse}}$ exists, which can be applied to the $\mathrm{CG}(3)$ projection $\gamma(M)$ of an SCM M in order to obtain the value of $\underline{\mathcal{A}}(M)$. This final condition requires that the evaluation of $\underline{A}$ can be obtained by operating at the $\mathrm{CG}(3)$ level of abstraction – which we argued is the level of knowledge at which data analysts can realistically operate. The above three conditions lead to structural ignorability, showing that structural ignorability is a rather natural notion, corresponding to the maximal lower bound (envelope) of conditional ignorability that can be evaluated based on the $\mathrm{CG}(3)$ projection of M.

As mentioned, structural ignorability intentionally abstracts away the internal structure of Z, making no claims about causal relationships within it. In our running examples, this means that cases (c) and (d) in Fig. 5 become indistinguishable under the $\mathrm{CG}(3)$ representation, falling into the same equivalence class. This is not a shortcoming but a deliberate feature: the method's scope is explicit and transparent, grounded in the language of structural causality.

Structural ignorability combines advantages from both the graphical and potential outcomes traditions:

(I) It substantially reduces the modeling burden compared to constructing a full causal diagram;

(II) The $\mathrm{CG}(3)$ template guides assumption elicitation through three clear questions (existence of bidirected edges $X \dashleftarrow\dashrightarrow Y$, $X \dashleftarrow\dashrightarrow Z$, $Y \dashleftarrow\dashrightarrow Z$), aiding transparency; and

(III) It delineates more realistic cases where adjustment is justified, avoiding hidden commitments that require overly detailed system knowledge.

The first point improves on classical graphical methods, while the latter two address limitations in the traditional potential outcomes perspective. Moreover, the broader lesson – that assessing counterfactual independence without a graphical model is limited – extends beyond ignorability to settings such as instrumental variables (see Supplement B). Finally, the $\mathrm{CG}(3)$ approach is a special case of the more general framework for clustered causal diagrams introduced in [1] and further discussed in [3, Sec. 5.7.2].

4. CONCLUSIONS

The potential outcomes (PO) framework, widely used in empirical disciplines, centers its identification strategy on conditional ignorability assumptions of the form $Y_x \perp\!\!\!\perp X \mid Z$. This formulation enables causal reasoning without explicit modeling of the underlying causal mechanisms. In practice, such statements are often interpreted as depending on broad categories of variables —- treatment X, outcome Y, and covariates Z — suggesting a kind of reasoning in blocks (i.e., grouping variables into modular units without specifying internal causal structure). While this abstraction is appealing, we show that it can be misleading: the validity of ignorability may hinge on fine-grained structural relationships within the covariates Z that are typically left unspecified. (Indeed, assessing ignorability may implicitly require reasoning over an exponential number of structural configurations – on the order of 4^{n^2} – which are rarely made explicit by analysts.) Our examples (e.g., Fig. 5) demonstrate that ignoring internal structure, even when the variable blocks are fixed, can lead to systematically incorrect causal conclusions.

Graphical models offer a contrasting perspective. Rooted in traditions from genetics, econometrics, and, more recently, computer science, this approach begins with the notion of mechanisms — formalized as structural causal models (SCMs) — from which one can derive a directed acyclic graph (DAG) and apply graphical criteria (e.g., the back-door criterion) to systematically evaluate identification strategies. In theory, this framework provides transparency and rigor. In practice, however, it requires the data scientist to specify a full causal diagram, including all direct causal and confounding relationships among observed variables. This task can be nontrivial, especially in high-dimensional settings or when domain knowledge is limited. Even though the DAG is already a coarsening of the underlying SCM, it often remains too fine-grained to be reliably elicited.

To address these limitations, we propose a new approach: instead of partitioning the space of SCMs Ω at the level of individual variables, according to a traditional DAG construction, we consider a coarser partition based on clusters of variables. Specifically, we introduce the class of cluster causal diagrams ($\mathrm{CG}(3)$), which group the covariates Z and focus attention on the high-level dependencies among X, Y, and Z. This leads to a new semantic notion — structural ignorability — which offers a structural counterpart to conditional ignorability. Although defined over the full SCM space Ω, structural ignorability can be assessed using back-door–style reasoning applied to the coarser $\mathrm{CG}(3)$ diagrams. By abstracting away internal structure within Z, this framework aligns more closely with the level of knowledge typically available to practitioners, while retaining a clear causal interpretation grounded in structural semantics.

Importantly, we show that conditional ignorability itself cannot be reliably assessed at the level of $\mathrm{CG}(3)$ abstraction. That is, no function defined over $\mathrm{CG}(3)$ diagrams can reproduce the judgments required for conditional ignorability across the SCM space. This impossibility result (Prop. 6) highlights the mismatch between the assumptions of the PO framework and the abstraction level at which analysts often operate. Structural ignorability, by contrast, is constructed to be compatible with this abstraction — it deliberately relaxes internal commitments while retaining testability via graphical tools. This result underscores the importance of aligning the level of abstraction in modeling with the level at which assumptions can be meaningfully articulated and judged.

