Formalizing Statistical Causality via Modal Logic

Yusuke Kawamoto[1,2], Tetsuya Sato[3], and Kohei Suenaga[4]

[1] AIST, Tokyo, Japan
[2] PRESTO, JST, Tokyo, Japan
[3] Tokyo Institute of Technology, Tokyo, Japan
[4] Kyoto University, Kyoto, Japan

Abstract. We propose a formal language for describing and explaining statistical causality. Concretely, we define *Statistical Causality Language* (StaCL) for expressing causal effects and specifying the requirements for causal inference. StaCL incorporates modal operators for interventions to express causal properties between probability distributions in different possible worlds in a Kripke model. We formalize axioms for probability distributions, interventions, and causal predicates using StaCL formulas. These axioms are expressive enough to derive the rules of Pearl's do-calculus. Finally, we demonstrate by examples that StaCL can be used to specify and explain the correctness of statistical causal inference.

1 Introduction

Statistical causality has been gaining significant importance in a variety of research fields. In particular, in life sciences, more and more researchers have been using statistical techniques to discover *causal relationships* from experiments and observations. However, these statistical methods can easily be misused or misinterpreted. In fact, it is reported that many research articles have serious errors in the applications and interpretations of statistical methods [8, 26].

A common mistake is to misinterpret statistical *correlation* as statistical *causality*. Notably, when we analyze observational data without experimental interventions, we may overlook some requirements for causal inference and make wrong calculations, leading to incorrect conclusions about the causality.

For this reason, the scientific community has developed guidelines on many requirements for statistical analyses [36, 28]. However, since there is no formal language to describe the entire procedures and their requirements, we refer to guidelines manually and cannot formally guarantee the correctness of analyses.

To address these problems, we propose a logic-based approach to formalizing and explaining the correctness of statistical causal inference. Specifically, we introduce a formal language called *statistical causality language* (StaCL) to formally describe and check the requirements for statistical causal inference. We consider this work as the first step to building a framework for formally guaranteeing and explaining the reliability of scientific research.

Contributions. Our main contributions are as follows:

- We propose *statistical causality language* (StaCL) for formalizing and explaining statistical causality by using modal operators for interventions.
- We define a *Kripke model for statistical causality*. To formalize not only statistical correlation but also statistical causality, we introduce a *data generator* in a possible world to model a causal diagram in a Kripke model.
- We introduce the notion of *causal predicates* to express statistical causality and interpret them using a data generator instead of a valuation in a Kripke model. In contrast, *(classical) predicates* are interpreted using a valuation in a Kripke model to express only *statistical correlations*.
- We introduce a sound deductive system $\mathbf{AX^{CP}}$ for StaCL with axioms for probability distributions, interventions, and causal predicates. These axioms are expressive enough to reason about all causal effects identifiable by Pearl's *do-calculus* [29]. We show that $\mathbf{AX^{CP}}$ can reason about the correctness of causal inference methods (e.g., backdoor adjustment). Unlike prior work, $\mathbf{AX^{CP}}$ does not aim to conduct causal inference about a specific causal diagram; rather, it concerns the correctness of the inference methods for any diagram. To the best of our knowledge, ours appears to be the first modal logic that can specify and reason about the requirements for causal inference.

Related Work. Many studies on causal reasoning rely on causal diagrams [30]. Whereas they aim to reason about a specific diagram, our logic-based approach aims to specify and reason about the requirements for causal inference methods.

Logic-based approaches for formalizing causal reasoning have been proposed. To name a few, Halpern and Pearl provide logic-based definitions of actual causes where logical formulas with events formalize counterfactuals [12, 13, 11]. Probabilistic logical languages [19] are proposed to axiomatize causal reasoning with observation, intervention, and counterfactual inference. Unlike our logic, however, their framework does not aim to syntactically derive the correctness of statistical causal inference. The causal calculus [27] is used to provide a logical representation [4, 3] of Pearl [30]'s structural causal model. The counterfactual-observational language [1] can reason about interventionist counterfactuals and has an axiomatization that is complete w.r.t. a causal team semantics. A modal logic in [2] integrates causal and epistemic reasoning. While these works deal with deterministic cases only, our StaCL can reason about statistical causality in probabilistic settings.

There have been studies on incorporating probabilities into team semantics [15]. For example, team semantics is used to deal with the dependence and independence among random variables [6, 5]. A probabilistic team semantics is provided for a first-order logic that can deal with conditional independence [7]. A team semantics is also introduced for logic with exact/approximate dependence and independence atoms [14]. Unlike our StaCL, however, these works do not allow for deriving the do-calculus or the correctness of causal inference methods.

Concerning the axiomatic characterization of causality, Galles and Pearl [9] prove that the axioms of composition, effectiveness, and reversibility are sound and complete with respect to the structural causal models. They also show that the reversibility axiom can be derived from the composition axiom if the causal diagram is acyclic (i.e., has no feedback loop). Halpern [10] provides axiomatizations for more general classes of causal models with feedback and with equations that may have no solutions. In contrast, our deductive system $\mathbf{AX}^{\mathbf{CP}}$ has axioms for causal predicates and two forms of interventions that can derive the rules of Pearl's do-calculus [29], while being equipped with axioms corresponding to the composition and effectiveness axioms mentioned above only for acyclic diagrams.

For the efficient computation of causal reasoning, constraint solving is applied [17, 18, 34]. Probabilistic logic programming is used to encode and reason about a specific causal diagram [31]. These are orthogonal to the goal of our work.

Finally, a few studies propose modal logic for statistical methods. Statistical epistemic logic [20–22] specifies various properties of machine learning. Belief Hoare logic [24, 25] can reason about statistical hypothesis testing programs. However, unlike our StaCL, these cannot reason about statistical causality.

2 Illustrating Example

We first present a simple example to explain our framework.

Example 1 (Drug's efficacy). We attempt to check a drug's efficacy for a disease by observing a situation where some patients take a drug and the others do not.

Table 1 shows the recovery rates and the numbers of patients treated with/without the drug. For both males and females, *more* patients recover by taking the drug. However, for the combined population, the recovery rate with the drug (0.73) is *less* than that without it (0.80). This inconsistency is called *Simpson's paradox* [33], showing the difficulty of identifying causality from observed data.

To model this, we define three variables: a *treatment* x (1 for drug, 0 for no-drug), an *outcome* y (1 for recovery, 0 for non-recovery), and a gender z. Fig. 1a depicts their causal dependency; the arrow $x \rightarrow y$ denotes that y depends on x. The *causal effect* $p(y|do(x=c))$ of a treatment $x=c$ on an outcome y [30] is defined as the distribution of y in case y were generated from $x = c$ (Fig. 1b).

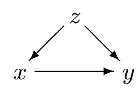
However, since the gender z influences the choice of the treatment x in reality (Fig. 1a), the causal effect $p(y|do(x=c))$ depends on the common cause z of x and y and differs from the correlation $p(y|x=c)$. Indeed, in Table 1, 80 % of females chose to take the drug ($x = 1$) while only 20 % of males did so; this dependency of x on the gender z leads to Simpson's paradox in Table 1. Thus, calculating the causal effect requires an "adjustment" for z, as explained below.

Overview of the Framework. We describe reasoning about the causal effect in Example 1 using logical formulas in our formal language StaCL (Section 5).

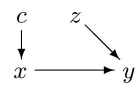
We define $\varphi_{\mathrm{RCT}} \stackrel{\mathrm{def}}{=} \lceil c/x \rceil (c_0 = y)$ to express a *randomized controlled trial (RCT)*, where we randomly divide the patients into two groups: one taking the drug ($x = 1$) and the other not ($x = 0$). This random choice of the treatment x is expressed by the intervention $\lceil c/x \rceil$ for $c = 0, 1$ in the diagram $G\lceil c/x \rceil$ (Fig. 1b). Since x is independent of z in $G\lceil c/x \rceil$, the causal effect $p(y|do(x=c))$ of x on the outcome y is given as y's distribution c_0 observed in the experiment in $G\lceil c/x \rceil$.

Table 1: Recovery rates of patients with/without taking a drug.

	Drug $x = 1$	No-drug $x = 0$
Male	**0.90** (18/20)	0.85 (68/80)
Female	**0.69** (55/80)	0.60 (12/20)
Total	0.73 (73/100)	**0.80** (80/100)



(a) The actual diagram G with a gender (confounder) z, a treatment x, and an outcome y.



(b) The diagram $G\lceil c/x \rceil$ with an intervention to x.

Fig. 1: Causal diagrams in Example 1.

In contrast, $\varphi_{\mathrm{BDA}} \stackrel{\mathrm{def}}{=} (f = y|_{z,x=c} \wedge\ c_1 = z \wedge\ c_0 = f(c_1) \downarrow_y)$ describes the inference about the causal effect from observation *without* intervention to x (Fig. 1a). This saves the cost of the experiment and avoids ethical issues in random treatments. Instead, to avoid Simpson's paradox, the inference φ_{BDA} conducts a *backdoor adjustment* (Section 7) to cope with the confounder z.

Concretely, the backdoor adjustment φ_{BDA} computes x's causal effect on y as follows. We first obtain the conditional distribution $f \stackrel{\mathrm{def}}{=} y|_{z,x=c}$ and the prior $c_1 \stackrel{\mathrm{def}}{=} z$. Then we conduct the adjustment by calculating the joint distribution $f(c_1)$ from f and c_1 and then taking the marginal distribution $c_0 \stackrel{\mathrm{def}}{=} f(c_1) \downarrow_y$. The resulting c_0 is the same as the c_0 in the RCT experiment φ_{RCT}; that is, the backdoor adjustment φ_{BDA} can compute the causal effect obtained by φ_{RCT}.

For this adjustment, we need to check the requirement $\mathit{pa}(z, x) \wedge \mathit{pos}(x :: z)$, that is, z is x's parent in the diagram G and the joint distribution $x :: z$ satisfies the positivity (i.e., it takes each value with a non-zero probability).

Now we formalize the *correctness* of this causal inference method (for any diagram G) as the judgment expressing that under the above requirements, the backdoor adjustment computes the same causal effect as the RCT experiment:

$$\mathit{pa}(z, x) \wedge \mathit{pos}(x :: z) \vdash_g \varphi_{\mathrm{RCT}} \leftrightarrow \varphi_{\mathrm{BDA}}. \tag{1}$$

By deriving this judgment in a deductive system called $\mathbf{AX^{CP}}$ (Section 6), we show the correctness of this causal inference method for any diagram (Section 7). We show all proofs of the technical results in Appendix.

3 Language for Data Generation

In this section, we introduce a language for describing data generation.

Constants and Causal Variables. We introduce a set Const of *constants* to denote probability distributions of data values and a set $\mathsf{dConst} \subseteq \mathsf{Const}$ of *deterministic constants*, each denoting a single data value (strictly speaking, denoting a distribution having a single data value with probability 1).

We introduce a finite set CVar of *causal variables*. A tuple $\langle x_1, \ldots, x_k \rangle$ of causal variables represents the joint distribution of k variables $x_1, \ldots, x_k$. We denote the set of all non-empty (resp. possibly empty) tuples of variables by CVar^{+} (resp. CVar^{*}). We use the bold font for a *tuple*; e.g., $\boldsymbol{x} = \langle x_1, \ldots, x_k \rangle$. We write *size*$(\boldsymbol{x})$ for the *dimension* k of a tuple $\boldsymbol{x}$. We assume that the variables in a tuple $\boldsymbol{x}$ are sorted lexicographically.

For disjoint tuples $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{x} :: \boldsymbol{y}$ denotes the *joint distribution* of $\boldsymbol{x}$ and $\boldsymbol{y}$. Formally, '::' is *not* a function symbol, but a meta-operator on CVar^{*}; $\boldsymbol{x} :: \boldsymbol{y}$ is the tuple obtained by merging $\boldsymbol{x}$ and $\boldsymbol{y}$ and sorting the variables lexicographically.

We use *conditional causal variables* $\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}}$ to denote the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{z}$ and $\boldsymbol{x} = \boldsymbol{c}$. We write FVar for the set of all conditional causal variables. For a conditional distribution $\boldsymbol{y}|_{\boldsymbol{x}}$ and a prior distribution $\boldsymbol{x}$, we write $\boldsymbol{y}|_{\boldsymbol{x}}(\boldsymbol{x})$ for the joint distribution $\boldsymbol{x} :: \boldsymbol{y}$.

Terms. We define *terms* to express how data are generated. Let Fsym be a set of *function symbols* denoting algorithms. We define the set CTerm of *causal terms* as the terms of depth at most 1; i.e., $u ::= c \mid f(v, \ldots, v)$ where $c \in \mathsf{Const}$, $f \in \mathsf{Fsym}$, and $v \in \mathsf{CVar} \cup \mathsf{Const}$. For example, $f(c)$ denotes a data generated by an algorithm f with input c. We denote the set of variables (resp. the set of constants) occurring in a term u by $\mathsf{fv}(u)$ (resp. $\mathsf{fc}(u)$).

We also define the set Term of *terms* by the BNF: $u ::= \boldsymbol{x} \mid c \mid f(u, \ldots, u)$, where $\boldsymbol{x} \in \mathsf{CVar}^{+}$, $c \in \mathsf{Const}$, and $f \in \mathsf{Fsym} \cup \mathsf{FVar}$. Unlike CTerm, terms in Term may repeatedly apply functions to describe multiple steps of data generation.

We introduce the special function symbol $\downarrow_{\boldsymbol{x}}$ for marginalization. $\boldsymbol{y} \downarrow_{\boldsymbol{x}}$ denotes the *marginal distribution* of $\boldsymbol{x}$ given a joint distribution $\boldsymbol{y}$; e.g., for a joint distribution $\boldsymbol{x} = \langle x_0, x_1 \rangle$, $\boldsymbol{x} \downarrow_{x_0}$ expresses the marginal distribution x_0. We also introduce the special constant $\bot$ for *undefined values*.

Data Generators. To describe how data are generated, we introduce the notion of a *data generator* as a function $g : \mathsf{CVar} \to \mathsf{CTerm} \cup \{\bot\}$ that maps a causal variable x to a causal term representing how the data assigned to x is generated. If $g(y) = u$ for $u \in \mathsf{CTerm}$ and $y \in \mathsf{CVar}$, we write $u \twoheadrightarrow_g y$. For instance, the data generator g in Fig. 2 models the situation in Example 1. To express that a variable x's value is generated by an algorithm f_1 with an input z, the data generator g maps x to $f_1(z)$, i.e., $f_1(z) \twoheadrightarrow_g x$. Since the causal term $f_1(z)$'s depth is at most 1, z represents the *direct cause* of x. We denote the set of all variables x satisfying $g(x) \neq \bot$ by *dom*(g), and the range of g by *range*(g).

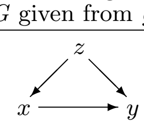
Data generator g	Causal diagram G given from g
dom$(g) = \{x, y, z\}$ $f_1(z) \twoheadrightarrow_g x$ $f_2(z, x) \twoheadrightarrow_g y$	

Fig. 2: The data generator and causal diagram for Example 1.

We assume the following *at-most-once* condition: Each function symbol and constant can be used at most once in a single data generator. This ensures that different sampling uses different randomness and is denoted by different symbols.

We say that a data generator g is *finite* if *dom*(g) is a finite set. We say that a data generator g is *closed* if no undefined variable occurs in the terms that g assigns to variables, namely, $\mathsf{fv}(\mathit{range}(g)) \subseteq \mathit{dom}(g)$.

We write $x \prec_g y$ iff y's value depends on x's, i.e., there are variables $z_1, \ldots, z_i$ $(i \geq 2)$ such that $z_1 = x$, $z_i = y$, and $z_j \in \mathsf{fv}(g(z_{j+1}))$ for $1 \leq j \leq i - 1$. A data generator g is *acyclic* if $\prec_g$ is a strict partial order over $dom(g)$. Then we can avoid the cyclic definitions of g. E.g., the data generator g_1 defined by $f(z) \rightarrow_{g_1} x$ and $f(c) \rightarrow_{g_1} z$ is acyclic, whereas g_2 by $f(z) \rightarrow_{g_2} x$ and $f(x) \rightarrow_{g_2} z$ is cyclic.

4 Kripke Model for Statistical Causality

In this section, we introduce a Kripke model for statistical causality.

We write $\mathcal{O}$ for the set of all data values we deal with, such as the Boolean values, integers, real numbers, and lists of data values. We write $\perp$ for the undefined value. For a set S, we denote the set of all probability distributions over S by $\mathbb{D}S$. For a probability distribution $m \in \mathbb{D}S$, we write $supp(m)$ for the set of m's non-zero probability elements.

Causal Diagrams. To model causal relations corresponding to a given data generator g, we consider a *causal diagram* $G = (U, V, E)$ [30] where $U \cup V$ is the set of all nodes and E is the set of all edges such that:

- $U \stackrel{\text{def}}{=} \mathsf{fc}(range(g)) \subseteq \mathsf{Const}$ is a set of symbols called *exogenous variables* that denote distributions of data;
- $V \stackrel{\text{def}}{=} dom(g) \subseteq \mathsf{CVar}$ is a set of symbols called *endogenous variables* that may depend on other variables;
- $E \stackrel{\text{def}}{=} \{x \rightarrow y \in V \times V \mid x \in \mathsf{fv}(g(y))\} \cup \{c \rightarrow y \in U \times V \mid c \in \mathsf{fc}(g(y))\}$ is the set of all *structural equations*, i.e., directed edges (arrows) denoting the direct causal relations between variables defined by the data generator g.

For instance, in Fig. 2, Example 1 is modeled as the causal diagram G.

Since a causal term's depth is at most 1, g specifies all information for defining G. By g's acyclicity, G is a directed acyclic graph (DAG) (See Proposition 4 in Appendix A.2 for details).

Pre-/Post-Intervention Distributions. For a causal diagram $G = (U, V, E)$ and a tuple $\boldsymbol{y} \subseteq V$, we write $P_G(\boldsymbol{y})$ for the joint distribution of $\boldsymbol{y}$ over $\mathcal{O}^{size(\boldsymbol{y})}$ generated according to G. As shown in the standard textbooks (e.g., [30]), $P_G(V)$ is factorized into conditional distributions according to G as follows:

$$P_G(V) \stackrel{\text{def}}{=} \prod_{y_i \in V} P_G(y_i \mid pa_G(y_i)), \tag{2}$$

where $pa_G(y_i)$ is the set of parent variables of y_i in G. For example, in Fig. 2, for $V = \{x, y, z\}$, $P_G(V) = P_G(y \mid x, z)\, P_G(x \mid z)\, P_G(z)$.

For tuples $\boldsymbol{x} \subseteq V$ and $\boldsymbol{o} \subseteq \mathcal{O}$ with $size(\boldsymbol{x}) = size(\boldsymbol{o})$, the *post-intervention distribution* $P_G(V \mid do(\boldsymbol{x}{=}\boldsymbol{o}))$ is the joint distribution of V after $\boldsymbol{x}$ is assigned $\boldsymbol{o}$ and all the variables dependent on $\boldsymbol{x}$ in G are updated by $\boldsymbol{x} := \boldsymbol{o}$ as follows:

$$P_G(V \mid do(\boldsymbol{x} = \boldsymbol{o})) \stackrel{\text{def}}{=} \begin{cases} \prod_{y_i \in V \setminus \boldsymbol{x}} P_G(y_i \mid pa_G(y_i)) & \\ \qquad \text{for values of } V \text{ consistent with } \boldsymbol{x} = \boldsymbol{o} & \\ 0 \qquad \text{otherwise.} & \end{cases}$$

For instance, in Fig. 2, $P_G(y, z | do(x = o)) = P_G(y | x = o, z)\, P_G(z)$ for any $o \in \mathcal{O}$.

Possible Worlds. We introduce the notion of a *possible world* to define the probability distribution of causal variables from a data generator. Formally, a possible world is a tuple (g, ξ, m) of (i) a finite and acyclic data generator $g : \mathsf{CVar} \to \mathsf{CTerm} \cup \{\perp\}$, (ii) an interpretation ξ that maps a function symbol in Fsym with arity $k \geq 0$ to a function from $\mathcal{O}^k$ to $\mathbb{D}\mathcal{O}$, and (iii) a memory m that maps a tuple of variables to a joint distribution of data values, which is determined by g and ξ. We denote these components of a world w by g_w, ξ_w, and m_w, and the set of all defined variables in w by $\mathsf{Var}(w) = dom(m_w)$.

The interpretation ξ can be constructed using a probability distribution I over an index set $\mathcal{I}$ and a family $\{\xi^r\}_{r\in\mathcal{I}}$ of interpretations each mapping a function symbol f with arity $k \geq 0$ to a deterministic function $\xi^r(f)$ from $\mathcal{O}^k$ to $\mathcal{O}$. Then $\xi(f)$ maps data values $\boldsymbol{o}$ to the probability distribution over $\mathcal{O}$ obtained by randomly drawing an index r from I and then computing $\xi^r(f)(\boldsymbol{o})$.

If $k = 0$, f is a constant and $\xi^r(f) \in \mathcal{O}$, hence $\xi(f) \in \mathbb{D}\mathcal{O}$ is a distribution of data values. For the undefined constant, we assume $\xi^r(\perp) = \perp$.

Interpretation of Terms. Terms are interpreted in a possible world $w = (\xi, g, m)$ as follows. First, for each index $r \in \mathcal{I}$, we define the *interpretation* $[\![\_]\!]^r_{\xi,g}$ that maps a tuple of k terms to k data values in $\mathcal{O}$ or $\perp$ by:

$$[\![\boldsymbol{x}]\!]^r_{\xi,g} = [\![g(\boldsymbol{x})]\!]^r_{\xi,g} \qquad [\![\langle u_1,\dots,u_k\rangle]\!]^r_{\xi,g} = ([\![u_1]\!]^r_{\xi,g},\ \dots,\ [\![u_k]\!]^r_{\xi,g})$$
$$[\![c]\!]^r_{\xi,g} = \xi^r(c) \qquad [\![f(u_1,\dots,u_k)]\!]^r_{\xi,g} = \xi^r(f)([\![\langle u_1,\dots,u_k\rangle]\!]^r_{\xi,g}).$$

For instance, in Fig. 2, we have $[\![x]\!]^r_{\xi,g} = [\![g(x)]\!]^r_{\xi,g} = [\![f_1(z)]\!]^r_{\xi,g} = \xi^r(f_1)([\![z]\!]^r_{\xi,g})$, where the interpretation of z does not depend on that of x due to g's acyclicity. We define the probability distribution $[\![u]\!]_w$ over $\mathcal{O}$ by randomly drawing r and then computing $[\![u]\!]^r_{\xi,g}$. Similarly, we define $[\![\langle u_1,\dots,u_k\rangle]\!]_w$ via $[\![\langle u_1,\dots,u_k\rangle]\!]^r_{\xi,g}$.

We remark that the interpretation $[\![\_]\!]_w$ defines the joint distribution P_{G_w} of all variables in the causal diagram G_w; e.g., $[\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_w = P_{G_w}(\boldsymbol{y} \mid \boldsymbol{z})$ (See Proposition 5 in Appendix A.2 for details). A function symbol f is interpreted as the function $\xi(f)$ that maps data values in $\mathcal{O}$ to the distribution over $\mathcal{O}$. We define the memory m by $m(\boldsymbol{x}) = [\![\boldsymbol{x}]\!]_w$ for all $\boldsymbol{x} \in \mathsf{CVar}^+$. Notice that $[\![\_]\!]_w$ is defined using g and ξ without using m.

We expand the interpretation $[\![\_]\!]_w$ to a conditional causal variable $\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} \in \mathsf{FVar}$ to interpret it as a function that maps a value $\boldsymbol{c}'$ of $\boldsymbol{z}$ to the distribution $[\![(\boldsymbol{x} :: \boldsymbol{y} :: \boldsymbol{z})|_{\boldsymbol{z}=\boldsymbol{c}',\boldsymbol{x}=\boldsymbol{c}}]\!]_w$. We then have $[\![\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}}(\boldsymbol{z}|_{\boldsymbol{x}=\boldsymbol{c}})]\!]_w = [\![\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}}]\!]_w([\![\boldsymbol{z}|_{\boldsymbol{x}=\boldsymbol{c}}]\!]_w)$.

For the sake of reasoning in Section 6, for each data generator g, $\boldsymbol{x} \in \mathsf{CVar}^+$, and $\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} \in \mathsf{FVar}$, we introduce a constant $c^{(g,\boldsymbol{x})}$ and a function symbol $f^{(g,\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}})}$. For brevity, we often omit the superscripts of these symbols.

Eager/Lazy Interventions. We introduce two forms of *interventions* and their corresponding *intervened worlds*. Intuitively, in a causal diagram, an *eager intervention* $\lceil c/x \rceil$ expresses the removal of all arrows pointing *to* a variable x by replacing x's value with c.

In contrast, a *lazy intervention* $\lfloor c/x \rfloor$ expresses the removal of all arrows emerging *from* x, which does not change the value of x itself but affects the values of the variables dependent on x, computed using $[\![c]\!]$ (instead of $[\![x]\!]$) as the value of x.

For instance, Fig. 3 shows how two interventions $\lceil {}^{c}\!/\!_{x} \rceil$ and $\lfloor {}^{c}\!/\!_{x} \rfloor$ change the data generator and the causal diagram in a world w that models Example 1.

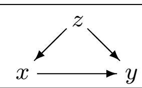
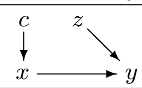
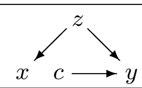
World	Data generator	Causal diagram
w	$f_1(z) \rightarrow x$; $f_2(z, x) \rightarrow y$	
$w\lceil {}^{c}\!/\!_{x} \rceil$	$c \rightarrow x$; $f_2(z, x) \rightarrow y$	
$w\lfloor {}^{c}\!/\!_{x} \rfloor$	$f_1(z) \rightarrow x$; $f_2(z, c) \rightarrow y$	

Fig. 3: Eager/lazy interventions.

For a world w and a $c \in \mathsf{dConst}$, we define an *eagerly intervened world* $w\lceil {}^{c}\!/\!_{x} \rceil$ as the world where $[\![c]\!]_w$ is assigned to x and is used to compute the other variables dependent on x. Formally, $w\lceil {}^{c}\!/\!_{x} \rceil$ is defined by $\xi_{w\lceil {}^{c}\!/\!_{x} \rceil} = \xi_w$, $g_{w\lceil {}^{c}\!/\!_{x} \rceil}(y) = c$ if $y = x$, and $g_{w\lceil {}^{c}\!/\!_{x} \rceil}(y) = g_w(y)$ if $y \neq x$. For instance, in Fig. 3, in the world $w\lceil {}^{c}\!/\!_{x} \rceil$, we use the value of c to compute $[\![x]\!]_{w\lceil {}^{c}\!/\!_{x} \rceil} = \xi_w(c)$ and $[\![y]\!]_{w\lceil {}^{c}\!/\!_{x} \rceil} = [\![f_2(z, x)]\!]_{w\lceil {}^{c}\!/\!_{x} \rceil} = [\![f_2(z, c)]\!]_w$.

Then the interpretation $[\![\_]\!]_{w\lceil {}^{c}\!/\!_{x} \rceil}$ defines the joint distribution of all variables in the causal diagram G_w after the intervention $\boldsymbol{x} := [\![\boldsymbol{c}]\!]_w$; e.g., $[\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_{w\lceil {}^{\boldsymbol{c}}\!/\!_{\boldsymbol{x}} \rceil} = P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x} = [\![\boldsymbol{c}]\!]_w), \boldsymbol{z})$ (See Proposition 5 in Appendix A.2 for details).

We next define a *lazily intervened world* $w\lfloor {}^{c}\!/\!_{x} \rfloor$ as the world where x's value is unchanged but the other variables dependent on x are computed using $[\![c]\!]_w$ instead of $[\![x]\!]_w$. Formally, $w\lfloor {}^{c}\!/\!_{x} \rfloor$ is defined by $\xi_{w\lfloor {}^{c}\!/\!_{x} \rfloor} = \xi_w$, $g_{w\lfloor {}^{c}\!/\!_{x} \rfloor}(y) = x$ if $y = x$, and $g_{w\lfloor {}^{c}\!/\!_{x} \rfloor}(y) = g_w(y)[x \mapsto c]$ if $y \neq x$. E.g., in Fig. 3, $[\![x]\!]_{w\lfloor {}^{c}\!/\!_{x} \rfloor} = [\![f_1(z)]\!]_w$.

For $\boldsymbol{x} = \langle x_1, \ldots, x_k \rangle$ and $\boldsymbol{c} = \langle c_1, \ldots, c_k \rangle$, we define $\lceil {}^{\boldsymbol{c}}\!/\!_{\boldsymbol{x}} \rceil$ from the simultaneous replacement $g_{w\lceil {}^{c_1}\!/\!_{x_1}, \ldots, {}^{c_k}\!/\!_{x_k} \rceil}$. We also define $\lfloor {}^{\boldsymbol{c}}\!/\!_{\boldsymbol{x}} \rfloor$ analogously.

Kripke Model. Let Psym be a set of predicate symbols. For a variable tuple $\boldsymbol{x}$ and a deterministic constant tuple $\boldsymbol{c}$, we introduce an *intervention relation* $w\mathcal{R}_{\lceil \boldsymbol{c}/\boldsymbol{x} \rceil}w'$ that expresses a transition from a world w to another w' by the intervention $\lceil {}^{\boldsymbol{c}}\!/\!_{\boldsymbol{x}} \rceil$; namely, $\mathcal{R}_{\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} = \{(w, w') \in \mathcal{W} \times \mathcal{W} \mid w' = w\lceil {}^{\boldsymbol{c}}\!/\!_{\boldsymbol{x}} \rceil\}$.

Then we define a *Kripke model for statistical causality* as a tuple $\mathfrak{M} = (\mathcal{W}, (\mathcal{R}_{\lceil \boldsymbol{c}/\boldsymbol{x} \rceil})_{\boldsymbol{x} \in \mathsf{CVar}^+, \boldsymbol{c} \in \mathsf{dConst}^+}, \mathcal{V})$ consisting of: (1) a set $\mathcal{W}$ of all possible worlds over the set CVar of causal variables; (2) for each $\boldsymbol{x} \in \mathsf{CVar}^+$ and $\boldsymbol{c} \in \mathsf{dConst}^+$, an *intervention relation* $\mathcal{R}_{\lceil \boldsymbol{c}/\boldsymbol{x} \rceil}$; (3) a valuation $\mathcal{V}$ that maps a k-ary predicate symbol $\eta \in \mathsf{Psym}$ to a set $\mathcal{V}(\eta)$ of k-tuples of distributions.

Notice that different worlds w and w' in $\mathcal{W}$ may have different data generators g_w and $g_{w'}$ corresponding to different causal diagrams; that is, $\mathcal{W}$ specifies all possible causal diagrams. Furthermore, different worlds w and w' may also have different interpretations ξ_w and $\xi_{w'}$ of function symbols if we do not have the knowledge of functions [23].

5 Statistical Causality Language

Predicates and Causal Predicates. Classical predicates in Psym describe *statistical correlation* among the distributions of variables, and are interpreted using a valuation $\mathcal{V}$. For example, $pos(x)$ expresses that x takes each value in the domain $\mathcal{O}$ with a non-zero probability. However, predicates cannot express the *statistical causality* among variables, whose interpretation relies on a causal diagram. Thus, we introduce a set CPsym of *causal predicates* (e.g., *dsep*, *nanc*, *allnanc*) and interpret them using a data generator g instead of a valuation $\mathcal{V}$.

Syntax and Semantics of StaCL. We define the set Fml of *formulas*: For $\eta \in \mathsf{Psym}$, $\chi \in \mathsf{CPsym}$, $\boldsymbol{x} \in \mathsf{Var}^+$, $\boldsymbol{u} \in \mathsf{Term}^+$, $\boldsymbol{c} \in \mathsf{Const}^+$, and $f \in \mathsf{Fsym} \cup \mathsf{FVar}$,

$$\varphi ::= \eta(\boldsymbol{x}, \dots, \boldsymbol{x}) \mid \chi(\boldsymbol{x}, \dots, \boldsymbol{x}) \mid \boldsymbol{u} = \boldsymbol{u} \mid f = f \mid \mathtt{true} \mid \neg\varphi \mid \varphi \wedge \varphi \mid \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi \mid \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \varphi.$$

Intuitively, $\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi$ (resp. $\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \varphi$) expresses that φ is satisfied in the eager (resp. lazy) intervened world. We assume that each variable appears at most once in $\boldsymbol{x}$ in $\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ and $\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor$. We use syntax sugar $\mathtt{false}$, $\vee$, $\rightarrow$, and $\leftrightarrow$ as usual. Note that the formulas have no quantifiers over variables.

We interpret a formula in a world w in a Kripke model $\mathfrak{M}$ by:

$$\mathfrak{M}, w \models \eta(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \text{ iff } (\llbracket x_1 \rrbracket_w, \dots, \llbracket x_k \rrbracket_w) \in \mathcal{V}(\eta)$$

$$\mathfrak{M}, w \models \boldsymbol{u} = \boldsymbol{u}' \text{ iff } \llbracket \boldsymbol{u} \rrbracket_w = \llbracket \boldsymbol{u}' \rrbracket_w \qquad \mathfrak{M}, w \models f = f' \text{ iff } \llbracket f \rrbracket_w = \llbracket f' \rrbracket_w$$

$$\mathfrak{M}, w \models \neg\varphi \text{ iff } \mathfrak{M}, w \not\models \varphi \qquad \mathfrak{M}, w \models \varphi \wedge \varphi' \text{ iff } \mathfrak{M}, w \models \varphi \text{ and } \mathfrak{M}, w \models \varphi'$$

$$\mathfrak{M}, w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi \text{ iff } \mathfrak{M}, w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \models \varphi \qquad \mathfrak{M}, w \models \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \varphi \text{ iff } \mathfrak{M}, w\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \models \varphi,$$

where $w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ and $w\lfloor \boldsymbol{u}/\boldsymbol{x} \rfloor$ are intervened worlds and the interpretation of atomic formulas with causal predicates χ is given below. For brevity, we often omit $\mathfrak{M}$.

Note that $\eta(x_1, \dots, x_k)$ represents a property of k independent distributions $\llbracket x_1 \rrbracket_w, \dots, \llbracket x_k \rrbracket_w$, where the randomness r_i in each $\llbracket x_i \rrbracket_w^{r_i}$ is chosen independently. In contrast, $\eta(\langle x_1, \dots, x_k \rangle)$ expresses a property of a single joint distribution, since the same r is used in all of $\llbracket x_1 \rrbracket_w^r, \dots, \llbracket x_k \rrbracket_w^r$.

Atomic formulas with causal predicates χ are interpreted using a causal diagram G_w corresponding to g_w. Let $\mathtt{ANC}(\boldsymbol{y})$ is the set of all ancestors of $\boldsymbol{y}$ in G_w, and $\mathtt{PA}(\boldsymbol{y})$ be the set of all parent variables of $\boldsymbol{y}$ in G_w. Then:

$$w \models \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \text{ iff } \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ are } d\text{-separated by } \boldsymbol{z} \text{ in } G_w$$

$$w \models \mathit{nanc}(\boldsymbol{x}, \boldsymbol{y}) \text{ iff } \boldsymbol{x} \cap \mathtt{ANC}(\boldsymbol{y}) = \emptyset \text{ and } \boldsymbol{x} \cap \boldsymbol{y} = \emptyset$$

$$w \models \mathit{allnanc}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \text{ iff } \boldsymbol{x} = \boldsymbol{y} \setminus \mathtt{ANC}(\boldsymbol{z})$$

$$w \models \mathit{pa}(\boldsymbol{x}, \boldsymbol{y}) \text{ iff } \boldsymbol{x} = \mathtt{PA}(\boldsymbol{y}) \text{ and } \boldsymbol{x} \cap \boldsymbol{y} = \emptyset,$$

where the d-separation [5] of $\boldsymbol{x}$ and $\boldsymbol{y}$ by $\boldsymbol{z}$ [35] is a sufficient condition for the conditional independence of $\boldsymbol{x}$ and $\boldsymbol{y}$ given $\boldsymbol{z}$ (See Appendix A for details).

Formalization of Causal Effect. Conventionally, the conditional probability of $\boldsymbol{y}$ given $\boldsymbol{z} = \boldsymbol{o_2}$ after an intervention $\boldsymbol{x} = \boldsymbol{o_1}$ is expressed using the *do*-operator by $P(\boldsymbol{y} \mid do(\boldsymbol{x} = \boldsymbol{o_1}), \boldsymbol{z} = \boldsymbol{o_2})$. This causal effect can be expressed using StaCL:

Proposition 1 (Causal effect) *Let w be a world, $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{Var}(w)^+$ be disjoint, $\boldsymbol{c} \in \mathsf{dConst}^+$, $\boldsymbol{c}' \in \mathsf{Const}^+$, and $f \in \mathsf{Fsym}$. Then:*

(i) $w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (\boldsymbol{c}' = \boldsymbol{y})$ iff there is a distribution P_{G_w} that is factorized according to G_w and satisfies $P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x} = \boldsymbol{c})) = \llbracket \boldsymbol{c}' \rrbracket_w$.

[5] An undirected path in a causal diagram G_w is said to be d-*separated* by $\boldsymbol{z}$ if it has either (a) a chain $v' \rightarrow v \rightarrow v''$ s.t. $v \in \boldsymbol{z}$, (b) a fork $v' \leftarrow v \rightarrow v''$ s.t. $v \in \boldsymbol{z}$, or (c) a collider $v' \rightarrow v \leftarrow v''$ s.t. $v \notin \boldsymbol{z} \cup \mathtt{ANC}(\boldsymbol{z})$. $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be d-*separated* by $\boldsymbol{z}$ if all undirected paths between variables in $\boldsymbol{x}$ and in $\boldsymbol{z}$ are d-separated by $\boldsymbol{z}$.

Axioms for probability distributions	
Eq$_{\mathrm{C}}$	$\vdash_g c^{(g,\boldsymbol{x})} = \boldsymbol{x}$
Eq$_{\mathrm{F}}$	$\vdash_g f^{(g,\boldsymbol{y}\|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}})} = \boldsymbol{y}\|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}}$
PD	$\vdash_g (pos(\boldsymbol{x}) \wedge c_0 = \boldsymbol{x} \wedge f = \boldsymbol{y}\|_{\boldsymbol{x}} \wedge c_1 = \boldsymbol{x} :: \boldsymbol{y}) \rightarrow c_1 = f(c_0)$
MPD	$\vdash_g \boldsymbol{x_1} \downarrow_{\boldsymbol{x_2}} = \boldsymbol{x_2}$ if $\boldsymbol{x_2} \subseteq \boldsymbol{x_1}$

Fig. 4: The axioms of **AX** for probability distributions, where $\boldsymbol{x}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{y} \in \mathsf{CVar}^+$ are disjoint, $c_0, c_1, c^{(g,\boldsymbol{x})} \in \mathsf{Const}$, $f, f^{(g,\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}})} \in \mathsf{Fsym}$.