In this way, structural ignorability strikes a balance between the PO and graphical frameworks: it avoids the hidden commitments of block-based conditional ignorability while requiring fewer assumptions than full DAG-based modeling. Crucially, it emphasizes that every ignorability-type assumption must rest on explicit structural semantics, grounded in assumptions about the underlying data-generating mechanisms. This principle realizes Haavelmo's foundational insight that "the model chosen must reflect, in a simplified form, the actual mechanism of the system investigated" [12]. Structural ignorability operationalizes this view by making such assumptions transparent and testable, thereby linking counterfactual independence with the logic of structural causality.[6] We believe the $\mathrm{CG}(3)$ approach offers a principled, operationally viable framework at a coarser modeling level — particularly appealing to practitioners who seek practical guidance without sacrificing formal guarantees.

REFERENCES

[1] Tara V Anand, Adèle H Ribeiro, Jin Tian, and Elias Bareinboim. Effect identification in cluster causal diagrams. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, 2023.

[2] Heejung Bang and James M Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 2005.

[3] Elias Bareinboim. *Causal Artificial Intelligence: A Roadmap for Building Causally Intelligent Systems*. 2025. Draft version, Apr 2025.

[4] Elias Bareinboim, Juan D. Correa, Duligur Ibeling, and Thomas Icard. On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, page 507–556. Association for Computing Machinery, New York, NY, USA, 1st edition, 2022.

[5] Elias Bareinboim and Judea Pearl. Causal inference by surrogate experiments: z-identifiability. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, UAI'12, page 113–120, Arlington, Virginia, USA, 2012. AUAI Press.

[6] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

[7] Nancy Cartwright. *Nature's Capacities and Their Measurement*. Oxford University Press, 1989.

[8] Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters, 2018.

[9] Lawrence J. Christiano, Martin Eichenbaum, and Sergio Rebelo. When is the government spending multiplier large? *Journal of Political Economy*, 119(1):78–121, 2009.

[10] Juan D. Correa and Elias Bareinboim. Counterfactual graphical models: Constraints and inference. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, 2025.

[11] Ronald Aylmer Fisher. Design of experiments. *British Medical Journal*, 1(3923):554, 1936.

[12] Trygve Haavelmo. The probability approach in econometrics. *Econometrica*, 12(Supplement):iii–115, 1944. Reprinted in *Econometrica* 12(Supplement).

[13] Yimin Huang and Marco Valtorta. Pearl's calculus of intervention is complete. *arXiv preprint arXiv:1206.6831*, 2012.

[14] Guido W Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, 86(1):4–29, 2004.

[15] Guido W Imbens. Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature*, 58(4):1129–1179, 2020.

[16] Marshall M Joffe, Wei Peter Yang, and Harold I Feldman. Selective ignorability assumptions in causal inference. *The International Journal of Biostatistics*, 6(2), 2010.

[17] Edward H Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2):3008–3049, 2023.

[18] Sanghack Lee, Juan D Correa, and Elias Bareinboim. General identifiability with arbitrary surrogate experiments. In *Uncertainty in artificial intelligence*, pages 389–398. PMLR, 2020.

[19] Robert E. Lucas. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46, 1976.

[20] Jacob Marschak. Economic measurements for policy and prediction. In William C. Hood and Tjalling C. Koopmans, editors, *Studies in Econometric Method*, pages 1–26. Wiley, New York, 1953.

[21] Jerzy Neyman. On the application of probability theory to agricultural experiments. essay on principles. *Ann. Agricultural Sciences*, pages 1–51, 1923.

[22] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[23] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

[24] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000. 2nd edition, 2009.

[25] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic books, 2018.

[26] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

[27] James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the american statistical association*, 90(429):106–121, 1995.

[28] Paul R Rosenbaum, P Rosenbaum, and Briskman. *Design of observational studies*, volume 10. Springer, 2010.

[29] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[30] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[31] Donald B Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58, 1978.

[32] Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[33] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models.

[6] See also Marschak [20]: "It is only through specifying the structure of the system—the mechanism that generates the data — that identification becomes possible."

In *Proceedings of the National Conference on Artificial Intelligence*, volume 21/2, page 1219. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[34] Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Aaai/iaai*, pages 567–573, 2002.

[35] Sewall Wright. Systems of mating. i. the biometric relations between parent and offspring. *Genetics*, 6(2):111, 1921.