Axioms for eager interventions	
DG$_{\mathrm{EI}}$	$\vdash_g \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi$ iff $\vdash_{g\lceil \boldsymbol{c}/\boldsymbol{x}\rceil} \varphi$
Effect$_{\mathrm{EI}}$	$\vdash_g \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (\boldsymbol{x} = \boldsymbol{c})$
Eq$_{\mathrm{EI}}$	$\vdash_g \boldsymbol{u_1} = \boldsymbol{u_2} \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (\boldsymbol{u_1} = \boldsymbol{u_2})$ if $\mathsf{fv}(\boldsymbol{u_1}) = \mathsf{fv}(\boldsymbol{u_2}) = \emptyset$
Split$_{\mathrm{EI}}$	$\vdash_g \lceil \boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2} \rceil \varphi \rightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil \varphi$
Simul$_{\mathrm{EI}}$	$\vdash_g \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil \varphi \rightarrow \lceil \boldsymbol{c'_1}/\boldsymbol{x'_1}, \boldsymbol{c_2}/\boldsymbol{x_2} \rceil \varphi$ if $\boldsymbol{x'_1} = \boldsymbol{x_1} \backslash \boldsymbol{x_2}$, $\boldsymbol{c'_1} = \boldsymbol{c_1} \backslash \boldsymbol{c_2}$
Rpt$_{\mathrm{EI}}$	$\vdash_g \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi \rightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi$
Cmp$_{\mathrm{EI}}$	$\vdash_g \big(\lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil (\boldsymbol{x_2} = \boldsymbol{c_2}) \wedge \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil (\boldsymbol{x_3} = \boldsymbol{u})\big) \rightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2} \rceil (\boldsymbol{x_3} = \boldsymbol{u})$
Distr$_{\mathrm{EI}}{}^{\neg}$	$\vdash_g (\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \neg\varphi) \leftrightarrow (\neg \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi)$
Distr$_{\mathrm{EI}}{}^{\wedge}$	$\vdash_g (\lceil \boldsymbol{c}/\boldsymbol{x} \rceil (\varphi_1 \wedge \varphi_2)) \leftrightarrow (\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi_1 \wedge \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi_2)$

Axioms for lazy interventions	
Cond$_{\mathrm{LI}}$	$\vdash_g (f = \boldsymbol{y}\|_{\boldsymbol{x}=\boldsymbol{c}}) \leftrightarrow \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor (f = \boldsymbol{y}\|_{\boldsymbol{x}=\boldsymbol{c}})$
Other axioms are analogous to eager interventions except for Effect$_{\mathrm{EI}}$.	

Axioms for the exchanges of eager and lazy interventions	
Expd$_{\mathrm{EILI}}$	$\vdash_g (\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \boldsymbol{c'} = \boldsymbol{y}) \leftrightarrow (\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \boldsymbol{c'} = \boldsymbol{y})$
Excd$_{\mathrm{EILI}}$	$\vdash_g pos(\boldsymbol{z}) \rightarrow \big((\lceil \boldsymbol{c}/\boldsymbol{x} \rceil f = \boldsymbol{y}\|_{\boldsymbol{z}}) \leftrightarrow (\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor f = \boldsymbol{y}\|_{\boldsymbol{z}})\big)$

Fig. 5: The axioms of **AX**, where $\boldsymbol{x}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{CVar}^+$ are disjoint, $f \in \mathsf{Fsym}$, $\boldsymbol{c}, \boldsymbol{c_1}, \boldsymbol{c_2} \in \mathsf{dConst}^+$, $\boldsymbol{c'} \in \mathsf{Const}^+$, $\boldsymbol{u}, \boldsymbol{u_1}, \boldsymbol{u_2} \in \mathsf{Term}^+$, and $\varphi, \varphi_1, \varphi_2 \in \mathsf{Fml}$.

(ii) $w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (f = \boldsymbol{y}|_{\boldsymbol{z}})$ *iff there is a distribution* P_{G_w} *that is factorized according to* G_w *and satisfies* $P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x} = \boldsymbol{c}), \boldsymbol{z}) = [\![f]\!]_w$.

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$, they are conditionally independent given $\boldsymbol{z}$ [35] (but not vice versa). StaCL can express this by $\models_g (dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z}) \rightarrow \boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} = \boldsymbol{y}|_{\boldsymbol{z}}$, where $pos(\boldsymbol{z})$ means that $\boldsymbol{z}$ takes each value with a positive probability, and $\models_g \varphi$ is defined as $w \models_g \varphi$ for all world w having the data generator g. Furthermore, if $[\![\boldsymbol{x}]\!]_w$ and $[\![\boldsymbol{y}]\!]_w$ are conditionally independent given $[\![\boldsymbol{z}]\!]_w$ for any world w with the data generator g_w, then they are d-separated by $\boldsymbol{z}$: $\models_g (pos(\boldsymbol{z}) \rightarrow \boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} = \boldsymbol{y}|_{\boldsymbol{z}})$ implies $\models_g dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ (See Proposition 15 in Appendix C.1).

6 Axioms for StaCL

We present a sound deductive system for StaCL in the Hilbert style. Our system consists of axioms and rules for the judgments of the form $\Gamma \vdash_g \varphi$.

The deductive system is stratified into two groups. The system **AX**, determined by the axioms in Figs. 4 and 5, concerns the derivation of $\Gamma \vdash_g \varphi$ that does not involve causal predicates (e.g., *pa*, *nanc*, *dsep*). The system $\mathbf{AX^{CP}}$, determined by the axioms in Fig. 6, concerns the derivation of a formula φ possibly equipped with causal predicates in a judgment $\Gamma \vdash_g \varphi$.

In these systems, we deal only with the reasoning that is independent of a causal diagram. Indeed, in Section 7, we will present examples of reasoning using the deductive system $\mathbf{AX^{CP}}$ that do not refer to a specific causal diagram.

Axioms of AX. Fig. 4 shows the axioms of the deductive system **AX**, where we omitted the axioms for propositional logic and equations (PT for the propositional tautologies, MP for the modus ponens, EQ1 for the reflexivity, and EQ2 for the substitutions for formulas). EQ_{C} and EQ_{F} represent the definitions of constants and function symbols corresponding to causal variables. PD describes the relationships among the prior distribution $\boldsymbol{x}$, the conditional distribution $\boldsymbol{y}|_{\boldsymbol{x}}$ of $\boldsymbol{y}$ given $\boldsymbol{x}$, and the joint distribution $\boldsymbol{x} :: \boldsymbol{y}$. MPD represents the computation $\downarrow_{\boldsymbol{x_2}}$ of the marginal distribution $\boldsymbol{x_2}$ from a joint distribution $\boldsymbol{x_1}$.

The axioms named with the subscript EI deal with eager intervention. Remarkably, DG_{EI} reduces the derivation of $\vdash_g \lceil \boldsymbol{c}/\boldsymbol{x} \rceil \varphi$, which involves an intervention modality $\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$, to the derivation of $\vdash_{g\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} \varphi$, which does not involve the modality under the modified data generator $g\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$. The axioms $\text{DISTR}_{\text{EI}}{}^{\neg}$ and $\text{DISTR}_{\text{EI}}{}^{\wedge}$ allow for pushing intervention operators outside logical connectives.

The axioms with the subscript LI deal with lazy intervention; they are analogous to the corresponding EI-rules. The axioms with the subscript EILI describe when an eager intervention can be exchanged with a lazy intervention.

Axioms of $\mathbf{AX^{CP}}$. Fig. 6 shows the axioms for $\mathbf{AX^{CP}}$. DSEPCI represents that d-separation implies conditional independence. DSEPSM, DSEPDC, DSEPWU, and DSEPCN are the *semi-graphoid* axioms [35], characterizing the d-separation. However, these well-known axioms are not sufficient to derive the relationships between d-separation and interventions. Therefore, we introduce two axioms DSEP_{EI} and DSEP_{LI} in Fig. 6 for the d-separation before/after interventions, and four axioms to reason about the relationships between the causal predicate *nanc* and the interventions/d-separation (named $\text{NANC}_{\{1,2,3,4\}}$ in Fig. 6). By ALLNANC, PANANC, and PADSEP, we transform the formulas using *allnanc* and *pa* into those with *nanc* or *dsep*.

Properties of Axiomatization. For a data generator g, a set $\Gamma \stackrel{\text{def}}{=} \{\psi_1, \ldots, \psi_n\}$ of formulas, and a formula φ, we write $\Gamma \vdash_g \varphi$ if there is a derivation of $\vdash_g (\psi_1 \wedge \cdots \wedge \psi_n) \rightarrow \varphi$ using axioms of **AX** or $\mathbf{AX^{CP}}$. We write $\Gamma \models_g \varphi$ if for all model $\mathfrak{M}$ and all world w having the data generator g, $\mathfrak{M}, w \models \varphi$. Then we obtain the *soundness* of **AX** and $\mathbf{AX^{CP}}$.

Theorem 1 (Soundness) *Let g be a finite, closed, and acyclic data generator. $\Gamma \subseteq \mathsf{Fml}$, and $\varphi \in \mathsf{Fml}$. If $\Gamma \vdash_g \varphi$ then $\Gamma \models_g \varphi$.*

Axioms for d-separation

DsepCI	$\vdash_g (dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \land pos(\boldsymbol{z})) \rightarrow \boldsymbol{y}\|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} = \boldsymbol{y}\|_{\boldsymbol{z}}$
DsepSm	$\vdash_g dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \leftrightarrow dsep(\boldsymbol{y},\boldsymbol{x},\boldsymbol{z})$
DsepDc	$\vdash_g dsep(\boldsymbol{x},\boldsymbol{y}\cup\boldsymbol{y}',\boldsymbol{z}) \rightarrow (dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \land dsep(\boldsymbol{x},\boldsymbol{y}',\boldsymbol{z}))$
DsepWu	$\vdash_g dsep(\boldsymbol{x},\boldsymbol{y}\cup\boldsymbol{v},\boldsymbol{z}) \rightarrow dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\cup\boldsymbol{v})$
DsepCn	$\vdash_g (dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \land dsep(\boldsymbol{x},\boldsymbol{v},\boldsymbol{z}\cup\boldsymbol{y})) \rightarrow dsep(\boldsymbol{x},\boldsymbol{y}\cup\boldsymbol{v},\boldsymbol{z})$

Axioms for d-separation with interventions

Dsep$_{\text{EI}}$	$\vdash_g (\lceil \boldsymbol{c}/\boldsymbol{z} \rceil dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})) \leftrightarrow dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$
Dsep$_{\text{LI}}$	$\vdash_g (\lfloor \boldsymbol{c}/\boldsymbol{z} \rfloor dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})) \leftrightarrow dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$

Axioms with other causal predicates

Nanc1	$\vdash_g (nanc(\boldsymbol{x},\boldsymbol{y}) \land nanc(\boldsymbol{x},\boldsymbol{z})) \rightarrow (f = \boldsymbol{y}\|_{\boldsymbol{z}} \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (f = \boldsymbol{y}\|_{\boldsymbol{z}}))$
Nanc2	$\vdash_g nanc(\boldsymbol{x},\boldsymbol{y}) \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil nanc(\boldsymbol{x},\boldsymbol{y})$
Nanc3	$\vdash_g nanc(\boldsymbol{x},\boldsymbol{y}) \rightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil dsep(\boldsymbol{x},\boldsymbol{y},\emptyset)$
Nanc4	$\vdash_g (nanc(\boldsymbol{x},\boldsymbol{z}) \land dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})) \rightarrow nanc(\boldsymbol{x},\boldsymbol{y})$
AllNanc	$\vdash_g allnanc(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \rightarrow nanc(\boldsymbol{x},\boldsymbol{z})$
PaNanc	$\vdash_g pa(\boldsymbol{x},\boldsymbol{y}) \rightarrow nanc(\boldsymbol{y},\boldsymbol{x})$
PaDsep	$\vdash_g pa(\boldsymbol{z},\boldsymbol{x}) \rightarrow \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$

Fig. 6: The additional axioms for $\mathbf{AX^{CP}}$ where $\boldsymbol{x},\boldsymbol{y},\boldsymbol{y}',\boldsymbol{z},\boldsymbol{v} \in \mathsf{CVar}^+$ are disjoint, $\boldsymbol{c} \in \mathsf{dConst}^+$, and $f \in \mathsf{Fsym}$.

See Appendices B and C for the proof. As shown in Section 7, $\mathbf{AX^{CP}}$ is expressive enough to derive the rules of Pearl's do-calculus [29]; it can reason about all causal effects identifiable by the do-calculus (without referring to a specific causal diagram). Furthermore, **AX** includes/derives the axioms used in the previous work [1] that are complete w.r.t. a different semantics without dealing with probability distributions. We leave investigating whether **AX** is complete w.r.t. our Kripke model for future work. We also remark that $\mathbf{AX^{CP}}$ has axioms corresponding to the composition and effectiveness axioms introduced by Galles and Pearl [9].

7 Reasoning About Statistical Causality

Deriving the Rules of the Do-Calculus. Using StaCL, we express the *do-calculus*'s rules [29], which are sufficient to compute all identifiable causal effects from observable quantities [16, 32]. Let $\mathsf{fv}(\varphi)$ be the set of all variables occurring in a formula φ, and $\mathsf{cdv}(\varphi)$ be the set of all *conditioning variables* in φ.

Proposition 2 (Do-calculus rules) *Let $\boldsymbol{v},\boldsymbol{x},\boldsymbol{y},\boldsymbol{z} \in \mathsf{CVar}^+$ be disjoint, $\boldsymbol{x_1},\boldsymbol{x_2} \in \mathsf{CVar}^+$, and $\boldsymbol{c_0},\boldsymbol{c_1},\boldsymbol{c_2} \in \mathsf{dConst}^+$. Let $S = \mathsf{cdv}(\varphi_0) \cup \mathsf{cdv}(\varphi_1)$.*

1. Do1. *Introduction/elimination of conditioning:*

$$\vdash_g \lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil (dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \land \textstyle\bigwedge_{\boldsymbol{s}\in S} pos(\boldsymbol{s})) \rightarrow ((\lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil \varphi_0) \leftrightarrow \lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil \varphi_1)$$

$$
\dfrac{
\dfrac{}{\vdash_g \psi_{\mathrm{d1}} \rightarrow ((\lceil c/x \rceil \psi_0) \leftrightarrow \psi_1)}\,\textsc{Do2}
\quad
\dfrac{\dfrac{}{\vdash_g \psi_{\mathrm{nanc}} \rightarrow ((\lceil c/x \rceil \psi_2) \leftrightarrow \psi_2)}\,\textsc{Nanc3}}{\vdash_g \psi_{\mathrm{d2}} \rightarrow ((\lceil c/x \rceil \psi_2) \leftrightarrow \psi_2)}\,\textsc{Do3}
\quad
\dfrac{}{\vdash_g (\lceil c/x \rceil \psi_3) \leftrightarrow \psi_3}\,\textsc{Eq}_{\textsc{EI}}
}{
\dfrac{\psi_{\mathrm{pre}} \vdash_g (\lceil c/x \rceil \psi_0 \wedge \lceil c/x \rceil \psi_2 \wedge \lceil c/x \rceil \psi_3) \leftrightarrow (\psi_1 \wedge \psi_2 \wedge \psi_3)}{
\dfrac{\psi_{\mathrm{pre}} \vdash_g (\lceil c/x \rceil (\psi_0 \wedge \psi_2 \wedge \psi_3)) \leftrightarrow (\psi_1 \wedge \psi_2 \wedge \psi_3)}{
\dfrac{\psi_{\mathrm{pre}} \vdash_g (\lceil c/x \rceil c_0 = (\boldsymbol{y}|_{\boldsymbol{z}}(\boldsymbol{z}))\!\downarrow_{\boldsymbol{y}}) \leftrightarrow (\psi_1 \wedge \psi_2 \wedge \psi_3)}{
\dfrac{\psi_{\mathrm{pre}} \vdash_g (\lceil c/x \rceil c_0 = (\boldsymbol{y} :: \boldsymbol{z})\!\downarrow_{\boldsymbol{y}}) \leftrightarrow (\psi_1 \wedge \psi_2 \wedge \psi_3)}{
\psi_{\mathrm{pre}} \vdash_g (\lceil c/x \rceil c_0 = \boldsymbol{y}) \leftrightarrow (\psi_1 \wedge \psi_2 \wedge \psi_3)
}\,\textsc{MPD}, \textsc{Eq2}
}\,\textsc{PD}, \textsc{Eq2}
}\,\textsc{Eq}_{\textsc{C}}, \textsc{Eq}_{\textsc{F}}, \textsc{Eq2}
}\,\textsc{Distr}_{\textsc{EI}}^{\wedge}
}
}
$$

Fig. 7: Sketch of a derivation tree for the correctness of the backdoor adjustment (Section 2) using $\mathbf{AX^{CP}}$ where $\psi_{\mathrm{pos}} \stackrel{\mathrm{def}}{=} pos(\boldsymbol{z} :: \boldsymbol{x})$, $\psi_{\mathrm{d1}} \stackrel{\mathrm{def}}{=} \lfloor c/x \rfloor\, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \psi_{\mathrm{pos}}$, $\psi_{\mathrm{d2}} \stackrel{\mathrm{def}}{=} \lceil c/x \rceil\, dsep(\boldsymbol{x}, \boldsymbol{z}, \emptyset) \wedge \psi_{\mathrm{pos}}$, $\psi_{\mathrm{nanc}} \stackrel{\mathrm{def}}{=} nanc(\boldsymbol{x}, \boldsymbol{z}) \wedge \psi_{\mathrm{pos}}$, $\psi_{\mathrm{pre}} \stackrel{\mathrm{def}}{=} \psi_{\mathrm{d1}} \wedge \psi_{\mathrm{nanc}}$, $\psi_0 \stackrel{\mathrm{def}}{=} (f = \boldsymbol{y}|_{\boldsymbol{z}})$, $\psi_1 \stackrel{\mathrm{def}}{=} (f = \boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c}})$, $\psi_2 \stackrel{\mathrm{def}}{=} (c_1 = \boldsymbol{z})$, and $\psi_3 \stackrel{\mathrm{def}}{=} (c_0 = f(c_1)\!\downarrow_{\boldsymbol{y}})$.

where φ_1 *is obtained by replacing some occurrences of* $\boldsymbol{y}|_{\boldsymbol{z}}$ *in* φ_0 *with* $\boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}}$;

2. Do2. *Exchange between intervention and conditioning:*

$$\vdash_g \lceil c_0/v \rceil \lfloor c_1/x \rfloor (dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \textstyle\bigwedge_{\boldsymbol{s} \in S} pos(\boldsymbol{s})) \rightarrow ((\lceil c_0/v, c_1/x \rceil \varphi_0) \leftrightarrow \lceil c_0/v \rceil \varphi_1)$$

where φ_1 *is obtained by replacing every occurrence of* $\boldsymbol{y}|_{\boldsymbol{z}}$ *in* φ_0 *with* $\boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}}$;

3. Do3 *Introduction/elimination of intervention:*

$$\vdash_g \lceil c_0/v \rceil (allnanc(\boldsymbol{x_1}, \boldsymbol{x}, \boldsymbol{y}) \wedge \lceil c_1/x_1 \rceil (dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z}))) \rightarrow ((\lceil c_0/v \rceil \varphi) \leftrightarrow \lceil c_0/v, c_1/x_1, c_2/x_2 \rceil \varphi)$$

where $\mathsf{fv}(\varphi) = \{\boldsymbol{y}|_{\boldsymbol{z}}\}$ *and* $\boldsymbol{x} \stackrel{def}{=} \boldsymbol{x_1} :: \boldsymbol{x_2}$.

By using the deductive system $\mathbf{AX^{CP}}$, we can derive those rules. Thanks to the modal operators for lazy interventions, our derivation of those rules is partly different from Pearl's [29] in that it does not use diagrams augmented with the intervention arc of the form $F_x \twoheadrightarrow x$ (See Appendix D for details).

Reasoning About Statistical Adjustment. We present how $\mathbf{AX^{CP}}$ can be used to reason about the correctness of the backdoor adjustment discussed in Section 2 (See Appendix A.6 for the details of the backdoor adjustment). Fig. 7 shows the derivation of the judgment:

$$\psi_{\mathrm{pre}} \vdash_g (\lceil c/x \rceil c_0 = \boldsymbol{y}) \leftrightarrow (\psi_1 \wedge \psi_2 \wedge \psi_3). \tag{3}$$

This judgment asserts the correctness of the backdoor adjustment in any causal diagram. Recall that $\varphi_{\mathrm{RCT}} \stackrel{\mathrm{def}}{=} (\lceil c/x \rceil c_0 = \boldsymbol{y})$ expresses the RCT and $\varphi_{\mathrm{BDA}} \stackrel{\mathrm{def}}{=} (\psi_1 \wedge \psi_2 \wedge \psi_3)$ expresses the backdoor adjustment. The correctness of the backdoor adjustment ($\varphi_{\mathrm{RCT}} \leftrightarrow \varphi_{\mathrm{BDA}}$) depends on the precondition ψ_{pre}.

By reading the derivation tree in a bottom-up manner, we observe that the proof first converts $(\lceil c/x \rceil c_0 = \boldsymbol{y})$ to a formula to which $\textsc{Eq}_{\textsc{C}}$ and $\textsc{Eq}_{\textsc{F}}$ are applicable. Then, the derived axioms Do2 and Do3 in Proposition 2 are used to complete the proof at the leaves of the derivation.

In Section 2, we stated the correctness of the backdoor adjustment in (1) using a simpler requirement $pa(z, x)$ instead of ψ_{d1} and ψ_{nanc}. We can derive the judgment (1) from (3), thanks to the axioms PaDsep and PaNanc.

The derivation does not mention the data generator g representing the causal diagram G. This exhibits that our logic successfully separates the reasoning about the properties of arbitrary causal diagrams from those depending on a specific causal diagram. Once we prove $\psi_{\mathrm{pre}} \vdash_g \varphi_{\mathrm{RCT}} \leftrightarrow \varphi_{\mathrm{BDA}}$ using $\mathbf{AX^{CP}}$, one can claim the correctness of the causal inference ($\varphi_{\mathrm{RCT}} \leftrightarrow \varphi_{\mathrm{BDA}}$) by checking that the requirement ψ_{pre} indeed holds for a specific causal diagram G.

8 Conclusion

We proposed statistical causality language (StaCL) to formally describe and explain the correctness of statistical causal inference. We introduced the notion of causal predicates and Kripke models equipped with data generators. We defined a sound deductive system $\mathbf{AX^{CP}}$ that can deduce all causal effects derived using Pearl's do-calculus. In ongoing and future work, we study the completeness of $\mathbf{AX}$ and $\mathbf{AX^{CP}}$ and develop a decision procedure for $\mathbf{AX^{CP}}$ for automated reasoning.

Acknowledgements. We thank Kenji Fukumizu for providing helpful information on the literature on causal inference. The authors are supported by ERATO HASUO Metamathematics for Systems Design Project (No. JPMJER1603), JST. Yusuke Kawamoto is supported by JST, PRESTO Grant Number JPMJPR2022, Japan, and by JSPS KAKENHI Grant Number 21K12028, Japan. Tetsuya Sato is supported by JSPS KAKENHI Grant Number 20K19775, Japan. Kohei Suenaga is supported by JST CREST Grant Number JPMJCR2012, Japan.

References

1. Barbero, F., Sandu, G.: Team semantics for interventionist counterfactuals: Observations vs. interventions. J. Philos. Log. **50**(3), 471–521 (2021). https://doi.org/10.1007/s10992-020-09573-6
2. Barbero, F., Schulz, K., Smets, S., Velázquez-Quesada, F.R., Xie, K.: Thinking about causation: A causal language with epistemic operators. In: Proc. the third International Workshop on Dynamic Logic (DaLí'20). LNCS, vol. 12569, pp. 17–32. Springer (2020). https://doi.org/10.1007/978-3-030-65840-3_2
3. Bochman, A.: A logical theory of causality. MIT Press (2021)
4. Bochman, A., Lifschitz, V.: Pearl's causality in a logical setting. In: Proc. the Twenty-Ninth AAAI Conference on Artificial Intelligence. pp. 1446–1452. AAAI Press (2015), http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9686
5. Corander, J., Hyttinen, A., Kontinen, J., Pensar, J., Väänänen, J.: A logical approach to context-specific independence. Ann. Pure Appl. Log. **170**(9), 975–992 (2019). https://doi.org/10.1016/j.apal.2019.04.004
6. Durand, A., Hannula, M., Kontinen, J., Meier, A., Virtema, J.: Approximation and dependence via multiteam semantics. In: Gyssens, M., Simari, G.R. (eds.)

Proc. the 9th International Symposium on the Foundations of Information and Knowledge Systems (FoIKS'16). LNCS, vol. 9616, pp. 271–291. Springer (2016). https://doi.org/10.1007/978-3-319-30024-5_15

7. Durand, A., Hannula, M., Kontinen, J., Meier, A., Virtema, J.: Probabilistic team semantics. In: Proc. the 10th International Symposium on the Foundations of Information and Knowledge Systems (FoIKS'18). LNCS, vol. 10833, pp. 186–206. Springer (2018). https://doi.org/10.1007/978-3-319-90050-6_11
8. Fernandes-Taylor, S., Hyun, J.K., Reeder, R.N., Harris, A.H.: Common statistical and research design problems in manuscripts submitted to high-impact medical journals. BMC Research Notes **4**(1), 304 (2011). https://doi.org/10.1186/1756-0500-4-304, https://doi.org/10.1186/1756-0500-4-304
9. Galles, D., Pearl, J.: An axiomatic characterization of causal counterfactuals. Foundations of Science **3**, 151–182 (1998)
10. Halpern, J.Y.: Axiomatizing causal reasoning. J. Artif. Intell. Res. **12**, 317–337 (2000). https://doi.org/10.1613/jair.648, https://doi.org/10.1613/jair.648
11. Halpern, J.Y.: A modification of the Halpern-Pearl definition of causality. In: Proc. IJCAI'15. pp. 3022–3033. AAAI Press (2015)
12. Halpern, J.Y., Pearl, J.: Causes and explanations: A structural-model approach - part II: explanations. In: Proc. IJCAI'01. pp. 27–34. Morgan Kaufmann (2001)
13. Halpern, J.Y., Pearl, J.: Causes and explanations: A structural-model approach: Part 1: Causes. In: Proc. UAI'01. pp. 194–202. Morgan Kaufmann (2001)
14. Hirvonen, Å., Kontinen, J., Pauly, A.: Continuous team semantics. In: Proc. the 15th Annual Conference on Theory and Applications of Models of Computation (TAMC'19). LNCS, vol. 11436, pp. 262–278. Springer (2019). https://doi.org/10.1007/978-3-030-14812-6_16
15. Hodges, W.: Compositional semantics for a language of imperfect information. Log. J. IGPL **5**(4), 539–563 (1997). https://doi.org/10.1093/jigpal/5.4.539
16. Huang, Y., Valtorta, M.: Pearl's calculus of intervention is complete. In: Proc. UAI'06. p. 217–224. AUAI Press (2006)
17. Hyttinen, A., Eberhardt, F., Järvisalo, M.: Constraint-based causal discovery: Conflict resolution with answer set programming. In: Proc. the Thirtieth Conference on Uncertainty in Artificial Intelligence (UAI'14). pp. 340–349. AUAI Press (2014)
18. Hyttinen, A., Eberhardt, F., Järvisalo, M.: Do-calculus when the true graph is unknown. In: Proc. the Thirty-First Conference on Uncertainty in Artificial Intelligence (UAI'15). pp. 395–404. AUAI Press (2015)
19. Ibeling, D., Icard, T.: Probabilistic reasoning across the causal hierarchy. In: Proc. the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI'20). pp. 10170–10177. AAAI Press (2020), https://aaai.org/ojs/index.php/AAAI/article/view/6577
20. Kawamoto, Y.: Statistical epistemic logic. In: The Art of Modelling Computational Systems: A Journey from Logic and Concurrency to Security and Privacy. LNCS, vol. 11760, pp. 344–362. Springer (2019). https://doi.org/10.1007/978-3-030-31175-9_20
21. Kawamoto, Y.: Towards logical specification of statistical machine learning. In: Proc. SEFM. pp. 293–311 (2019). https://doi.org/10.1007/978-3-030-30446-1_16
22. Kawamoto, Y.: An epistemic approach to the formal specification of statistical machine learning. Software and Systems Modeling **20**(2), 293–310 (2020). https://doi.org/10.1007/s10270-020-00825-2
23. Kawamoto, Y., Mano, K., Sakurada, H., Hagiya, M.: Partial knowledge of functions and verification of anonymity. Transactions of the Japan

Society for Industrial and Applied Mathematics **17**(4), 559–576 (2007). https://doi.org/10.11540/jsiamt.17.4_559
24. Kawamoto, Y., Sato, T., Suenaga, K.: Formalizing statistical beliefs in hypothesis testing using program logic. In: Proc. KR'21. pp. 411–421 (2021). https://doi.org/10.24963/kr.2021/39
25. Kawamoto, Y., Sato, T., Suenaga, K.: Sound and relatively complete belief Hoare logic for statistical hypothesis testing programs. CoRR **abs/2208.07074** (2022)
26. Makin, T.R., de Xivry, J.J.O.: Science forum: Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. Elife **8**, e48175 (2019)
27. McCain, N., Turner, H.: Causal theories of action and change. In: Proc. the Fourteenth National Conference on Artificial Intelligence and Ninth Innovative Applications of Artificial Intelligence Conference (AAAI'97/IAAI'97). pp. 460–465. AAAI Press / The MIT Press (1997)
28. Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzsche, P.C., Devereaux, P., Elbourne, D., Egger, M., Altman, D.G.: Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. International journal of surgery **10**(1), 28–55 (2012)
29. Pearl, J.: Causal diagrams for empirical research. Biometrika **82**(4), 669–688 (1995), http://www.jstor.org/stable/2337329
30. Pearl, J.: Causality. Cambridge university press (2009)
31. Rückschloß, K., Weitkämper, F.: Exploiting the full power of Pearl's causality in probabilistic logic programming. In: Proc. the 9th Workshop on Probabilistic Logic Programming (PLP'22). CEUR Workshop Proceedings, vol. 3193. CEUR-WS.org (2022), http://ceur-ws.org/Vol-3193/paper1PLP.pdf
32. Shpitser, I., Pearl, J.: Identification of conditional interventional distributions. In: Proc. UAI'06. p. 437–444. AUAI Press (2006)
33. Simpson, E.H.: The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society. Series B (Methodological) **13**(2), 238–241 (1951), http://www.jstor.org/stable/2984065
34. Triantafillou, S., Tsamardinos, I.: Constraint-based causal discovery from multiple interventions over overlapping variable sets. J. Mach. Learn. Res. **16**, 2147–2205 (2015)
35. Verma, T., Pearl, J.: Causal networks: semantics and expressiveness. In: Proc. UAI'88. pp. 69–78. North-Holland (1988)
36. Von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P.: The strengthening the reporting of observational studies in epidemiology (strobe) statement: guidelines for reporting observational studies. Bulletin of the World Health Organization **85**, 867–872 (2007)

Appendix

We present the following technical details:

- Appendix A presents the details of our models and causality.
- Appendix B proves the soundness of the deductive system $\mathbf{AX}$ for StaCL.
- Appendix C shows the soundness of the deductive system $\mathbf{AX}^{\mathbf{CP}}$ for StaCL with causal predicates.
- Appendix D presents the details of the reasoning and the explanation about statistical causality using StaCL.

We first introduce notations. We present key notations in Tables 2 and 3.

For tuples $\boldsymbol{x}$ and $\boldsymbol{x'}$ of variables, we write $\boldsymbol{x} \subseteq \boldsymbol{x'}$ iff every variable in $\boldsymbol{x}$ appears in $\boldsymbol{x'}$. $\boldsymbol{x} \setminus \boldsymbol{y}$ is the tuple of variables obtained by removing all variables in $\boldsymbol{y}$ from $\boldsymbol{x}$. As with ::, the symbol '$\setminus$' is a meta-operator on sets of variables, and *not* a function symbol. For brevity, we identify a singleton tuple $\langle x \rangle$ as its element x.

For $u, u' \in \mathsf{Term}$ and $x \in \mathsf{CVar}$, the *substitution* $u[x \mapsto u']$ is the term obtained by replacing every occurrence of x in u with u'.

We recall that a *memory* is a joint probability distribution $m \in \mathbb{D}(\mathsf{Var} \to \mathcal{O} \cup \{\perp\})$ of data values of all variables in Var. We write $m(\boldsymbol{x})$ for the joint distribution of all variables in $\boldsymbol{x}$.

A Details on Models and Causality

In this section, we show a couple of remarks on the interpretation of terms (Appendix A.1), the relationships between data generators and causal diagrams (Appendix A.2), and properties on memories (Appendix A.3). Then we present more details on causal predicates (Appendix A.4), causal effects (Appendix A.5), and causal diagrams (Appendix A.6).

Table 2: Notations in syntax.

Symbol	Description
CVar	Set of causal variable
FVar	Set of conditional causal variables
$\mathsf{fv}(\varphi)$	Set of all free variables in a formula φ
$\mathsf{cdv}(\varphi)$	Set of all conditioning variables in a formula φ
Fsym	Set of function symbols
pFsym	Set of probabilistic function symbols
dFsym	Set of deterministic function symbols
Const	Set of constants
dConst	Set of deterministic constants
$\perp$	Constant denoting the undefined value
CTerm	Set of causal terms
Term	Set of terms
Psym	Set of predicates
CPsym	Set of causal predicates
Fml	Set of (causality) formulas

A.1 Remarks on the Interpretation of Terms

We present remarks on the *at-most-once condition*, on deterministic functions, and on the well-definedness of the interpretation of terms.

Table 3: Notations in semantics.

Symbol	Description
$\mathfrak{M}$	Kripke model
$\mathcal{W}$	Set of all possible worlds
$\mathcal{R}_{\lceil \boldsymbol{c}/\boldsymbol{x} \rceil}$	Intervention relation
$\mathcal{V}$	Valuation
$\mathcal{O}$	Domain of data values
$\mathbb{D}\mathcal{O}$	Set of all probability distributions over $\mathcal{O}$
w	Possible world
$w\lceil {}^{c}/_{x} \rceil$	Eagerly intervened world
$w\lfloor {}^{c}/_{x} \rfloor$	Lazily intervened world
g_w	Data generator in a world w
$x \prec_g y$	y's value depends on x's in a data generator g
$\xi_w = \{\xi_w^r\}_{r \sim I}$	Interpretation of function symbols in a world w
m_w	Memory on variables in a world w
$G_w = (U, V, E)$	Causal diagram in a world w
$pa_{G_w}(v)$	All parent variables of v in G_w
$P_{G_w}(V)$	Joint distribution of all variables V in G_w

Remark on At-Most-Once Condition We remark on the at-most-once condition. In Section 3, we assumed that a data generator satisfies the following *at-most-once* condition: Each function symbol f and each constant c can be used at most once in a single data generator. For example, we may consider the data generator g_3 defined by $c \rightarrow_{g_3} z$ and $f(z,z) \rightarrow_{g_1} y$ as it is. Then g_3 rewrites y into $f(c,c)$ after substitutions. In contrast, the data generator g_4 defined by

$$c \rightarrow_{g_4} z_1,\ c \rightarrow_{g_4} z_2,\ f(z_1, z_2) \rightarrow_{g_4} y \tag{4}$$

also rewrites y into $f(c,c)$, but does not satisfy the at-most-once condition. Thus, the two calls of the constant c should be distinguished and replaced with two symbols c_1 and c_2 (denoting the same distribution of data values as c):

$$c_1 \rightarrow_{g_4} z_1,\ c_2 \rightarrow_{g_4} z_2,\ f(z_1, z_2) \rightarrow_{g_4} y. \tag{5}$$

Then g_4 rewrites y into $f(c_1, c_2)$. This at-most-once condition clarifies that an occurrence of a probabilistic constant c_i represents a single independent sampling. In the former definition (4) of g_4, which does not satisfy the at-most-once condition, it is not clear whether (i) c is sampled once and its (single) value is assigned to both y_1 and y_2, or (ii) it is sampled twice and the drawn (two) values are assigned to y_1 and y_2. By imposing the at-most-once condition as in the latter definition (5), we can clarify that there are two occurrences of sampling that may use different randomness.

Remark on Interpretation of Terms We remark that two copies of the same data value are obtained from the distribution $[\![\langle c, c\rangle]\!]_w$, whereas two different values may be drawn from $[\![\langle c_1, c_2\rangle]\!]_w$ for constants c_1, c_2 denoting the same distribution, i.e., $[\![c_1]\!]_w = [\![c_2]\!]_w$. Indeed, for a randomly chosen r, the former results in $[\![\langle c, c\rangle]\!]^r_{\xi,g} = (\xi^r(c), \xi^r(c))$. In contrast, the latter results in $[\![\langle c_1, c_2\rangle]\!]^r_{\xi,g} = (\xi^r(c_1), \xi^r(c_2))$, where two data values are sampled independently. Notice that this definition is consistent with the at-most-once condition on a data generator g.

Remark on Deterministic Functions As a remark, we consider probabilistic and deterministic functions. Let $\mathsf{pFsym} \subseteq \mathsf{Fsym}$ be the set of all *probabilistic function symbols*, each denoting a randomized algorithm that produces a data value using a randomness drawn internally (e.g., a function that returns a number obtained by adding a random number to an input). Let $\mathsf{dFsym} \stackrel{\text{def}}{=} \mathsf{Fsym} \setminus \mathsf{pFsym}$ be the set of all *deterministic function symbols*, denoting deterministic algorithms (e.g., $+$ and $-$).

If $f \in \mathsf{dFsym}$, the interpretation of f is independent of the randomness; i.e., $\xi^r(f) = \xi^{r'}(f)$ for all $r, r' \in \mathcal{I}$, hence $\xi(f)$ maps k data values $\boldsymbol{o}$ to the Dirac distribution $\delta_{\xi^r(f)(\boldsymbol{o})}$, having a single value $\xi^r(f)(\boldsymbol{o})$ with probability 1.

We may relax the at-most-once condition in Section 3 so that deterministic function symbols do not have to satisfy the condition. This is because the interpretation of a deterministic function symbol f is the same in every occurrence of f.

Well-definedness of the Interpretation We next show that the interpretation $[\![\_]\!]_w$ of terms in a world w is unique thanks to the assumption on the strict partial order $\prec$ over the defined causal variables $dom(g_w)$ as follows.

Proposition 3 (Well-definedness of $[\![\_]\!]_w$) *Let u be a term and $w = (\xi, g, m)$ be a world such that g is closed. Then we have $[\![u]\!]_w \in \mathbb{D}\mathcal{O}^k$.*

Proof. Since g is closed, we have $\mathsf{fv}(range(g)) \subseteq dom(g)$. By the definition of $[\![u]\!]_w$ in Section 4, it suffices to show $[\![u]\!]^r_{\xi,g} \in \mathcal{O}^k$ for each $r \in \mathcal{I}$.

By the definition of the possible world w, g is finite and acyclic. Then, $\prec_g$ is the strict partial order over $dom(g)$ defined in Section 3. For a tuple $\boldsymbol{x} = \langle x_1, \ldots, x_k\rangle$ of variables, let $cnt_{\mathsf{v}}(\boldsymbol{x})$ be the number of variables z such that $z \prec_g x_i$ for some $i = 1, \ldots, k$. Since $dom(g)$ is a finite set, $cnt_{\mathsf{v}}(\boldsymbol{x})$ is finite.

Let $r \in \mathcal{I}$. We first show that for any $\boldsymbol{x} = \langle x_1, \ldots, x_k\rangle \in dom(g)^k$, we have $[\![\boldsymbol{x}]\!]^r_{\xi,g} \in \mathcal{O}^k$ by induction on $cnt_{\mathsf{v}}(\boldsymbol{x})$.