APPENDIX A: PROOFS

PROOF OF PROP. 4. Suppose that for an SCM M structural ignorability does not hold, which can happen for two reasons. Firstly, it may be due to the existence of a latent U_{xy} affecting both X, Y. In this case, there exists a bidirected edge $X \leftarrow\!\!\dashrightarrow Y$, and therefore Z is not back-door for (X, Y). Secondly, it may be that case that there exists a pair U_1, U_2 such that U_1 is an argument of both f_X, f_Z, while U_2 is an argument of f_Y, f_Z. In this case, the existence of such U_1 implies the bidirected edge $X \leftarrow\!\!\dashrightarrow Z$, while U_2 implies the bidirected edge $Y \leftarrow\!\!\dashrightarrow Z$. Therefore, we again conclude Z is not back-door for (X, Y), since conditioning on Z would leave the path $X \leftarrow\!\!\dashrightarrow Z \leftarrow\!\!\dashrightarrow Y$ open. Thus, we can conclude that the back-door criterion in CG(3) implies structural ignorability. □

PROOF OF PROP. 6. To prove that no function $\mathcal{A}^{\text{coarse}}$ exists, consider the following construction of two SCMs M_1, M_2 over $n \geq 4$ variables $\{Z_1, \dots, Z_{n-2}, X, Y\}$. For both SCMs, each Z_i for $1 \leq i \leq n-2$ is an input of f_X, f_Y. In the SCM M_1, there is a shared noise variable $U_{1,X}$ that is an input to f_{Z_1}, f_X; there is also a noise variable $U_{n-2,Y}$ that is an input to $f_{Z_{n-2}}, f_Y$. In M_1, no other exogenous U_i is an input to more than one causal mechanism. In M_2, however, there are additional noise variables $U_{i,i+1}$ that are inputs to $f_{Z_i}, f_{Z_{i+1}}$, for each $1 \leq i \leq n-3$. First, note that both M_1, M_2 correspond to the same CG(3) model, namely in Fig. 8(b). The bidirected edge $X \leftarrow\!\!\dashrightarrow Z$ exists in both cases due to the existence of the noise variable $U_{1,X}$; the edge $Y \leftarrow\!\!\dashrightarrow Z$ exists due to $U_{n-2,Y}$. Crucially, however, conditional ignorability holds in M_1, since conditioning on Z blocks all the back-door paths. In M_2, however, conditioning on Z opens the path $X \leftarrow\!\!\dashrightarrow Z_1 \leftarrow\!\!\dashrightarrow \dots \leftarrow\!\!\dashrightarrow Z_{n-2} \leftarrow\!\!\dashrightarrow Y$, and therefore conditional ignorability does not hold. Therefore, no mapping $\mathcal{A}^{\text{coarse}}$ satisfying the required properties can exist since $\gamma(M_1) = \gamma(M_2)$ but $\mathcal{A}(M_1) \neq \mathcal{A}(M_2)$. □

PROOF OF PROP. 7. Denote by $\gamma_1, \dots, \gamma_8$ the eight CG(3) diagrams in Figs. 8(a)-8(h). Suppose that M_1, M_2 are such that $\gamma(M_1) = \gamma(M_2) = \gamma_i$. Then, the existence of $\mathcal{A}^{\text{coarse}}$ such that $\mathcal{A}^{\text{coarse}}(\gamma(M)) = \underline{\mathcal{A}}(M)$ implies that $\underline{\mathcal{A}}(M_1) = \underline{\mathcal{A}}(M_2)$. Therefore, the function $\underline{\mathcal{A}}$ is constant along each set $\gamma^{-1}(\gamma_i)$, where γ^{-1} denotes the preimage of γ.

For each $\gamma_1, \dots, \gamma_8$, we now consider two cases. In the first case, suppose that for γ_i there exists M such that $\gamma(M) = \gamma_i$ and $\mathcal{A}(M) = 0$ (meaning that conditional ignorability does not hold for the SCM M). Due to the requirement $\underline{\mathcal{A}}(M) \leq \mathcal{A}(M)\ \forall M$, it follows that $\underline{\mathcal{A}}(M) = 0$. Using the fact that $\underline{\mathcal{A}}$ must be constant along $\gamma^{-1}(\gamma_i)$, it follows that $\underline{\mathcal{A}}(M) = 0\ \forall M \in \gamma^{-1}(\gamma_i)$. In the second case, suppose that for γ_i for each M such that $\gamma(M) = \gamma_i$ we have $\mathcal{A}(M) = 1$ (meaning that conditional ignorability holds for each SCM M). Then, by the condition that $\int_\Omega |\underline{\mathcal{A}} - \mathcal{A}| d\mu$ is minimal, we can see that we must have $\underline{\mathcal{A}}(M) = 1\ \forall M \in \gamma^{-1}(\gamma_i)$, since otherwise $\int_\Omega |\underline{\mathcal{A}} - \mathcal{A}| d\mu$ can be made smaller.

Therefore, $\underline{\mathcal{A}}$ evaluates to 0 along $\gamma^{-1}(\gamma_i)$ whenever the set has an element with $\mathcal{A}(M) = 0$; it evaluates to 1 whenever each element of it satisfies $\mathcal{A}(M) = 1$. It remains to verify which values $\mathcal{A}$ takes on each of the sets $\gamma^{-1}(\gamma_1), \dots, \gamma^{-1}(\gamma_8)$. For $\gamma^{-1}(\gamma_1), \dots, \gamma^{-1}(\gamma_5)$ with corresponding CG(3) models in Figs. 8(a)-8(e), it is not difficult to construct a model M such that $\gamma(M) = \gamma_i$ and $\mathcal{A}(M) = 0$ (we omit the full details of these constructions, since the key ideas appear in the main text and the proof of Prop. 6). Therefore, we conclude that $\underline{\mathcal{A}} = 0$ on $\gamma^{-1}(\gamma_1), \dots, \gamma^{-1}(\gamma_5)$. For $\gamma^{-1}(\gamma_6), \gamma^{-1}(\gamma_7), \gamma^{-1}(\gamma_8)$ corresponding to CG(3) models in Figs. 8(f)-8(h), we can see that for each M in these sets, in $\gamma(M)$ the back-door criterion is satisfied; implying that $\mathcal{A}(M) = 1$ for any such M. Therefore, we conclude $\underline{\mathcal{A}} = 1$ on $\gamma^{-1}(\gamma_6), \gamma^{-1}(\gamma_7), \gamma^{-1}(\gamma_8)$. This construction determines the values of $\underline{\mathcal{A}}$ on the entire space Ω, and we can see that $\underline{\mathcal{A}}$ corresponds exactly to the notion of structural ignorability. □

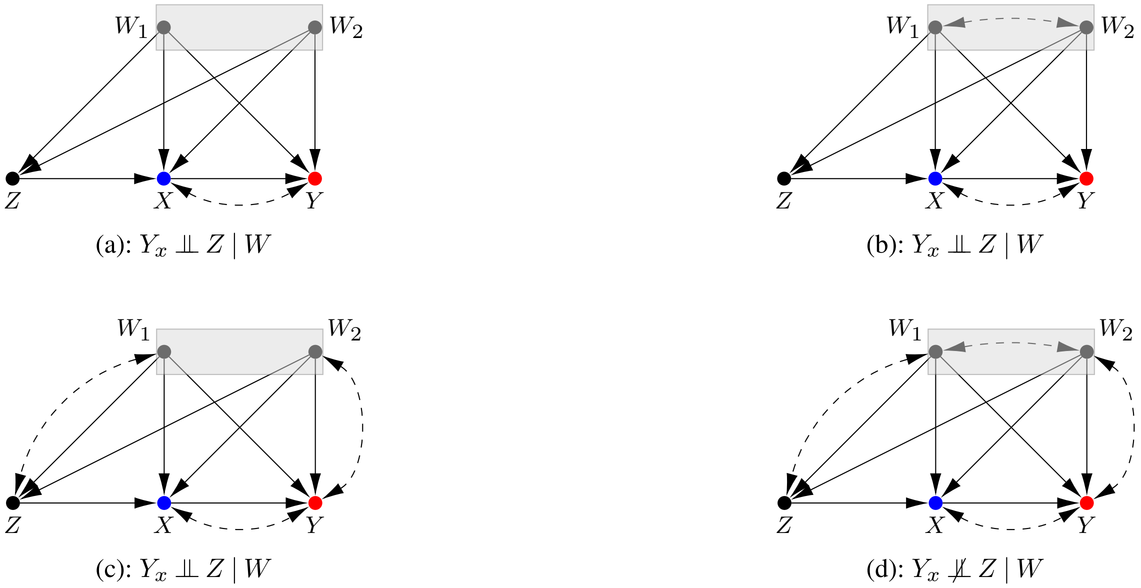
APPENDIX B: STRUCTURAL ACCOUNT OF INSTRUMENTAL VARIABLES

In this section, we extend the argument in the main text to the setting of instrumental variables. In the context of a conditional instrument, we show that evaluating the independence restriction $Y_x \perp\!\!\!\perp Z \mid W$ — where X and Y are the treatment and outcome variables, Z is the instrument, and W is a set of covariates — faces essentially the same challenges as assessing confounding in the classic setting discussed in the main body, as illustrated next.

EXAMPLE 5. *Consider the models in Fig. 9(a–d), where X is the treatment, Y the outcome, Z the instrument, and W_1, W_2 observed confounders. The goal is to assess whether the independence constraint $Y_x \perp\!\!\!\perp Z \mid W$ holds in each scenario. In this case, a cluster causal diagram over four blocks (denoted* CG(4)*) may be considered, with*

$$B_1 = \{Y\}, \quad B_2 = \{X\}, \quad B_3 = \{Z\}, \quad B_4 = \{W_1, W_2\}. \tag{23}$$

We analyze the pairwise relations between blocks B_1-B_4 for each model to determine the corresponding cluster diagram CG(4)*. In all models, the edges $X \to Y$, $X \leftarrow\!\!\dashrightarrow Y$, $W \to X$, $W \to Y$, $W \to Z$, and $Z \to X$ are present. For models (a) and (b), the edges $Z \to Y$, $Z \leftarrow\!\!\dashrightarrow W$, and $Y \leftarrow\!\!\dashrightarrow W$ are absent. Therefore, both models (a) and (b) map to the*