– Case $cnt_{\mathsf{v}}(\boldsymbol{x}) = 0$. By definition, there is no variable z such that $z \prec_{g_w} x_i$ for any $i = 1, \ldots, k$; hence $\mathsf{fv}(g(\boldsymbol{x})) = \emptyset$. Since g is closed, for each $i = 1, \ldots, k$, $[\![g(x_i)]\!]_w$ is defined and represented as $[\![c_i]\!]_w$ or $[\![f_i(c_{i1}, \ldots, c_{il})]\!]_w$ for constants $c_i, c_{i1}, \ldots, c_{il}$. Therefore,

$$[\![\boldsymbol{x}]\!]^r_{\xi,g} = [\![\langle g(x_1), \ldots, g(x_k)\rangle]\!]^r_{\xi,g} \in \mathcal{O}^k$$

follows immediately from the definition of ξ^r.

– Case $cnt_{\mathsf{v}}(\boldsymbol{x}) > 0$. Let $\boldsymbol{z} \stackrel{\text{def}}{=} \langle z_1, \ldots, z_l \rangle \subseteq \mathsf{fv}(g(\boldsymbol{x}))$ for $l \geq 1$. From the strict partial order structure of $\prec_g$, we have $cnt_{\mathsf{v}}(\boldsymbol{z}) < cnt_{\mathsf{v}}(\boldsymbol{x})$. By induction hypothesis, $[\![\boldsymbol{z}]\!]^r_{\xi,g} \in \mathcal{O}^l$. Then, by the definition of $\mathsf{fv}(g(\boldsymbol{x}))$, for each $i = 1, 2, \ldots, l$, $g(x_i)$ can be represented as a constant c_i or a causal term $f_i(v_1, \ldots v_l)$ where $v_j \in \mathsf{CVar} \cup \mathsf{Const}$ for each $j = 1, 2, \ldots, l$. In the former case, $[\![g(x_i)]\!]^r_{\xi,g} = [\![c_i]\!]^r_{\xi,g} = \xi^r(c_i) \in \mathcal{O}$ by definition. In the latter case, $[\![g(x_i)]\!]^r_{\xi,g} = [\![f_i(v_1, \ldots v_l)]\!]^r_{\xi,g}$. If $v_i \in \mathsf{Const}$ then $[\![v_i]\!]^r_{\xi,g} \in \mathcal{O}$ is immediate; if $v_j \in \mathsf{CVar}$ then we obtain $v_j \in \mathsf{fv}(g(\boldsymbol{x}))$, and hence $[\![v_j]\!]^r_{\xi,g} \in \mathcal{O}$. Therefore, we conclude:

$$\begin{aligned}[\![f_i(v_1, \ldots, v_l)]\!]^r_{\xi,g} &= \xi^r(f_i)([\![\langle v_1, \ldots, v_l \rangle]\!]^r_{\xi,g}) \\ &= \xi^r(f_i)([\![v_1]\!]^r_{\xi,g}, \cdots, [\![v_l]\!]^r_{\xi,g}) \in \mathcal{O}.\end{aligned}$$

The rest of the proof is immediately by induction on the structures of tuples of terms. □

A.2 Relationships Between Data Generators and Causal Diagrams

We show that each data generator corresponds to a DAG (directed acyclic graph) of the causal model as follows.

Proposition 4 (Acyclicity of G) *Let G be the causal diagram corresponding to a finite and acyclic data generator g. Then G is a finite directed acyclic graph.*

Proof. Let $G = (U, V, E)$. Since g is finite, $dom(g)$ and $range(g)$ are finite. By definition, for each $x \rightarrow y \in E$, we have $x \prec_g y$. Since $\prec_g$ is a strict partial order, so is $\rightarrow$. Therefore, G is a finite directed acyclic graph. □

Next, we show that the interpretation $[\![\_]\!]_w$ defines the joint distribution P_{G_w} of all variables in the causal diagram G_w.

Proposition 5 (Relationship between $[\![\_]\!]_w$ and P_{G_w}) *Let w be a world, G_w be the causal diagram corresponding to the data generator g_w, $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{CVar}^+$, $\boldsymbol{z} \in \mathsf{CVar}^*$, and $\boldsymbol{c} \in \mathsf{dConst}^+$. Then there is a joint distribution P_{G_w} that factorizes according to G_w, and that satisfies:*

1. $[\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_w = P_{G_w}(\boldsymbol{y} \,|\, \boldsymbol{z})$
2. $[\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} = P_{G_w}(\boldsymbol{y} \,|\, do(\boldsymbol{x} = [\![\boldsymbol{c}]\!]_w),\, \boldsymbol{z})$.

Proof. We fix a world w. For brevity, we write $G \stackrel{\text{def}}{=} G_w$. We recall that G is the triple (U, V, E) consisting of $U \stackrel{\text{def}}{=} \mathsf{fc}(range(g_w))$, $V \stackrel{\text{def}}{=} dom(g_w)$, and the set E of structural equations defined by g_w. Let $P_G(V)$ be the probability distribution defined in (2).

We inductively define the set Leaf^k of variables of depth k from leaves and its subset LF^k having a parent variable by:

$$\begin{aligned} V^0 &= V,\\ E^0 &= E,\\ \mathsf{Leaf}^k &= \{v \in V^k \mid \text{ for all } v' \in V^k,\ v \to v' \notin E^k\},\\ \mathsf{LF}^k &= \{v \in \mathsf{Leaf}^k \mid \text{ there is a } v'' \in V^k,\ v'' \to v \notin E^k\},\\ V^{k+1} &= V^k \setminus \mathsf{LF}^k,\\ E^{k+1} &= E^k \cap ((U \cup V^{k+1}) \times V^{k+1}). \end{aligned}$$

This procedure terminates when $\mathsf{LF}^k = \emptyset$. Since g_w is finite and acyclic, the above procedure terminates in a finite number N of steps.

Notice that, by the definition of causal terms, for any $k < N$ and and any $v \in V^k$, $g_w(v)$ is of the form $f(z_1, \ldots, z_l)$ for some $f \in \mathsf{Fsym}$ and $z_1, \ldots, z_l \in \mathsf{CVar}$.

For each $k = 0, 1. \ldots, N$, we consider the restricted diagram $G^k = (U, V^k, E^k)$.

We first claim that $[\![V^k]\!]_w = P_{G^k}(V^k)$ for any $k = 0, 1, \ldots, N$. We prove this by induction on k as follows.

The base case is $k = N$. Since no variable in V^N has a parent variable, there is a $\boldsymbol{c} \in \mathsf{Const}^+$ such that $g_w(V^N) = \boldsymbol{c}$. Hence,

$$P_G(V^N) = [\![\boldsymbol{c}]\!]_w = [\![g_w(V^N)]\!]_w = [\![V^N]\!]_w.$$

Next, we consider the case $k < N$. Since the above procedure terminates exactly at N steps, we obtain $\mathsf{LF}^k \neq \emptyset$. By applying the induction hypothesis (the case of $k+1$), we obtain $P_{G^{k+1}}(V^k \setminus \mathsf{LF}^k) = P_{G^{k+1}}(V^{k+1}) = [\![V^{k+1}]\!]_w = [\![V^k \setminus \mathsf{LF}^k]\!]_w$. Since each causal term is of depth at most 1 by definition, the set $pa_{G^k}(v)$ of all parents of a variable v in the causal diagram G^k is given by $\mathsf{fv}(g_w(v))$. Hence,

$$\begin{aligned} P_{G^k}(V^k) &= P_{G^{k+1}}(\mathsf{LF}^k) \cdot P_{G^{k+1}}(V^k \setminus \mathsf{LF}^k)\\ &= \Big(\prod_{v \in \mathsf{LF}^k} P_{G^k}(v \mid pa_{G^k}(v))\Big) \cdot P_{G^{k+1}}(V^k \setminus \mathsf{LF}^k)\\ &= \Big(\prod_{v \in \mathsf{LF}^k} P_{G^k}(v \mid pa_{G^k}(v))\Big) \cdot [\![V^k \setminus \mathsf{LF}^k]\!]_w\\ &\qquad \text{(by induction hypothesis)}\\ &= [\![V^k]\!]_w.\\ &\qquad \text{(since each causal term is of depth at most 1)} \end{aligned}$$

Therefore, we conclude:

$$P_G(V) = P_{G^0}(V^0) = [\![V^0]\!]_w = [\![V]\!]_w.$$

Now the first equation in the proposition is obtained as follows.

$$\begin{aligned}
[\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_w &= \tfrac{[\![\boldsymbol{y}::\boldsymbol{z}]\!]_w}{[\![\boldsymbol{z}]\!]_w} \\
&= \tfrac{([\![V]\!]_w)|_{\boldsymbol{y}::\boldsymbol{z}}}{([\![V]\!]_w)|_{\boldsymbol{z}}} \\
&= \tfrac{(P_{G_w}(V))|_{\boldsymbol{y}::\boldsymbol{z}}}{(P_{G_w}(V))|_{\boldsymbol{z}}} \\
&= \tfrac{P_{G_w}(\boldsymbol{y}::\boldsymbol{z})}{P_{G_w}(\boldsymbol{z})} \\
&= P_{G_w}(\boldsymbol{y} \mid \boldsymbol{z}).
\end{aligned}$$

Similarly, the second equation in the proposition is obtained as follows. Let G'_w be the causal diagram obtained by an intervention $\boldsymbol{x} := [\![\boldsymbol{c}]\!]_w$ in G_w.

$$\begin{aligned}
[\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil} &= \tfrac{[\![\boldsymbol{y}::\boldsymbol{z}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil}}{[\![\boldsymbol{z}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil}} \\
&= \tfrac{([\![V]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil})|_{\boldsymbol{y}::\boldsymbol{z}}}{([\![V]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil})|_{\boldsymbol{z}}} \\
&= \tfrac{(P_{G'_w}(V))|_{\boldsymbol{y}::\boldsymbol{z}}}{(P_{G'_w}(V))|_{\boldsymbol{z}}} \\
&= \tfrac{P_{G_w}(\boldsymbol{y}::\boldsymbol{z} \mid do(\boldsymbol{x} = [\![\boldsymbol{c}]\!]_w))}{P_{G_w}(\boldsymbol{z} \mid do(\boldsymbol{x} = [\![\boldsymbol{c}]\!]_w))} \\
&= P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x} = [\![\boldsymbol{c}]\!]_w),\, \boldsymbol{z}).
\end{aligned}$$

□

A.3 Properties on Memories

We present basic properties on memories. Intuitively, we show that:

(i) the same formulas $\varphi_{\boldsymbol{y}}$ with free variables $\boldsymbol{y}$ are satisfied in any worlds w and w' having the same memory on $\boldsymbol{y}$;
(ii) an eager and a lazy interventions to $\boldsymbol{x}$ result in the same distribution of the variables $\boldsymbol{y}$ disjoint from $\boldsymbol{x}$;
(iii) disjoint worlds w and w' have data generators g_w and $g_{w'}$ with disjoint domains.

Proposition 6 (Properties on m_w) *Let w and w' be worlds, $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{Var}^+$, $\boldsymbol{c} \in \mathsf{dConst}^+$, and $\varphi_{\boldsymbol{y}} \in \mathsf{Fml}$ with $\mathsf{fv}(\varphi_{\boldsymbol{y}}) = \boldsymbol{y}$.*

(i) If $m_w(\boldsymbol{y}) = m_{w'}(\boldsymbol{y}) \neq \bot$, then $w \models \varphi_{\boldsymbol{y}}$ iff $w' \models \varphi_{\boldsymbol{y}}$.
(ii) If $\boldsymbol{x} \cap \boldsymbol{y} = \emptyset$, then $m_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil}(\boldsymbol{y}) = m_{w\lfloor \boldsymbol{c}/\boldsymbol{x}\rfloor}(\boldsymbol{y})$.
(iii) If $\mathsf{Var}(w) \cap \mathsf{Var}(w') = \emptyset$ (namely, $dom(m_w) \cap dom(m_{w'}) = \emptyset$), then $dom(g_w) \cap dom(g_{w'}) = \emptyset$.

Proof. (i) Assume that $m_w(\boldsymbol{y}) = m_{w'}(\boldsymbol{y}) \neq \bot$. Then we prove $w \models \varphi_{\boldsymbol{y}}$ iff $w' \models \varphi_{\boldsymbol{y}}$ by induction on $\varphi_{\boldsymbol{y}}$.

If $\varphi_{\boldsymbol{y}} \stackrel{\text{def}}{=} \eta(\boldsymbol{y})$ for $\eta \in \mathsf{Psym}$, then:

$$\begin{aligned} & w \models \eta(\boldsymbol{y}) \\ \text{iff}\ \ & m_w(\boldsymbol{y}) \in \mathcal{V}(\eta) \\ \text{iff}\ \ & m_{w'}(\boldsymbol{y}) \in \mathcal{V}(\eta) && (\text{by } m_w(\boldsymbol{y}) = m_{w'}(\boldsymbol{y}) \neq \bot) \\ \text{iff}\ \ & w' \models \eta(\boldsymbol{y}). \end{aligned}$$

The other cases are straightforward by definitions.

(ii) Assume $\boldsymbol{x} \cap \boldsymbol{y} = \emptyset$. By definition, we obtain:

$$\begin{aligned} m_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil}(\boldsymbol{y}) &= [\![g_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil}(\boldsymbol{y})]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil} \\ &= [\![g_w(\boldsymbol{y})]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil} && (\text{by } \boldsymbol{x} \cap \boldsymbol{y} = \emptyset) \\ &= [\![g_w(\boldsymbol{y})[\boldsymbol{x} \mapsto \boldsymbol{c}]]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil} \\ &= [\![g_w(\boldsymbol{y})[\boldsymbol{x} \mapsto \boldsymbol{c}]]\!]_{w\lfloor \boldsymbol{c}/\boldsymbol{x}\rfloor} && (\text{by } \boldsymbol{x} \cap \mathsf{fv}(g_w(\boldsymbol{y})[\boldsymbol{x} \mapsto \boldsymbol{c}]) = \emptyset) \\ &= [\![g_{w\lfloor \boldsymbol{c}/\boldsymbol{x}\rfloor}(\boldsymbol{y})]\!]_{w\lfloor \boldsymbol{c}/\boldsymbol{x}\rfloor} \\ &= m_{w\lfloor \boldsymbol{c}/\boldsymbol{x}\rfloor}(\boldsymbol{y}). \end{aligned}$$

(iii) By the definition of possible worlds, we have $dom(g_w) \subseteq dom(m_w)$ and $dom(g_{w'}) \subseteq dom(m_{w'})$. Therefore, we obtain $dom(g_w) \cap dom(g_{w'}) \subseteq dom(m_w) \cap dom(m_{w'}) = \emptyset$.

□

A.4 Causal Predicates

Next, we present more details on causal predicates. Among the causal predicates listed below, our deduction system $\mathbf{AX^{CP}}$ requires only *dsep*, *nanc*, and *allnanc*. For the sake of convenience, we can use *pa*, but it is sufficient for us to derive the formulas equipped with *pa* from those with *nanc*. Thus, we do not deal with axioms of the other predicates in this paper.

We show a list of causal predicates as follows.

- $dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ $\boldsymbol{x}$ and $\boldsymbol{y}$ are *d*-separated by $\boldsymbol{z}$;
- $pa(\boldsymbol{x}, \boldsymbol{y})$ $\boldsymbol{x}$ is the set of all parents of variables in $\boldsymbol{y}$;
- $npa(\boldsymbol{x}, \boldsymbol{y})$ $\boldsymbol{x}$ is a set of non-parents of variables in $\boldsymbol{y}$;
- $anc(\boldsymbol{x}, \boldsymbol{y})$ $\boldsymbol{x}$ is the set of all ancestors of variables in $\boldsymbol{y}$;
- $nanc(\boldsymbol{x}, \boldsymbol{y})$ $\boldsymbol{x}$ is a set of non-ancestors of variables in $\boldsymbol{y}$;
- $allnanc(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ $\boldsymbol{x}$ is the set of all variables in $\boldsymbol{y}$ that are not ancestors of any variables in $\boldsymbol{z}$.

These causal predicates are interpreted using a data generator g_w in a world w as follows.

Definition 1 (Semantics of *dsep*, *pa*, *npa*, *anc*, *nanc*, *allnanc*). Let w be a world, and G_w be the causal diagram corresponding to g_w. Let $\mathtt{PA}(\boldsymbol{y})$ be the set of all parent variables of $\boldsymbol{y}$:

$$\mathtt{PA}(\boldsymbol{y}) = \{x \in \mathsf{Var}(w) \mid x \rightarrow y',\ y' \in \boldsymbol{y}\}.$$

Let $\mathtt{ANC}(\boldsymbol{y})$ is the set of all ancestors variables of $\boldsymbol{y}$:

$$\mathtt{ANC}(\boldsymbol{y}) = \{x \in \mathsf{Var}(w) \mid x \twoheadrightarrow^{+} y',\ y' \in \boldsymbol{y}\}$$

where $\twoheadrightarrow^{+}$ is the transitive closure of $\twoheadrightarrow$. The interpretations of the causal predicates *dsep*, *pa*, *npa*, *anc*, *nanc*, *allnanc* are given as follows:

$$\begin{aligned}
w \models \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) &\text{ iff } \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ are } d\text{-separated by } \boldsymbol{z} \text{ in } G_w \\
w \models \mathit{pa}(\boldsymbol{x}, \boldsymbol{y}) &\text{ iff } \boldsymbol{x} = \mathtt{PA}(\boldsymbol{y}) \text{ and } \boldsymbol{x} \cap \boldsymbol{y} = \emptyset \\
w \models \mathit{npa}(\boldsymbol{x}, \boldsymbol{y}) &\text{ iff } \boldsymbol{x} \cap \mathtt{PA}(\boldsymbol{y}) = \emptyset \text{ and } \boldsymbol{x} \cap \boldsymbol{y} = \emptyset \\
w \models \mathit{anc}(\boldsymbol{x}, \boldsymbol{y}) &\text{ iff } \boldsymbol{x} = \mathtt{ANC}(\boldsymbol{y}) \text{ and } \boldsymbol{x} \cap \boldsymbol{y} = \emptyset \\
w \models \mathit{nanc}(\boldsymbol{x}, \boldsymbol{y}) &\text{ iff } \boldsymbol{x} \cap \mathtt{ANC}(\boldsymbol{y}) = \emptyset \text{ and } \boldsymbol{x} \cap \boldsymbol{y} = \emptyset \\
w \models \mathit{allnanc}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) &\text{ iff } \boldsymbol{x} = \boldsymbol{y} \setminus \mathtt{ANC}(\boldsymbol{z}),
\end{aligned}$$

where we recall the notion of d-separation in Appendix A.6.

Proposition 7 (Relationships among causal predicates) *The causal predicates pa, npa, anc, and nanc satisfy the relationships:*

1. $\models \mathit{pa}(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \mathit{anc}(\boldsymbol{x}, \boldsymbol{y})$.
2. $\models \mathit{anc}(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \mathit{nanc}(\boldsymbol{y}, \boldsymbol{x})$.
3. $\models \mathit{nanc}(\boldsymbol{y}, \boldsymbol{x}) \rightarrow \mathit{npa}(\boldsymbol{y}, \boldsymbol{x})$.

Proof. These claim are straightforward from Definition 1. □

A.5 Causal Effect

We show that the causal effect can be expressed using a StaCL formula as follows.

Proposition 1 (Causal effect) *Let w be a world, $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{Var}(w)^{+}$ be disjoint, $\boldsymbol{c} \in \mathsf{dConst}^{+}$, $\boldsymbol{c}' \in \mathsf{Const}^{+}$, and $f \in \mathsf{Fsym}$. Then:*

(i) $w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (\boldsymbol{c}' = \boldsymbol{y})$ *iff there is a distribution P_{G_w} that is factorized according to G_w and satisfies $P_{G_w}(\boldsymbol{y} \mid \mathit{do}(\boldsymbol{x} = \boldsymbol{c})) = [\![\boldsymbol{c}']\!]_w$.*

(ii) $w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (f = \boldsymbol{y}|_{\boldsymbol{z}})$ *iff there is a distribution P_{G_w} that is factorized according to G_w and satisfies $P_{G_w}(\boldsymbol{y} \mid \mathit{do}(\boldsymbol{x} = \boldsymbol{c}), \boldsymbol{z}) = [\![f]\!]_w$.*

Proof. We show the first claim as follows. By Proposition 5, there is a joint distribution P_{G_w} that is factorized according to G_w and that satisfies $[\![\boldsymbol{y}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} = P_{G_w}(\boldsymbol{y} \mid \mathit{do}(\boldsymbol{x} = [\![\boldsymbol{c}]\!]_w))$. Thus, we obtain:

$$\begin{aligned}
& w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (\boldsymbol{c}' = \boldsymbol{y}) \\
\text{iff } & w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \models \boldsymbol{c}' = \boldsymbol{y} \\
\text{iff } & [\![\boldsymbol{y}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} = [\![\boldsymbol{c}']\!]_w \\
\text{iff } & P_{G_w}(\boldsymbol{y} \mid \mathit{do}(\boldsymbol{x} = [\![\boldsymbol{c}]\!]_w)) = [\![\boldsymbol{c}']\!]_w.
\end{aligned}$$

Analogously, the second claim is obtained as follows. By Proposition 5, there is a joint distribution P_{G_w} that is factorized according to G_w and that satisfies $[\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil} = P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x}=[\![\boldsymbol{c}]\!]_w),\, \boldsymbol{z})$. Thus, we obtain:

$$
\begin{aligned}
& w \models \lceil \boldsymbol{c}/\boldsymbol{x}\rceil(\boldsymbol{c}' = \boldsymbol{y}|_{\boldsymbol{z}}) \\
\text{iff}\quad & w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil \models \boldsymbol{c}' = \boldsymbol{y}|_{\boldsymbol{z}} \\
\text{iff}\quad & [\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x}\rceil} = [\![\boldsymbol{c}']\!]_w \\
\text{iff}\quad & P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x}=\boldsymbol{c}), \boldsymbol{z}) = [\![\boldsymbol{c}']\!]_w.
\end{aligned}
$$

□

A.6 Details on Causal Diagrams

Next, we recall the notion of d-separation [35] as follows.

Definition 2 (d-separation). Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ be disjoint sets of variables, and $\mathtt{ANC}_*(\boldsymbol{z}) \stackrel{\text{def}}{=} \boldsymbol{z} \cup \mathtt{ANC}(\boldsymbol{z})$ be the union of $\boldsymbol{z}$ and the set of $\boldsymbol{z}$'s all ancestors. An undirected path p is said to be *d-separated* by $\boldsymbol{z}$ if it satisfies one of the following conditions:

(a) p has a chain $v' \rightarrow v \rightarrow v''$ s.t. $v \in \boldsymbol{z}$.
(b) p has a fork $v' \leftarrow v \rightarrow v''$ s.t. $v \in \boldsymbol{z}$.
(c) p has a collider $v' \rightarrow v \leftarrow v''$ s.t. $v \notin \mathtt{ANC}_*(\boldsymbol{z})$.

$\boldsymbol{x}$ and $\boldsymbol{y}$ are *d-separated* by $\boldsymbol{z}$ if all undirected paths between variables in $\boldsymbol{x}$ and in $\boldsymbol{z}$ are d-separated by $\boldsymbol{z}$.

We also recall the notion of back-door path as follows.

Definition 3 (Back-door path). For variables x and y, a *back-door path* from x to y in a causal diagram G is an arbitrary undirected path between x and y in G that starts with an arrow pointing to x (i.e., an undirected path of the form $x \leftarrow v \cdots y$). For tuples of variables $\boldsymbol{x}$ and $\boldsymbol{y}$, a *back-door path* from $\boldsymbol{x}$ to $\boldsymbol{y}$ is an arbitrary back-door path from $x \in \boldsymbol{x}$ to $y \in \boldsymbol{y}$.

We remark on the relationships between back-door paths and two kinds of interventions as follows.

Remark 1. An eager intervention $\lceil \boldsymbol{c}/\boldsymbol{x}\rceil$ can remove *all back-door paths from $\boldsymbol{x}$ to $\boldsymbol{y}$*, because all of these paths have arrows pointing to $\boldsymbol{x}$ and $\lceil \boldsymbol{c}/\boldsymbol{x}\rceil$ removes all such arrows.

In contrast, a lazy intervention $\lfloor \boldsymbol{c}/\boldsymbol{x}\rfloor$ can remove all undirected paths between x and y *except for all back-door paths from x to y*, because $\lfloor \boldsymbol{c}/\boldsymbol{x}\rfloor$ removes all arrows emerging from $\boldsymbol{x}$ while keeping all arrows pointing to $\boldsymbol{x}$. Thus, $\lfloor \boldsymbol{c}/\boldsymbol{x}\rfloor\, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ represents that all back-door paths from $\boldsymbol{x}$ to $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$.

These relationships are used to reason about the causality, e.g., when we derive the second rule of Pearl's do-calculus using our StaCL (Proposition 2).

Now we recall the back-door criteria and the back-door adjustment in Pearl's causal model.

Definition 4 (Back-door criterion). For two sets $\boldsymbol{x}$ and $\boldsymbol{y}$ of variables, a set $\boldsymbol{z}$ of variables satisfies the *back-door criterion* in a causal diagram G if (i) no variable in $\boldsymbol{z}$ is a descendent of an element of $\boldsymbol{x}$ in G and (ii) all back-door paths from $\boldsymbol{x}$ to $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$ in G.

The back-door criterion is expressed as the following StaCL formula:

$$nanc(\boldsymbol{x}, \boldsymbol{z}) \wedge \lfloor c/\boldsymbol{x} \rfloor \, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}).$$

When $\boldsymbol{z}$ satisfies the back-door criterion in a causal diagram G, then the causal effect of $\boldsymbol{x}$ on $\boldsymbol{y}$ is given by:

$$P_G(\boldsymbol{y} \mid do(\boldsymbol{x})) = \sum_{\boldsymbol{z}} P_G(\boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{x}) \, P_G(\boldsymbol{z}).$$

In Fig. 7 in Section 7, we show a derivation tree for the correctness of computing the causal effect using the backdoor adjustment.

B Proof for the Soundness of AX

We show that the deductive system **AX** of StaCL is sound w.r.t. the Kripke semantics for statistical causality (Theorem 1).

We first remark that **AX** satisfies the deduction theorem.

Proposition 8 (Deduction) *Let $\Gamma \subseteq \mathsf{Fml}$, and $\varphi_1, \varphi_2 \in \mathsf{Fml}$. Then $\Gamma \vdash_g \varphi_1 \rightarrow \varphi_2$ iff $\Gamma, \varphi_1 \vdash_g \varphi_2$.*

Proof. The direction from left to right is straightforward by the application of MP. The other direction is shown as usual by induction on the derivation. □

We prove the soundness of **AX** as follows. We show the validity of the axioms for basic constructs (Appendix B.1), for eager interventions (Appendix B.2), for lazy interventions (Appendix B.3), and for the exchanges of eager/lazy interventions (Appendix B.4).

B.1 Validity of the Basic Axioms

Here are the basic axioms of **AX** without interventions.

PT	$\vdash_g \varphi$ for a propositional tautology φ
MP	$\varphi_1, \varphi_1 \rightarrow \varphi_2 \vdash_g \varphi_2$
EQ1	$\vdash_g \boldsymbol{x} = \boldsymbol{x}$
EQ2	$\vdash_g \boldsymbol{x} = \boldsymbol{y} \rightarrow (\varphi_1 \rightarrow \varphi_2)$ where φ_2 is the formula obtained by replacing any number of occurrences of $\boldsymbol{x}$ in φ_1 with $\boldsymbol{y}$
EQC	$\vdash_g c^{(g,\boldsymbol{x})} = \boldsymbol{x}$
EQF	$\vdash_g f^{(g,\boldsymbol{y}\|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}})} = \boldsymbol{y}\|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}}$
PD	$\vdash_g (pos(\boldsymbol{x}) \wedge c_0{=}\boldsymbol{x} \wedge f{=}\boldsymbol{y}\|_{\boldsymbol{x}} \wedge c_1{=}\boldsymbol{x} :: \boldsymbol{y}) \rightarrow c_1{=}f(c_0)$
MPD	$\vdash_g \boldsymbol{x_1} \downarrow_{\boldsymbol{x_2}} = \boldsymbol{x_2}$ if $\boldsymbol{x_2} \subseteq \boldsymbol{x_1}$

The validity of the rules PT, MP, EQ1, EQ2, is straightforward. The validity of EQ$_{\text{C}}$ and EQ$_{\text{F}}$ is by the definition of the interpretation of the constants and function symbols introduced for the purpose of reasoning (Section 4):

$$\xi^r(c^{(g,\boldsymbol{x})}) = [\![\boldsymbol{x}]\!]^r_{\xi,g}$$
$$\xi^r(f^{(g,\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}})}) = [\![\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}}]\!]^r_{\xi,g}.$$

We show the validity of PD and MPD as follows.

Proposition 9 (Probability distributions) *Let* $\boldsymbol{x}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{y} \in \mathsf{CVar}^+$, $c_0, c_1 \in$ *Const, and* $f \in$ *Fsym.*

(i) PD $\vdash_g (pos(\boldsymbol{x}) \wedge c_0{=}\boldsymbol{x} \wedge f{=}\boldsymbol{y}|_{\boldsymbol{x}} \wedge c_1{=}\boldsymbol{x} :: \boldsymbol{y})$
$\rightarrow c_1{=}f(c_0)$.

(ii) MPD $\vdash_g \boldsymbol{x_1} \downarrow_{\boldsymbol{x_2}} = \boldsymbol{x_2}$ *if* $\boldsymbol{x_2} \subseteq \boldsymbol{x_1}$.

Proof. Let $w = (g_w, \xi_w, m_w)$ be a world such that $\boldsymbol{x}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{y} \in \mathsf{Var}(w)^+$.

(i) We show the validity of PD as follows. Suppose that $w \models pos(\boldsymbol{x}) \wedge c_0{=}\boldsymbol{x} \wedge f{=}\boldsymbol{y}|_{\boldsymbol{x}} \wedge c_1{=}\boldsymbol{x} :: \boldsymbol{y}$. Then we have $[\![\boldsymbol{x}]\!]_w(o'_{\boldsymbol{x}}) > 0$ for all $o'_{\boldsymbol{x}} \in \mathcal{O}^{|\boldsymbol{x}|}$, $[\![c_0]\!]_w = [\![\boldsymbol{x}]\!]_w$, $[\![c_1]\!]_w = [\![\boldsymbol{x} :: \boldsymbol{y}]\!]_w$, and $[\![f]\!]_w = [\![\boldsymbol{y}|_{\boldsymbol{x}}]\!]_w$. Since $[\![\boldsymbol{x}]\!]_w(o'_{\boldsymbol{x}}) > 0$, we have:

$$([\![\boldsymbol{y}|_{\boldsymbol{x}}]\!]_w(o'_{\boldsymbol{x}})) = \textstyle\sum_{o'_{\boldsymbol{y}}} \frac{[\![\boldsymbol{x} :: \boldsymbol{y}]\!]_w(o'_{\boldsymbol{x}}, o'_{\boldsymbol{y}})}{[\![\boldsymbol{x}]\!]_w(o'_{\boldsymbol{x}})} \cdot \delta_{(o'_{\boldsymbol{x}}, o'_{\boldsymbol{y}})}.$$

Thus, for each $o_{\boldsymbol{x}} \in \mathcal{O}^{|\boldsymbol{x}|}$ and $o_{\boldsymbol{y}} \in \mathcal{O}^{|\boldsymbol{y}|}$, we have:

$$\begin{aligned}
&[\![f(c_0)]\!]_w(o_{\boldsymbol{x}}, o_{\boldsymbol{y}}) \\
&= ([\![f]\!]_w [\![c_0]\!]_w)(o_{\boldsymbol{x}}, o_{\boldsymbol{y}}) \\
&= ([\![\boldsymbol{y}|_{\boldsymbol{x}}]\!]_w [\![\boldsymbol{x}]\!]_w)(o_{\boldsymbol{x}}, o_{\boldsymbol{y}}) \\
&= \Big(\sum_{o'_{\boldsymbol{x}}} \sum_{o'_{\boldsymbol{y}}} \tfrac{[\![\boldsymbol{x} :: \boldsymbol{y}]\!]_w(o'_{\boldsymbol{x}}, o'_{\boldsymbol{y}})}{[\![\boldsymbol{x}]\!]_w(o'_{\boldsymbol{x}})} [\![\boldsymbol{x}]\!]_w(o'_{\boldsymbol{x}}) \cdot \delta_{(o'_{\boldsymbol{x}}, o'_{\boldsymbol{y}})} \Big)(o_{\boldsymbol{x}}, o_{\boldsymbol{y}}) \\
&= [\![\boldsymbol{x} :: \boldsymbol{y}]\!]_w(o_{\boldsymbol{x}}, o_{\boldsymbol{y}}) \\
&= [\![c_1]\!]_w(o_{\boldsymbol{x}}, o_{\boldsymbol{y}}).
\end{aligned}$$

Therefore, we obtain $w \models c_1{=}f(c_0)$.

(ii) We show the validity of MPD as follows. Let $\otimes$ be the product of probability distributions of data values. Let $\boldsymbol{x_1} = \{x_1, \ldots, x_k\}$. Assume that $\emptyset \neq \boldsymbol{x_2} \subseteq \boldsymbol{x_1}$. Then we may write $\boldsymbol{x_2} = \{x_{l(1)}, \ldots, x_{l(k')}\}$ for some $1 \leq k' \leq k$ and monotone increasing function $l\colon \{1, \ldots, k'\} \to \{1, \ldots, k\}$. Using this, we obtain:

$$\begin{aligned}
&[\![\boldsymbol{x_1} \downarrow_{\boldsymbol{x_2}}]\!]_w \\
&= [\![\downarrow_{\boldsymbol{x_2}}]\!]_w [\![\boldsymbol{x_1}]\!]_w \\
&= ((o_1, \ldots, o_k) \mapsto (o_{l(1)}, \ldots, o_{l(k')}))[\![\langle x_1, \ldots, x_k\rangle]\!]_w \\
&= ((o_1, \ldots, o_k) \mapsto (o_{l(1)}, \ldots, o_{l(k')}))[\![x_1]\!]_w \otimes \cdots \otimes [\![x_k]\!]_w \\
&= [\![x_{l(1)}]\!]_w \otimes \cdots \otimes [\![x_{l(k')}]\!]_w \\
&= [\![\langle x_{l(1)}, \ldots, x_{l(k')}\rangle]\!]_w = [\![\boldsymbol{x_2}]\!]_w.
\end{aligned}$$

Therefore, we obtain $w \models \boldsymbol{x_1} \downarrow_{\boldsymbol{x_2}} = \boldsymbol{x_2}$ if $\boldsymbol{x_2} \subseteq \boldsymbol{x_1}$. □

B.2 Validity of the Axioms for Eager Interventions

Here are the axioms of **AX** with the eager interventions $\lceil\cdot\rceil$.

DG$_{\text{EI}}$	$\vdash_g \lceil\boldsymbol{c}/\boldsymbol{x}\rceil\varphi$ iff $\vdash_{g\lceil\boldsymbol{c}/\boldsymbol{x}\rceil} \varphi$
Effect$_{\text{EI}}$	$\vdash_g \lceil\boldsymbol{c}/\boldsymbol{x}\rceil(\boldsymbol{x} = \boldsymbol{c})$
Eq$_{\text{EI}}$	$\vdash_g \boldsymbol{u_1} = \boldsymbol{u_2} \leftrightarrow \lceil\boldsymbol{c}/\boldsymbol{x}\rceil(\boldsymbol{u_1} = \boldsymbol{u_2})$ if $\mathsf{fv}(\boldsymbol{u_1}) = \mathsf{fv}(\boldsymbol{u_2}) = \emptyset$
Split$_{\text{EI}}$	$\vdash_g \lceil\boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2}\rceil\varphi \to \lceil\boldsymbol{c_1}/\boldsymbol{x_1}\rceil\lceil\boldsymbol{c_2}/\boldsymbol{x_2}\rceil\varphi$
Simul$_{\text{EI}}$	$\vdash_g \lceil\boldsymbol{c_1}/\boldsymbol{x_1}\rceil\lceil\boldsymbol{c_2}/\boldsymbol{x_2}\rceil\varphi \to \lceil\boldsymbol{c'_1}/\boldsymbol{x'_1}, \boldsymbol{c_2}/\boldsymbol{x_2}\rceil\varphi$ for $\boldsymbol{x'_1} = \boldsymbol{x_1} \setminus \boldsymbol{x_2}$ and $\boldsymbol{c'_1} = \boldsymbol{c_1} \setminus \boldsymbol{c_2}$
Rpt$_{\text{EI}}$	$\vdash_g \lceil\boldsymbol{c}/\boldsymbol{x}\rceil\varphi \to \lceil\boldsymbol{c}/\boldsymbol{x}\rceil\lceil\boldsymbol{c}/\boldsymbol{x}\rceil\varphi$
Cmp$_{\text{EI}}$	$\vdash_g \big(\lceil\boldsymbol{c_1}/\boldsymbol{x_1}\rceil(\boldsymbol{x_2} = \boldsymbol{c_2}) \land \lceil\boldsymbol{c_1}/\boldsymbol{x_1}\rceil(\boldsymbol{x_3} = \boldsymbol{u})\big) \to \lceil\boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2}\rceil(\boldsymbol{x_3} = \boldsymbol{u})$
Distr$_{\text{EI}}{}^{\neg}$	$\vdash_g (\lceil\boldsymbol{c}/\boldsymbol{x}\rceil\neg\varphi) \leftrightarrow (\neg\lceil\boldsymbol{c}/\boldsymbol{x}\rceil\varphi)$
Distr$_{\text{EI}}{}^{\land}$	$\vdash_g (\lceil\boldsymbol{c}/\boldsymbol{x}\rceil(\varphi_1 \land \varphi_2)) \leftrightarrow (\lceil\boldsymbol{c}/\boldsymbol{x}\rceil\varphi_1 \land \lceil\boldsymbol{c}/\boldsymbol{x}\rceil\varphi_2)$

Next, we show basic laws of eager interventions as follows.

Proposition 10 (Basic laws of $\lceil\cdot\rceil$) *Let $\boldsymbol{x}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{CVar}^+$ be disjoint, $\boldsymbol{c}, \boldsymbol{c_1}, \boldsymbol{c_2} \in \mathsf{dConst}^+$, $\boldsymbol{u}, \boldsymbol{u_1}, \boldsymbol{u_2} \in \mathsf{Term}^+$, and $\varphi \in \mathsf{Fml}$.*

1. DG$_{\text{EI}}$ $\models_g \lceil\boldsymbol{c}/\boldsymbol{x}\rceil\varphi$ *iff* $\models_{g\lceil\boldsymbol{c}/\boldsymbol{x}\rceil} \varphi$.
2. Effect$_{\text{EI}}$ $\models \lceil\boldsymbol{c}/\boldsymbol{x}\rceil(\boldsymbol{x} = \boldsymbol{c})$.
3. Eq$_{\text{EI}}$ $\models \boldsymbol{u_1} = \boldsymbol{u_2} \leftrightarrow \lceil\boldsymbol{c}/\boldsymbol{x}\rceil(\boldsymbol{u_1} = \boldsymbol{u_2})$ *if* $\mathsf{fv}(\boldsymbol{u_1}) = \mathsf{fv}(\boldsymbol{u_2}) = \emptyset$.