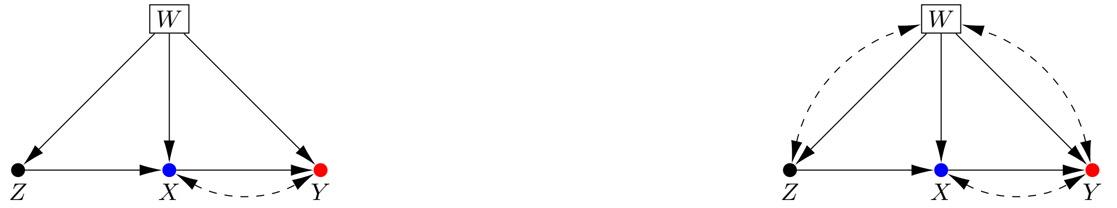
(a): $Y_x \perp\!\!\!\perp Z \mid W$

(b): $Y_x \perp\!\!\!\perp Z \mid W$

(c): $Y_x \perp\!\!\!\perp Z \mid W$

(d): $Y_x \not\perp\!\!\!\perp Z \mid W$

Figure 9: Causal diagrams (a)–(d).



(a) Model (a)

(b) Model (b)

Figure 10: Cluster diagrams discussed in the example, where (a) represents causal diagrams in Fig. 9(a-b), and (b) the ones in Fig. 9(c-d).

same $\mathrm{CG}(4)$ *model shown in Fig. 10(a), where* W *is a cluster consisting of* W_1, W_2*. This occurs because all inter-block relationships are identical for the two models.*

In both models (a) and (b), the independence constraint holds. This follows from the back-door criterion applied to the mutilated graph $\mathcal{G}_{\underline{X}}$*, obtained by removing all outgoing edges from* X*. The result is invariant to the internal relationship between* W_1 *and* W_2 *since they are treated as a single block* B_4 *in* $\mathrm{CG}(4)$*. Graphically, the shaded area over* W_1 *and* W_2 *indicates that the modeler remains agnostic to their mutual dependence.*

Consider now the models in Fig. 9(c–d). For these models, moving to the $\mathrm{CG}(4)$ *representation requires adding the edges* $Z \dashleftarrow\!\!\dashrightarrow W$ *(due to* $Z \dashleftarrow\!\!\dashrightarrow W_1$*) and* $Y \dashleftarrow\!\!\dashrightarrow W$ *(due to* $Y \dashleftarrow\!\!\dashrightarrow W_2$*). Therefore, both models (c) and (d) correspond to the same* $\mathrm{CG}(4)$ *model shown in Fig. 10(b). The relations within the shaded area remain unchanged: in models (a) and (c),* W_1 *and* W_2 *are independent, whereas in models (b) and (d), they share an unobserved common cause (dashed arrow* $W_1 \dashleftarrow\!\!\dashrightarrow W_2$*).*

When analyzing cases (c) and (d), one might be tempted to abstract away the structure between W_1 *and* W_2 *(within block* B_4*), as was done successfully in cases (a) and (b). However, this abstraction fails here. In model (c), conditioning on* W *blocks all back-door paths between* Z *and* Y*, whereas in model (d), conditioning on* W *leaves the path* $Z \dashleftarrow\!\!\dashrightarrow W_1 \dashleftarrow\!\!\dashrightarrow W_2 \dashleftarrow\!\!\dashrightarrow Y$ *open. Consequently, the independence* $Y_x \perp\!\!\!\perp Z \mid W$ *holds in model (c) but not in model (d). The critical observation is that the inter-block relationships are identical in models (c) and (d), yet the independence constraint holds only in (c). The difference lies entirely within block* B_4*, which the statement* $Y_x \perp\!\!\!\perp Z \mid W$ *fails to capture. As expected, the corresponding* $\mathrm{CG}(4)$ *in Fig. 10 properly reflects this subtlety.* □

As the above example illustrates, the key observation in the main text, namely, that ignorability cannot account for fine-grained structure, is not limited to adjustment settings, but applies more broadly to the assessment of counterfactual relationships expressed as independencies. Once again, our proposal is to use a template model over four nodes (labeled $\mathrm{CG}(4)$) and assess the following:

(i) whether a directed edge $Z \rightarrow Y$ exists,
(ii) whether a bidirected edge $Z \dashleftarrow\!\!\dashrightarrow Y$ exists,
(iii) whether a bidirected edge $Z \dashleftarrow\!\!\dashrightarrow W$ exists,
(iv) whether a bidirected edge $Y \dashleftarrow\!\!\dashrightarrow W$ exists.

Eliciting assumptions through these steps provides a natural, transparent way to encode causal knowledge, and a simple application of the back-door criterion then determines whether the independence constraint holds, based on the cluster diagram $\mathrm{CG}(4)$.

APPENDIX C: ON THE NON-LOCAL NATURE OF IGNORABILITY STATEMENTS

There is a common belief that the lack of unobserved confounding between variables X and Y implies the corresponding ignorability statement (such as in Eq. 10), but this perspective is inaccurate. This is so since ignorability is a global statement between the treatment and the outcome variables, while confounding is a more local and fine-grained property of the causal system. To illustrate this point, we revisit the previous examples through structural lenses.