4. SPLIT$_{\mathrm{EI}}$ $\models \lceil {c_1}/{x_1},\, {c_2}/{x_2} \rceil \varphi \rightarrow \lceil {c_1}/{x_1} \rceil \lceil {c_2}/{x_2} \rceil \varphi$.
5. SIMUL$_{\mathrm{EI}}$ $\models \lceil {c_1}/{x_1} \rceil \lceil {c_2}/{x_2} \rceil \varphi \rightarrow \lceil {c'_1}/{x'_1},\, {c_2}/{x_2} \rceil \varphi$ *for* $x'_1 = x_1 \setminus x_2$ *and* $c'_1 = c_1 \setminus c_2$.
6. RPT$_{\mathrm{EI}}$ $\models \lceil {c}/{x} \rceil \varphi \rightarrow \lceil {c}/{x} \rceil \lceil {c}/{x} \rceil \varphi$.
7. CMP$_{\mathrm{EI}}$ $\models \big(\lceil {c_1}/{x_1} \rceil (x_2 = c_2) \wedge \lceil {c_1}/{x_1} \rceil (x_3 = u) \big) \rightarrow \lceil {c_1}/{x_1}, {c_2}/{x_2} \rceil (x_3 = u)$.

Proof. Let $w = (g_w, \xi_w, m_w)$ be a world such that $x, x_1, x_2, x_3, y, z \in \mathsf{Var}(w)^+$.

1. Assume that $\models_{g\lceil c/x \rceil} \varphi$. Then for any world w' having the data generator g, we have $w'\lceil c/x \rceil \models \varphi$, hence $w' \models \lceil c/x \rceil \varphi$. Therefore, $\models_g \lceil c/x \rceil \varphi$. The other direction is also shown analogously.
2. By the definition of an eagerly intervened world in Section 4, we have $g_{w\lceil c/x \rceil}(x) = c$. By $\xi_w = \xi_{w\lceil c/x \rceil}$, $[\![c]\!]_w = [\![c]\!]_{w\lceil c/x \rceil}$.
   Then $[\![x]\!]_{w\lceil c/x \rceil} = [\![g_{w\lceil c/x \rceil}(x)]\!]_{w\lceil c/x \rceil} = [\![c]\!]_{w\lceil c/x \rceil}$. Hence $w\lceil c/x \rceil \models x = c$. Therefore, $w \models \lceil c/x \rceil (x = c)$.
3. Assume that $\mathsf{fv}(u_1) = \mathsf{fv}(u_2) = \emptyset$. Then for each $i = 1, 2$, $[\![u_i]\!]_w = [\![u_i]\!]_{w\lceil c/x \rceil}$. Hence, $\models u_1 = u_2 \leftrightarrow \lceil c/x \rceil (u_1 = u_2)$.
4. The proof is straightforward from the definition.
5. The proof is straightforward from the definition.
6. The proof is straightforward from the definition.
7. Assume that $w \models \lceil {c_1}/{x_1} \rceil (x_2 = c_2) \wedge \lceil {c_1}/{x_1} \rceil (x_3 = u)$. Then $g_{w\lceil c_1/x_1 \rceil}(x_2) = c_2$ and $g_{w\lceil c_1/x_1 \rceil}(x_3) = u$. Let $w' = w\lceil {c_1}/{x_1} \rceil$. Thus,

$$
\begin{aligned}
& [\![g_{w\lceil c_1/x_1, c_2/x_2 \rceil}(x_3)]\!]_{w\lceil c_1/x_1, c_2/x_2 \rceil} \\
&= [\![g_{w'\lceil c_2/x_2 \rceil}(x_3)]\!]_{w'\lceil c_2/x_2 \rceil} \\
&= [\![g_{w'}(x_3)]\!]_{w'\lceil c_2/x_2 \rceil} \\
&= [\![u]\!]_{w\lceil c_1/x_1, c_2/x_2 \rceil}.
\end{aligned}
$$

   Therefore, $w \models \lceil {c_1}/{x_1}, {c_2}/{x_2} \rceil (x_3 = u)$. □

The eager intervention operator $\lceil \cdot \rceil$ is distributive w.r.t. logical connectives.

Proposition 11 (Distributive laws of $\lceil \cdot \rceil$) *Let $x \in \mathit{CVar}^+$, $c \in \mathit{dConst}^+$, and $\varphi, \varphi' \in \mathit{Fml}$.*

(i) DISTR$_{\mathrm{EI}}^{\neg}$ $\models \lceil c/x \rceil \neg\varphi \leftrightarrow \neg \lceil c/x \rceil \varphi$.
(ii) DISTR$_{\mathrm{EI}}^{\rightarrow}$ $\models \lceil c/x \rceil (\varphi \rightarrow \varphi') \leftrightarrow \big(\lceil c/x \rceil \varphi \rightarrow \lceil c/x \rceil \varphi' \big)$.

Similarly, the eager intervention operator $\lceil \cdot \rceil$ is distributive w.r.t. $\vee$ and $\wedge$.

Proof. Let w be a world such that $x \in \mathsf{Var}(w)^+$.

(i)

$$
\begin{aligned}
w \models \lceil c/x \rceil \neg\varphi \quad &\text{iff} \quad w\lceil c/x \rceil \models \neg\varphi \\
&\text{iff} \quad w\lceil c/x \rceil \not\models \varphi \\
&\text{iff} \quad w \not\models \lceil c/x \rceil \varphi \\
&\text{iff} \quad w \models \neg \lceil c/x \rceil \varphi.
\end{aligned}
$$

(ii) We first show the direction from left to right as follows.

$$\begin{aligned} & w \models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil (\varphi \to \varphi') \text{ and } w \models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \varphi \\ \Longrightarrow\ & w\lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \models \varphi \to \varphi' \text{ and } w\lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \models \varphi \\ \Longrightarrow\ & w\lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \models \varphi' \\ \Longrightarrow\ & w \models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \varphi'. \end{aligned}$$

We next show the other direction as follows. Assume that $w \models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \varphi \to \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \varphi'$. Then:

$$\begin{aligned} & w\lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \models \varphi \\ \Longrightarrow\ & w \models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \varphi \\ \Longrightarrow\ & w \models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \varphi' && \text{(by assumption)} \\ \Longrightarrow\ & w\lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \models \varphi'. \end{aligned}$$

Hence $w \models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil (\varphi \to \varphi')$ iff $w \models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \varphi \to \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \varphi'$. □

B.3 Validity of the Axioms for Lazy Interventions

Here are the axioms of **AX** with the lazy interventions $\lfloor \cdot \rfloor$.

$\mathrm{DG_{LI}}$	$\vdash_g \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi$ iff $\vdash_{g\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor} \varphi$
$\mathrm{COND_{LI}}$	$\vdash_g (f = \boldsymbol{y}\lvert_{\boldsymbol{x}=\boldsymbol{c}}) \leftrightarrow \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor (f = \boldsymbol{y}\lvert_{\boldsymbol{x}=\boldsymbol{c}})$
$\mathrm{EQ_{LI}}$	$\vdash_g \boldsymbol{u_1} = \boldsymbol{u_2} \leftrightarrow \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor (\boldsymbol{u_1} = \boldsymbol{u_2})$ if $\mathsf{fv}(\boldsymbol{u_1}) = \mathsf{fv}(\boldsymbol{u_2}) = \emptyset$
$\mathrm{SPLIT_{LI}}$	$\vdash_g \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}},\, {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor \varphi \to \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor \lfloor {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor \varphi$
$\mathrm{SIMUL_{LI}}$	$\vdash_g \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor \lfloor {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor \varphi \to \lfloor {}^{\boldsymbol{c'_1}}/_{\boldsymbol{x'_1}},\, {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor \varphi$ if $\boldsymbol{x'_1} = \boldsymbol{x_1} \setminus \boldsymbol{x_2}$ and $\boldsymbol{c'_1} = \boldsymbol{c_1} \setminus \boldsymbol{c_2}$
$\mathrm{RPT_{LI}}$	$\vdash_g \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi \to \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi$
$\mathrm{CMP_{LI}}$	$\vdash_g \big(\lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor (\boldsymbol{x_2} = \boldsymbol{c_2}) \wedge \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor (\boldsymbol{x_3} = \boldsymbol{u})\big) \to \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}}, {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor (\boldsymbol{x_3} = \boldsymbol{u})$
$\mathrm{DISTR_{LI}}^{\neg}$	$\vdash_g (\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \neg\varphi) \leftrightarrow (\neg \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi)$
$\mathrm{DISTR_{LI}}^{\wedge}$	$\vdash_g (\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor (\varphi_1 \wedge \varphi_2)) \leftrightarrow (\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi_1 \wedge \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi_2)$

Proposition 12 (Basic properties of $\lfloor \cdot \rfloor$) *Let $\boldsymbol{x}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{y} \in \mathsf{CVar}^+$ be disjoint, $\boldsymbol{c}, \boldsymbol{c_1}, \boldsymbol{c_2} \in \mathsf{dConst}^+$, $\boldsymbol{u}, \boldsymbol{u_1}, \boldsymbol{u_2} \in \mathsf{Term}^+$, $f \in \mathsf{Fsym}$, and $\varphi \in \mathsf{Fml}$.*

1. $\mathrm{DG_{LI}}$ $\models_g \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi$ *iff* $\vdash_{g\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor} \varphi$.
2. $\mathrm{COND_{LI}}$ $\models (f = \boldsymbol{y}\lvert_{\boldsymbol{x}=\boldsymbol{c}}) \leftrightarrow \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor (f = \boldsymbol{y}\lvert_{\boldsymbol{x}=\boldsymbol{c}})$.
3. $\mathrm{EQ_{LI}}$ $\models \boldsymbol{u_1} = \boldsymbol{u_2} \leftrightarrow \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor (\boldsymbol{u_1} = \boldsymbol{u_2})$ *if* $\mathsf{fv}(\boldsymbol{u_1}) = \mathsf{fv}(\boldsymbol{u_2}) = \emptyset$.
4. $\mathrm{SPLIT_{LI}}$ $\models \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}},\, {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor \varphi \to \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor \lfloor {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor \varphi$
5. $\mathrm{SIMUL_{LI}}$ $\models \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor \lfloor {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor \varphi \to \lfloor {}^{\boldsymbol{c'_1}}/_{\boldsymbol{x'_1}},\, {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor \varphi$ *if* $\boldsymbol{x'_1} = \boldsymbol{x_1} \setminus \boldsymbol{x_2}$ *and* $\boldsymbol{c'_1} = \boldsymbol{c_1} \setminus \boldsymbol{c_2}$.

6. RPT$_{\mathrm{LI}}$ $\models \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi \rightarrow \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi$.

7. CMP$_{\mathrm{LI}}$ $\models \big(\lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor (\boldsymbol{x_2} = \boldsymbol{c_2}) \wedge \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor (\boldsymbol{x_3} = \boldsymbol{u}) \big) \rightarrow \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}}, {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor (\boldsymbol{x_3} = \boldsymbol{u})$.

Proof. Let $w = (g_w, \xi_w, m_w)$ be a world such that $\boldsymbol{x}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \boldsymbol{y} \in \mathsf{Var}(w)^+$.

1. The proof is similar to Proposition 10.
2. Let $g_w(\boldsymbol{x}) = \boldsymbol{u}$, G_w be the causal diagram corresponding to g_w, and $\mathtt{DEC}(\boldsymbol{x})$ be the set of all descendant variables of $\boldsymbol{x}$. Let $\boldsymbol{y_0}, \boldsymbol{y_1}$ be possibly empty tuples of variables such that $\boldsymbol{y} = \boldsymbol{y_0} :: \boldsymbol{y_1}$, $\boldsymbol{y_0} \subseteq \mathtt{DEC}(\boldsymbol{x})$, and $\boldsymbol{y_1} \cap \mathtt{DEC}(\boldsymbol{x}) = \emptyset$. Then on every undirected path between $\boldsymbol{y_0}$ and $\boldsymbol{y_1}$ in G_w, $\boldsymbol{x}$ are on chains or forks. Hence $P_{G_w}(\boldsymbol{y_0} :: \boldsymbol{y_1} | \boldsymbol{x}) = P_{G_w}(\boldsymbol{y_0} | \boldsymbol{x}) \cdot P_{G_w}(\boldsymbol{y_1} | \boldsymbol{x})$. Thus, we obtain:
$$\begin{aligned} & [\![\boldsymbol{y}|_{\boldsymbol{x}=\boldsymbol{c}}]\!]_w \\ = \; & P_{G_w}(\boldsymbol{y_0} :: \boldsymbol{y_1} | \boldsymbol{x} = \boldsymbol{c}) \\ = \; & P_{G_w}(\boldsymbol{y_0} | \boldsymbol{x} = \boldsymbol{c}) \cdot P_{G_w}(\boldsymbol{y_1} | \boldsymbol{x} = \boldsymbol{c}) \\ = \; & P_{G_{w\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor}}(\boldsymbol{y_0} | \boldsymbol{x} = \boldsymbol{c}) \cdot P_{G_{w\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor}}(\boldsymbol{y_1} | \boldsymbol{x} = \boldsymbol{c}) \\ = \; & P_{G_{w\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor}}(\boldsymbol{y_0} :: \boldsymbol{y_1} | \boldsymbol{x} = \boldsymbol{c}) \\ = \; & [\![\boldsymbol{y}|_{\boldsymbol{x}=\boldsymbol{c}}]\!]_{w\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor}. \end{aligned}$$
Therefore, $w \models f = \boldsymbol{y}|_{\boldsymbol{x}=\boldsymbol{c}}$ iff $w \models \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor f = \boldsymbol{y}|_{\boldsymbol{x}=\boldsymbol{c}}$.
3. Assume that $\mathsf{fv}(\boldsymbol{u_1}) = \mathsf{fv}(\boldsymbol{u_2}) = \emptyset$. Then for each $i = 1, 2$, $[\![\boldsymbol{u}_i]\!]_w = [\![\boldsymbol{u}_i]\!]_{w\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor}$. Hence, $\models \boldsymbol{u_1} = \boldsymbol{u_2} \leftrightarrow \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor (\boldsymbol{u_1} = \boldsymbol{u_2})$.
4. The proof is straightforward from the definition.
5. The proof is straightforward from the definition.
6. The proof is straightforward from the definition.
7. Assume that $w \models \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor (\boldsymbol{x_2} = \boldsymbol{c_2}) \wedge \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor (\boldsymbol{x_3} = \boldsymbol{u})$. Then $g_{w\lfloor \boldsymbol{c_1}/\boldsymbol{x_1} \rfloor}(\boldsymbol{x_2}) = \boldsymbol{c_2}$ and $g_{w\lfloor \boldsymbol{c_1}/\boldsymbol{x_1} \rfloor}(\boldsymbol{x_3}) = \boldsymbol{u}$. Let $w' = w\lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}} \rfloor$. Thus,
$$\begin{aligned} & [\![g_{w\lfloor \boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2} \rfloor}(\boldsymbol{x_3})]\!]_{w\lfloor \boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2} \rfloor} \\ & = [\![g_{w'\lfloor \boldsymbol{c_2}/\boldsymbol{x_2} \rfloor}(\boldsymbol{x_3})]\!]_{w'\lfloor \boldsymbol{c_2}/\boldsymbol{x_2} \rfloor} \\ & = [\![g_{w'}(\boldsymbol{x_3})[\boldsymbol{x_2} \mapsto \boldsymbol{c_2}]]\!]_{w'\lfloor \boldsymbol{c_2}/\boldsymbol{x_2} \rfloor} \\ & = [\![\boldsymbol{u}[\boldsymbol{x_2} \mapsto \boldsymbol{c_2}]]\!]_{w'\lfloor \boldsymbol{c_2}/\boldsymbol{x_2} \rfloor} \\ & = [\![\boldsymbol{u}]\!]_{w'\lfloor \boldsymbol{c_2}/\boldsymbol{x_2} \rfloor}. \\ & = [\![\boldsymbol{u}]\!]_{w\lfloor \boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2} \rfloor}. \end{aligned}$$
Therefore, $w \models \lfloor {}^{\boldsymbol{c_1}}/_{\boldsymbol{x_1}}, {}^{\boldsymbol{c_2}}/_{\boldsymbol{x_2}} \rfloor (\boldsymbol{x_3} = \boldsymbol{u})$. □

The lazy intervention operator $\lfloor \cdot \rfloor$ is also distributive w.r.t. logical connectives.

Proposition 13 (Distributive laws of $\lfloor \cdot \rfloor$) *Let $\boldsymbol{x} \in \mathit{CVar}^+$, $\boldsymbol{c} \in \mathit{dConst}^+$, and $\varphi, \varphi' \in \mathit{Fml}$.*

(i) DISTR$_{\mathrm{LI}}{}^{\neg}$ $\models \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \neg \varphi \leftrightarrow \neg \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi$.

(ii) DISTR$_{\mathrm{LI}}{}^{\rightarrow}$ $\models \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor (\varphi \rightarrow \varphi') \leftrightarrow \big(\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi \rightarrow \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \varphi' \big)$.

Similarly, the lazy intervention operator $\lfloor \cdot \rfloor$ is distributive w.r.t. $\vee$ and $\wedge$.

Proof. The proofs are analogous to those for Proposition 11. □

B.4 Validity of the Exchanges of Eager/Lazy Interventions

Here are the axioms of **AX** for the exchanges of eager/lazy interventions.

$$\text{EXPD}_{\text{EILI}} \ \vdash_g (\lceil c/x \rceil c' = y) \leftrightarrow (\lfloor c/x \rfloor c' = y)$$
$$\text{EXCD}_{\text{EILI}} \ \vdash_g pos(z) \rightarrow \big((\lceil c/x \rceil f = y|_z) \leftrightarrow (\lfloor c/x \rfloor f = y|_z)\big)$$

Proposition 14 (Exchanges of $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$) *Let $x, y, z \in \mathsf{CVar}^+$, $c, c' \in \mathsf{dConst}^+$, and $f \in \mathsf{Fsym}$.*

(i) $\text{EXPD}_{\text{EILI}} \models (\lceil c/x \rceil c' = y) \leftrightarrow (\lfloor c/x \rfloor c' = y)$.
(ii) $\text{EXCD}_{\text{EILI}} \models pos(z) \rightarrow \big((\lceil c/x \rceil f = y|_z) \leftrightarrow (\lfloor c/x \rfloor f = y|_z)\big)$.

Proof. (i) Let w be a world. By Proposition 6 (ii), we have: $[\![y]\!]_{w\lceil c/x \rceil} = m_{w\lceil c/x \rceil}(y) = m_{w\lfloor c/x \rfloor}(y) = [\![y]\!]_{w\lfloor c/x \rfloor}$. Therefore, we obtain the claim.

(ii) Let w be a world such that $w \models pos(z)$. Let $c_1, c_2 \in \mathsf{dConst}^+$. By the first claim, we have $w \models (\lceil c/x \rceil c_1 = y :: z) \leftrightarrow (\lfloor c/x \rfloor c_1 = y :: z)$ and $w \models (\lceil c/x \rceil c_2 = z) \leftrightarrow (\lfloor c/x \rfloor c_2 = z)$. Then we have:

$$\begin{aligned} [\![y|_z]\!]_{w\lceil c/x \rceil} &= \frac{[\![y :: z]\!]_{w\lceil c/x \rceil}}{[\![z]\!]_{w\lceil c/x \rceil}} \\ &= \frac{[\![y :: z]\!]_{w\lfloor c/x \rfloor}}{[\![z]\!]_{w\lfloor c/x \rfloor}} \\ &= [\![y|_z]\!]_{w\lfloor c/x \rfloor} \end{aligned}$$

Therefore, we obtain the claim. □

B.5 Remarks on Axioms and Invalid Formulas

From our axioms, we can derive the following formulas that are considered as axioms in the previous work [1].

$$\text{UNQ} \quad \vdash_g \lceil u/x \rceil (y = d) \rightarrow \lceil u/x \rceil (y \neq d') \ \text{ for } d \neq d'$$

We show examples formulas that are not valid in our model. The following formulas suggest the difference between the intervention $\lceil c/x \rceil \varphi$ and the conditioning $(x = c) \rightarrow \varphi$.

- Strengthened intervention:
  $\lceil u_1/x_1 \rceil \varphi \not\models \lceil u_1/x_1, u_2/x_2 \rceil \varphi$.
- Pseudo transitivity:
  $(\lceil u/x \rceil y = d) \wedge \lceil d/y \rceil \varphi \not\models \lceil u/x \rceil \varphi$.
- Weak pseudo transitivity:
  $(\lceil u/x \rceil y = d) \wedge \lceil u/x, d/y \rceil \varphi \not\models \lceil u/x \rceil \varphi$.
- Pseudo contraposition:
  $\lceil u/x \rceil d = y \not\models \lceil d/y \rceil (x = u)$.