EXAMPLE 6. *We consider again the causal diagrams in Fig. 5(a-d) and the following SCMs $\mathcal{M}^{(i)}$ with structural mechanisms $\mathcal{F}$:*

$$\mathcal{F}: \begin{cases} Z_1 \leftarrow f_{Z_1}(U_{Z_1}) & (24)\\ Z_2 \leftarrow f_{Z_2}(U_{Z_2}) & (25)\\ X \leftarrow f_X(Z_1, Z_2, U_X) & (26)\\ Y \leftarrow f_Y(X, Z_1, Z_2, U_Y), & (27) \end{cases}$$

where the distributions over the exogenous variables $P^{(i)}(U_{Z_1}, U_{Z_2}, U_X, U_Y)$ factorize in each model as:

$$\mathcal{M}^{(1)} : P^{(1)}(U_{Z_1})P^{(1)}(U_{Z_2})P^{(1)}(U_X)P^{(1)}(U_Y), \tag{28}$$

$$\mathcal{M}^{(2)} : P^{(2)}(U_{Z_1}, U_{Z_2})P^{(2)}(U_X)P^{(2)}(U_Y), \tag{29}$$

$$\mathcal{M}^{(3)} : P^{(3)}(U_{Z_1})P^{(3)}(U_{Z_2})P^{(3)}(U_X|U_{Z_1})P^{(3)}(U_Y|U_{Z_2}), \tag{30}$$

$$\mathcal{M}^{(4)} : P^{(4)}(U_{Z_1}, U_{Z_2})P^{(4)}(U_X|U_{Z_1})P^{(4)}(U_Y|U_{Z_2}). \tag{31}$$

In the model $\mathcal{M}^{(1)}$ corresponding to Fig. 5(a), it follows through Eq. 28 that all unobserved factors are marginally independent. In particular, since U_X and U_Y are marginally independent in this case, this means that there is no unobserved confounding between X and Y. The same holds true with Z_1 and Z_2, following from the independence of their exogenous variables U_{Z_1} and U_{Z_2}.

To further understand the implications of such independences, we consider the factual and counterfactual variables involved in the independence relation that is being evaluated in the ignorability case, namely:

$$X \leftarrow f_X(Z_1, Z_2, U_X) \tag{32}$$

$$Y_x \leftarrow f_Y(x, Z_1, Z_2, U_Y) \tag{33}$$

Now we re-write these expressions replacing the confounders with the corresponding exogenous variables (i.e., substituting Eqs. 24-25 into Eqs. 32-33) which leads to:

$$X \leftarrow f_X(f_{Z_1}(U_{Z_1}), f_{Z_2}(U_{Z_2}), U_X) \tag{34}$$

$$Y_x \leftarrow f_Y(x, f_{Z_1}(U_{Z_1}), f_{Z_2}(U_{Z_2}), U_Y). \tag{35}$$

Note that the source of randomness for the factual X and the counterfactual Y_x comes from the distribution of exogenous variables (U_{Z_1}, U_{Z_2}, U_X) and (U_{Z_1}, U_{Z_2}, U_Y), respectively. It is then immediate to see that despite the fact that $U_{Z_1} \perp\!\!\!\perp U_{Z_2}$ and $U_X \perp\!\!\!\perp U_Y$ hold in this case, the variables X and Y_x share certain exogenous variations, namely, U_{Z_1}, U_{Z_2}, which means that a priori they are not independent.

Interestingly, there is no unobserved confounding between X and Y, but still, they share sources of exogenous variations through Z_1, Z_2. We note through Eq. 34 that what precludes ignorability, $Y_x \perp\!\!\!\perp X$, is not the relationship between U_X and U_Y in this case, since these are independent, but the variations coming from Z_1, Z_2. Therefore, to control for these confounding variations, we need to condition on variables Z_1, Z_2, which could imply that $Y_x \perp\!\!\!\perp X \mid Z = z$. To show this is the case, we will show that

$$P(X = x', Y_x = y \mid Z = z) = P(X = x' \mid Z = z)P(Y_x = y \mid Z = z). \tag{36}$$

In doing so, we wish to leverage the independences implied by the factorization of the distribution over the exogenous variables, $P(u)$. In particular, model $\mathcal{M}^{(1)}$ implies:

$$P(u_x, u_y, u_z) = P(u_x)P(u_y, u_z), \tag{37}$$

$$P(u_x, u_z) = P(u_x)P(u_z). \tag{38}$$

Using the definition of conditional probabilities, l.h.s. of Eq. 36 can be written as

$$P(X = x', Y_x = y \mid Z = z) = \frac{P(X = x', Y_x = y, Z = z)}{P(Z = z)}. \tag{39}$$

Based on Def. 7, the numerator $P(X = x', Y_x = y, Z = z)$ can be expanded as

$$\sum_u \Big(\mathbb{1}(Z(u) = z)\mathbb{1}(X(u) = x')\mathbb{1}(Y_x(u) = y)\Big)P(u) \tag{40}$$

$$= \sum_u \Big(\mathbb{1}(Z(u_z) = z)\mathbb{1}(X(z, u_x) = x')\mathbb{1}(Y_x(u_z, u_y) = y)\Big)P(u) \qquad \textit{(using the specific SCM)} \tag{41}$$

$$= \sum_u \Big(\mathbb{1}(X(z, u_x) = x')\mathbb{1}(Y_x(z, u_y) = y)\mathbb{1}(Z(u_z) = z)\Big)P(u_x)P(u_z, u_y) \qquad (\textit{Eq. } 37) \tag{42}$$

$$= \underbrace{\sum_{u_x} \mathbb{1}(X(z, u_x) = x')P(u_x)}_{\textit{Term I}} \underbrace{\sum_{u_z, u_y} \mathbb{1}(Y_x(u_z, u_y) = y)\mathbb{1}(Z(u_z) = z)P(u_z, u_y)}_{\textit{Term II}} \quad \textit{(factorizing)}. \tag{43}$$

Using Def. 7 we can recognize Term II as $P(Y_x = y, Z = z)$. We next multiply Term I with $P(Z = z)$ and obtain

$$\textit{Term I} * P(Z = z) = \sum_{u_x} \mathbb{1}(X(z, u_x) = x')P(u_x) \sum_{u_z'} \mathbb{1}(Z(u_z') = z)P(u_z') \tag{44}$$

$$= \sum_{u_x, u_z'} \Big(\mathbb{1}(X(z, u_x) = x')\mathbb{1}(Z(u_z') = z)\Big)P(u_x)P(u_z') \tag{45}$$

$$= \sum_{u_x, u_z'} \Big(\mathbb{1}(X(z, u_x) = x')\mathbb{1}(Z(u_z') = z)\Big)P(u_x, u_z') \qquad (\textit{Eq. } 38) \tag{46}$$

$$= \sum_{u_x, u_z'} \Big(\mathbb{1}(X(u_z', u_x) = x')\mathbb{1}(Z(u_z') = z)\Big)P(u_x, u_z') \quad \textit{(using the specific SCM)} \tag{47}$$

$$= P(X = x', Z = z). \tag{48}$$

Thus, Term I equals $P(X = x' \mid Z = z)$, and plugging it back into Eq. 39 yields

$$P(X = x', Y_x = y \mid Z = z) = \frac{P(Y_x = y, Z = z)}{P(Z = z)} P(X = x' \mid Z = z) \tag{49}$$

$$= P(X = x' \mid Z = z)P(Y_x = y \mid Z = z), \tag{50}$$

which shows that ignorability holds. Interestingly, the independences in Eqs. 37-38 used above also hold for the model $\mathcal{M}^{(2)}$ corresponding to Fig. 5(b), and therefore the same proof can be used for verifying ignorability in the model $\mathcal{M}^{(2)}$.

Now, for the models $\mathcal{M}^{(3)}$, $\mathcal{M}^{(4)}$ in Fig. 5(c), (d), the independence statements in Eqs. 37-38 do not hold. However, for the model $\mathcal{M}^{(3)}$, we know that

$$P(u_{z_1}, u_{z_2}, u_x, u_y) = P(u_{z_1}, u_x)P(u_{z_2}, u_y), \tag{51}$$

$$P(u_{z_1}, u_{z_2}) = P(u_{z_1})P(u_{z_2}). \tag{52}$$

Using these independences, a slightly different approach is possible:

$$\sum_u P(u)\Big(\mathbb{1}(X(u)=x')\mathbb{1}(Y_x(u)=y)\mathbb{1}(Z(u)=z)\Big) \tag{53}$$

$$= \sum_u P(u)\Big(\mathbb{1}(X(z,u_x)=x')\mathbb{1}(Y_x(u_z,u_y)=y)\mathbb{1}(Z_1(u_{z_1})=z_1)\mathbb{1}(Z_2(u_{z_2})=z_2)\Big) \tag{54}$$

(using the specific SCM)

$$= \sum_u P(u_x,u_{z_1})P(u_{z_2},u_y)\Big(\mathbb{1}(X(z,u_x)=x')\mathbb{1}(Y_x(z,u_y)=y)\mathbb{1}(Z_1(u_{z_1})=z_1)\mathbb{1}(Z_2(u_{z_2})=z_2)\Big) \tag{55}$$

(*Eq.* 51)

$$= \underbrace{\sum_{u_x,u_{z_1}} P(u_x,u_{z_1})\Big(\mathbb{1}(X(z,u_x)=x')\mathbb{1}(Z_1(u_{z_1})=z_1)\Big)}_{\textit{Term III}} \tag{56}$$

$$\times \underbrace{\sum_{u_{z_2},u_y} P(u_{z_2},u_y)\Big(\mathbb{1}(Y_x(u_z,u_y)=y)\mathbb{1}(Z_2(u_{z_2})=z_2)\Big)}_{\textit{Term IV}}.$$

Now, with almost the same reasoning as in Eqs. 44-48, we can show that

$$\textit{Term III} * P(z_2) = P(X=x',Z_1=z_1,Z_2=z_2) \tag{57}$$

$$\textit{Term IV} * P(z_1) = P(Y_x=y,Z_1=z_1,Z_2=z_2). \tag{58}$$

Therefore, it follows that

$$P(X=x',Y_x=y \mid Z=z) = \frac{P(X=x',Z=z)P(Y_x=y,Z=z)}{P(Z_1=z_1)P(Z_2=z_2)P(Z=z)} \tag{59}$$

$$= \frac{P(X=x',Z=z)}{P(Z=z)}\frac{P(Y_x=y,Z=z)}{P(Z=z)} \tag{60}$$

$$= P(X=x' \mid Z=z)P(Y_x=y \mid Z=z) \tag{61}$$

also using that $P(Z=z)=P(Z_1=z_1)P(Z_2=z_2)$ in $\mathcal{M}^{(3)}$. Therefore, ignorability also holds in the model $\mathcal{M}^{(3)}$, although for a different reason. Independences in Eqs. 51-52, however, do not hold for the model $\mathcal{M}^{(4)}$, and therefore the same proof does not apply, and the ignorability statement $Y_x \perp\!\!\!\perp X \mid Z=z$ does not hold. □

This example illustrates an important point, namely that the confounding structure as described by the bidirected edges is a local property, while ignorability evokes a more intricate, non-local type of judgement involving constraints related to other variables, in this case Z_1 and Z_2. This line of reasoning can be extended naturally, for more general graphs, as done formally in the following proposition:

PROPOSITION 8 (Adjustment Validity and Ignorability). *Let $\mathbb{G}^n$ be the space of Semi-Markovian causal diagrams G over n endogenous variables $V=\{Z_1,\dots,Z_{n-2},X,Y\}$, with X,Y labeling the last two variables in G. Let $\mathcal{A}:\mathbb{G}^n \mapsto \{0,1\}$ be the adjustment validity operator indicating if ignorability holds in a diagram G. Let $\mathbf{B}=(B_1,B_2,B_3)=(\{Z_1,\dots,Z_{n-2}\},X,Y)$ represent the block structure following the ignorability statement in Eq. 10. Let $\mathcal{B}(G)\in\mathbb{B}$ represent the block specification for the diagram G, which determines, for each pair of blocks B_i,B_j, (i) whether there is a directed edge between B_i,B_j in G; (ii) whether there is hidden confounding between the B_i,B_j, i.e., if there is a bidirected arrow between B_i,B_j in G. Six possible relations are considered (three possible directed arrows, and three possible confounding arrows), and thus the space of block specifications $\mathbb{B}=\{0,1\}^6$.*

Then, for any number of endogenous variables $|V|=n\geq 3$, the following statements hold:

(a) $\exists\,\mathcal{B}^{(i)}\in\mathbb{B}$ *such that*

$$\exists\, G_1,G_2 \textit{ such that} \tag{62}$$

$$\mathcal{B}(G_1) = \mathcal{B}(G_2) = \mathcal{B}^{(i)} \tag{63}$$

$$\mathcal{A}(G_1) \neq \mathcal{A}(G_2). \tag{64}$$

Therefore, there is no valid mapping from the space of block-specifications $\mathbb{B}$ *to the adjustment validity decisions* $\{0,1\}$.

(b) $\exists\, \mathcal{B}^{(i)} \in \mathbb{B}$ *such that*

$$\forall\, G_1, G_2: \quad \mathcal{B}(G_1) = \mathcal{B}(G_2) = \mathcal{B}^{(i)} \implies \mathcal{A}(G_1) = \mathcal{A}(G_2). \tag{65}$$

□

The significance of the proposition can be understood as follows. We consider all diagrams over n variables, assuming that treatment X and outcome Y are the last two variables in the topological order, preceded by a set of confounders $Z_1, \ldots, Z_{n-2}$ (see Fig. 7 as an illustration). The first assertion (a) shows that two graphs G_1, G_2 can always be found such that they have the same block specification, but disagree on the assessment of ignorability. This implies mathematically that no valid mapping from the space of block specifications to adjustment validity decisions exists, generally. In other words, when abstracting away the structure within blocks, we sometimes may be mixing instances in which ignorability is true with those where it is not. This was the case with models in Fig. 5(c) and (d), which have the same block specification but differ with respect to ignorability.

The second assertion of the proposition, in (b), shows that for some specific block specifications (labeled $\mathcal{B}^{(i)}$), it may happen that all graphs G compatible with $\mathcal{B}^{(i)}$ actually agree on the ignorability assessment. In other words, for some block specification, abstracting away the structure within block may still be sufficient for assessing ignorability, i.e., there is no loss of information in the abstraction process. Mathematically, this means there may be a locally valid mapping, that maps $\mathcal{B}^{(i)}$ to a correct adjustment validity decision in $\{0,1\}$. Importantly, though, this is not always the case. This was the case with models in Fig. 5(a) and (b), which have the same block specification and also agree on the ignorability statement.