- Replacing conjunction with intervention:
  $\boldsymbol{x} = \boldsymbol{u} \wedge \varphi \not\models \lceil \boldsymbol{u}/\boldsymbol{x} \rceil \varphi$.
- Pseudo Modus Ponens:
  $(\boldsymbol{x} = \boldsymbol{u} \wedge \lceil \boldsymbol{u}/\boldsymbol{x} \rceil \varphi) \not\models \varphi$.
- Pseudo Modus Tollens:
  $(\neg\varphi \wedge \lceil \boldsymbol{u}/\boldsymbol{x} \rceil \varphi) \not\models (\boldsymbol{x} \neq \boldsymbol{u})$.

Similar formulas are not valid also in [1], which does not deal with probability distributions.

B.6 Remark on Defining Lazy Interventions as Syntax Sugar

We remark that a lazy intervention $\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor$ can be defined as syntax sugar if we expand data generators.

Recall that $\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \varphi$ expresses that φ is satisfied in the lazy intervened world:

$$\mathfrak{M}, w \models \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \varphi \;\text{ iff }\; \mathfrak{M}, w\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \models \varphi.$$

To define $\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \varphi$ as syntax sugar, we *expand* the data generator g_w as follows. For each $x \in \mathit{dom}(g_w)$, we introduce a fresh auxiliary variable x', add $g_w(x') = x$, and replace every occurrence of x in $\mathit{range}(g_w)$ with x'. Then the corresponding causal diagram has arrows $x \rightarrow x'$ and $x' \rightarrow y$ instead of $x \rightarrow y$. Now the lazy intervention $\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \varphi$ can be defined as the eager intervention $\lceil \boldsymbol{c}/\boldsymbol{x}' \rceil \varphi$.

In summary, we can replace the lazy intervention $\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor \varphi$ with its corresponding eager intervention $\lceil \boldsymbol{c}/\boldsymbol{x}' \rceil \varphi$ by considering a model $\mathfrak{M}$ that have possible worlds equipped only with expanded data generators.

C Proof for the Soundness of $\mathbf{AX^{CP}}$

In this section, we prove the soundness of $\mathbf{AX^{CP}}$ w.r.t. the Kripke semantics for statistical causality by showing the validity of the axioms with the d-separation predicate *dsep* (Appendix C.1), with the non-ancestor causal predicate *nanc* (Appendix C.2), and with other causal predicates (Appendix C.3)

C.1 Validity of the Axioms with d-Separation

Here are the axioms of $\mathbf{AX^{CP}}$ with the d-separation *dsep*.

DsepCI	$\vdash_g (\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{pos}(\boldsymbol{z})) \rightarrow \boldsymbol{y}\|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c}} = \boldsymbol{y}\|_{\boldsymbol{z}}$
DsepSm	$\vdash_g \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \leftrightarrow \mathit{dsep}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$
DsepDc	$\vdash_g \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y} \cup \boldsymbol{y}', \boldsymbol{z}) \rightarrow (\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}', \boldsymbol{z}))$
DsepWu	$\vdash_g \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y} \cup \boldsymbol{v}, \boldsymbol{z}) \rightarrow \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cup \boldsymbol{v})$
DsepCn	$\vdash_g (\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{dsep}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{z} \cup \boldsymbol{y})) \rightarrow \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y} \cup \boldsymbol{v}, \boldsymbol{z})$
Dsep_{EI}	$\vdash_g (\lceil \boldsymbol{c}/\boldsymbol{z} \rceil \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})) \leftrightarrow \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$
Dsep_{LI}	$\vdash_g (\lfloor \boldsymbol{c}/\boldsymbol{z} \rfloor \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})) \leftrightarrow \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$
Dsep_{LIC}	$\vdash_g \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cup \boldsymbol{z}') \rightarrow \lfloor \boldsymbol{c}/\boldsymbol{z} \rfloor \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}')$

We first show the validity of DsepCI. It is well-known that the d-separation in a causal diagram G implies the conditional independence, but not vice versa [30]. However, if $[\![\boldsymbol{x}]\!]_w$ and $[\![\boldsymbol{y}]\!]_w$ are conditionally independent given $[\![\boldsymbol{z}]\!]_w$ for any interpretation $[\![\_]\!]_w$ factorizing G (i.e., for any world w with the data generator g_w corresponding to G), then they are d-separated by $\boldsymbol{z}$.

Proposition 15 (d-separation and conditional independence) *Let $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{CVar}^+$ and $\boldsymbol{z} \in \mathsf{CVar}^*$ be disjoint. Let $\boldsymbol{c} \in \mathsf{dConst}^+$.*

1. DsepCI
   $\models (dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z})) \rightarrow \boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} = \boldsymbol{y}|_{\boldsymbol{z}}$.
2. *For any finite, closed, acyclic data generator g, we have:*

$$\models_g (pos(\boldsymbol{z}) \rightarrow \boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} = \boldsymbol{y}|_{\boldsymbol{z}}) \quad \textit{implies} \quad \models_g dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}). \tag{6}$$

Proof. We show the first claim as follows. Let w be a world. Assume that $w \models dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z})$. Then in the causal diagram G_w, $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$. Thus, $\boldsymbol{x}$ and $\boldsymbol{y}$ are conditionally independent given $\boldsymbol{z}$. Therefore, $w \models \boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} = \boldsymbol{y}|_{\boldsymbol{z}}$.

We show the second claim as follows. Assume that $\models_g (pos(\boldsymbol{z}) \rightarrow \boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} = \boldsymbol{y}|_{\boldsymbol{z}})$. Then, for any world w with a data generator g, $[\![\boldsymbol{x}]\!]_w$ and $[\![\boldsymbol{y}]\!]_w$ are conditionally independent given $[\![\boldsymbol{z}]\!]_w$. Let G be the causal diagram corresponding to g. We recall that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are conditionally independent given $\boldsymbol{z}$ for any joint distribution P_G factorized according to G, then they are d-separated by $\boldsymbol{z}$ in G (see e.g., [30]). Therefore, we obtain $\models_g dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. □

d-separation is known to satisfy the *semi-graphoid* axioms [35], which we can describe using our logic as follows:

Proposition 16 (Semi-graphoid) *Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y}' \in \mathsf{CVar}^+$ and $\boldsymbol{z}, \boldsymbol{v} \in \mathsf{CVar}^*$ be disjoint. Then dsep satisfies:*

1. DsepSm *(symmetry):*
   $\models dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \leftrightarrow dsep(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z})$.
2. DsepDc *(decomposition):*
   $\models dsep(\boldsymbol{x}, \boldsymbol{y} \cup \boldsymbol{y}', \boldsymbol{z}) \rightarrow (dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge dsep(\boldsymbol{x}, \boldsymbol{y}', \boldsymbol{z}))$.
3. DsepWu *(weak union):*
   $\models dsep(\boldsymbol{x}, \boldsymbol{y} \cup \boldsymbol{v}, \boldsymbol{z}) \rightarrow dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cup \boldsymbol{v})$.
4. DsepCn *(contraction):*
   $\models (dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge dsep(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{z} \cup \boldsymbol{y})) \rightarrow dsep(\boldsymbol{x}, \boldsymbol{y} \cup \boldsymbol{v}, \boldsymbol{z})$.

The causal predicates and interventions satisfy the following axioms, which are later used in Appendix D to prove the soundness of Pearl's do-calculus rules (Proposition 2). We prove the validity of these axioms and an additional property Dsep$_{\mathrm{LIC}}$ as follows.

Proposition 17 (Relationships between $dsep$ and $\lceil\cdot\rceil$) *Let $\boldsymbol{x}, \boldsymbol{y} \in \mathsf{CVar}^+$ and $\boldsymbol{z}, \boldsymbol{z}' \in \mathsf{CVar}^*$ be disjoint, and $\boldsymbol{c} \in \mathsf{dConst}^+$.*

1. $\textsc{Dsep}_{\mathrm{EI1}}$ $\models (\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil\, dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})) \rightarrow dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$.
2. $\textsc{Dsep}_{\mathrm{EI2}}$ $\models dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \rightarrow \lceil{}^{c}\!/\!_{\boldsymbol{x}}\rceil\, dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$.
3. $\textsc{Dsep}_{\mathrm{LI1}}$ $\models (\lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor\, dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})) \rightarrow dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$.
4. $\textsc{Dsep}_{\mathrm{LI2}}$ $\models dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) \rightarrow \lfloor{}^{c}\!/\!_{\boldsymbol{x}}\rfloor\, dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$.
5. $\textsc{Dsep}_{\mathrm{LIC}}$ $\models dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\cup\boldsymbol{z}') \rightarrow \lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor\, dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}')$.

Proof. Let w be a world such that $\boldsymbol{x},\boldsymbol{y} \in \mathsf{Var}(w)^+$ and $\boldsymbol{z},\boldsymbol{z}' \in \mathsf{Var}(w)^*$. Recall that a data generator corresponds to a causal diagram that is defined as a directed acyclic graph (DAG) in Section 4. Let G be the causal diagram corresponding to the data generator g_w in the world w.

Then the causal diagram $G\lceil{}^{c}\!/\!_{\boldsymbol{x}}\rceil$ corresponding to $g_{w\lceil c/\boldsymbol{x}\rceil}$ is obtained by removing all arrows pointing to $\boldsymbol{x}$ in G. Similarly, the causal diagram $G\lfloor{}^{c}\!/\!_{\boldsymbol{x}}\rfloor$ corresponding to $g_{w\lfloor c/\boldsymbol{x}\rfloor}$ is obtained by removing all arrows emerging from $\boldsymbol{x}$ in G.

1. Assume that $w \models \lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil\, dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. Then $w\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil \models dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. Let p be an undirected path between $\boldsymbol{x}$ and $\boldsymbol{y}$ in $G\lceil{}^{c}\!/\!_{\boldsymbol{x}}\rceil$. Since $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$ in the diagram $G\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil$, we have:
   (a) there is no path p in $G\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil$ that has a chain $v' \rightarrow v \rightarrow v''$ s.t. $v \in \boldsymbol{z}$;
   (b) there is no path p in $G\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil$ that has a fork $v' \leftarrow v \rightarrow v''$ s.t. $v \in \boldsymbol{z}$;
   (c) if $G\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil$ has a path with a collider $v' \rightarrow v \leftarrow v''$, then $v \notin \mathtt{ANC}_*(\boldsymbol{z})$.
   By (a), if G has an undirected path with a chain $v' \rightarrow v \rightarrow v''$, then $v \in \boldsymbol{z}$, because $v \notin \boldsymbol{z}$ contradicts (a).
   By (b), if G has an undirected path with a fork $v' \leftarrow v \rightarrow v''$, then $v \in \boldsymbol{z}$, because $v \notin \boldsymbol{z}$ contradicts (b).
   Let p be an undirected path in $G\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil$ that has a collider $v' \rightarrow v \leftarrow v''$. By (c), we have $v \notin \mathtt{ANC}_*(\boldsymbol{z})$ in $G\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil$. Then, G also has the same path p, and the arrows connecting with v in G are the same as those in $G\lceil{}^{c}\!/\!_{\boldsymbol{z}}\rceil$. Hence, we obtain $v \notin \mathtt{ANC}_*(\boldsymbol{z})$ in G.
   Therefore, $w \models dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$.
2. Assume that $w \models dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. Then in the diagram G, $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$. By definition, $G\lceil{}^{c}\!/\!_{\boldsymbol{x}}\rceil$ is the same as G except that it has no arrows pointing to $\boldsymbol{x}$. Hence, also in $G\lceil{}^{c}\!/\!_{\boldsymbol{x}}\rceil$, $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$. Therefore, $w \models \lceil{}^{c}\!/\!_{\boldsymbol{x}}\rceil\, dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$.
3. Assume that $w \models \lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor\, dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. Then $w\lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor \models dsep(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$. Let p be an undirected path between $\boldsymbol{x}$ and $\boldsymbol{y}$ in $G\lfloor{}^{c}\!/\!_{\boldsymbol{x}}\rfloor$. Since $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$ in the diagram $G\lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor$, we have:
   (a) there is no path p in $G\lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor$ that has a chain $v' \rightarrow v \rightarrow v''$ s.t. $v \in \boldsymbol{z}$;
   (b) there is no path p in $G\lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor$ that has a fork $v' \leftarrow v \rightarrow v''$ s.t. $v \in \boldsymbol{z}$;
   (c) if $G\lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor$ has a path with a collider $v' \rightarrow v \leftarrow v''$, then $v \notin \mathtt{ANC}_*(\boldsymbol{z})$.
   By (a), if G has an undirected path with a chain $v' \rightarrow v \rightarrow v''$, then $v \in \boldsymbol{z}$, because $v \notin \boldsymbol{z}$ contradicts (a).
   By (b), if G has an undirected path with a fork $v' \leftarrow v \rightarrow v''$, then $v \in \boldsymbol{z}$, because $v \notin \boldsymbol{z}$ contradicts (b).
   Let p be an undirected path in $G\lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor$ that has a collider $v' \rightarrow v \leftarrow v''$. By (c), we obtain $v \notin \mathtt{ANC}_*(\boldsymbol{z})$ in $G\lfloor{}^{c}\!/\!_{\boldsymbol{z}}\rfloor$. Then, G also has the path p, and may

have additional arrows pointing to v and no arrows pointing from v. Hence, we obtain $v \notin \texttt{ANC}_*(\boldsymbol{z})$ in G.
Therefore, $w \models \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$.
4. The proof for Claim 4 is analogous to that for Claim 2.
5. Assume that $w \models \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cup \boldsymbol{z}')$. Then in the diagram G, $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z} \cup \boldsymbol{z}'$. By definition, $G\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{z}} \rfloor$ has no arrows emerging from $\boldsymbol{z}$. If $G\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor$ has no *undirected* path between $\boldsymbol{x}$ and $\boldsymbol{y}$, then $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}'$, hence $w \models \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}')$.
Otherwise, let p be an undirected path between $\boldsymbol{x}$ and $\boldsymbol{y}$ in $G\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor$. Since $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z} \cup \boldsymbol{z}'$, we fall into one of the three cases in Definition 2.
   (a) If p has a chain $v' \to v \to v''$ s.t. $v \in \boldsymbol{z} \cup \boldsymbol{z}'$, then $v \in \boldsymbol{z}'$, because $G\lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{z}} \rfloor$ has no arrows pointing from $\boldsymbol{z}$. Hence, p is d-separated by $\boldsymbol{z}'$.
   (b) For the same reason as (a), if p has a fork $v' \leftarrow v \to v''$ s.t. $v \in \boldsymbol{z} \cup \boldsymbol{z}'$, then $v \in \boldsymbol{z}'$. Hence, p is d-separated by $\boldsymbol{z}'$.
   (c) If p has a collider $v' \to v \leftarrow v''$ s.t. $v \notin \texttt{ANC}_*(\boldsymbol{z} \cup \boldsymbol{z}')$, then $v \notin \texttt{ANC}_*(\boldsymbol{z}')$. Thus p is d-separated by $\boldsymbol{z}'$.
   Therefore, $w \models \lfloor {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rfloor \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}')$. □

Remark 2. In contrast with Claim 5 in Proposition 17, there exists a world w s.t.

$$w \not\models \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cup \boldsymbol{z}') \to \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{z}} \rceil \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}').$$

To see this, assume that $w \models \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cup \boldsymbol{z}')$. Suppose that w has a causal diagram G where there is an undirected path p between $\boldsymbol{x}$ and $\boldsymbol{y}$ that has a fork $v' \leftarrow v \to v''$ s.t. $v \in \boldsymbol{z}$ and no other variable in $\boldsymbol{z} \cup \boldsymbol{z}'$ appears on p. Then $G\lceil {}^{\boldsymbol{c}}/_{\boldsymbol{z}} \rceil$ also has the path p, because the intervention $\lceil {}^{\boldsymbol{c}}/_{\boldsymbol{z}} \rceil$ removes no arrows in p. Hence, p is d-separated by $\boldsymbol{z}$ but not by $\boldsymbol{z}'$ in $G\lceil {}^{\boldsymbol{c}}/_{\boldsymbol{z}} \rceil$. Therefore, $w \not\models \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{z}} \rceil \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}')$.

C.2 Validity of the Axioms with *nanc*

Here are the axioms of $\mathbf{AX}^{\mathbf{CP}}$ with the non-anscestor predicate *nanc* and a property Nanc0.

Nanc0	$\vdash_g \mathit{nanc}(\boldsymbol{x}, \boldsymbol{y}) \to ((\boldsymbol{c}' = \boldsymbol{y}) \leftrightarrow \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil (\boldsymbol{c}' = \boldsymbol{y}))$
Nanc1	$\vdash_g (\mathit{nanc}(\boldsymbol{x}, \boldsymbol{y}) \land \mathit{nanc}(\boldsymbol{x}, \boldsymbol{z})) \to (f{=}\boldsymbol{y}\vert_{\boldsymbol{z}} \leftrightarrow \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil (f{=}\boldsymbol{y}\vert_{\boldsymbol{z}}))$
Nanc2	$\vdash_g \mathit{nanc}(\boldsymbol{x}, \boldsymbol{y}) \leftrightarrow \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \mathit{nanc}(\boldsymbol{x}, \boldsymbol{y})$
Nanc3	$\vdash_g \mathit{nanc}(\boldsymbol{x}, \boldsymbol{y}) \to \lceil {}^{\boldsymbol{c}}/_{\boldsymbol{x}} \rceil \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \emptyset)$
Nanc4	$\vdash_g (\mathit{nanc}(\boldsymbol{x}, \boldsymbol{z}) \land \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})) \to \mathit{nanc}(\boldsymbol{x}, \boldsymbol{y})$

Concerning *nanc*, the axioms Nanc1 to Nanc4 are sufficient for us to derive the rules of Pearl's do-calculus.

Proposition 18 (Validity of axioms with *nanc***)** *Let* $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{CVar}^+$ *be disjoint,* $\boldsymbol{c} \in \mathsf{dConst}^+$*,* $\boldsymbol{c}' \in \mathsf{Const}^+$*, and* $f \in \mathsf{Fsym}$*.*

1. Nanc0
$\models nanc(\boldsymbol{x}, \boldsymbol{y}) \rightarrow (\boldsymbol{c}' = \boldsymbol{y} \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (\boldsymbol{c}' = \boldsymbol{y}))$.
2. Nanc1
$\models (nanc(\boldsymbol{x}, \boldsymbol{y}) \wedge nanc(\boldsymbol{x}, \boldsymbol{z})) \rightarrow (f{=}\boldsymbol{y}|_{\boldsymbol{z}} \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (f{=}\boldsymbol{y}|_{\boldsymbol{z}}))$.
3. Nanc2
$\models nanc(\boldsymbol{x}, \boldsymbol{y}) \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil\, nanc(\boldsymbol{x}, \boldsymbol{y})$.
4. Nanc3
$\models nanc(\boldsymbol{x}, \boldsymbol{y}) \rightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil\, dsep(\boldsymbol{x}, \boldsymbol{y}, \emptyset)$.
5. Nanc4
$\models (nanc(\boldsymbol{x}, \boldsymbol{z}) \wedge dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})) \rightarrow nanc(\boldsymbol{x}, \boldsymbol{y})$.

Proof. Let $w = (g_w, \xi_w, m_w)$ be a world such that $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{Var}(w)^+$.

1. Assume that $w \models nanc(\boldsymbol{x}, \boldsymbol{y})$. Let G_w be the causal diagram corresponding to the data generator g_w. Then $\boldsymbol{x} \cap \mathtt{ANC}(\boldsymbol{y}) = \emptyset$ in G_w. This means that the value of $\boldsymbol{y}$ does not depend on that of $\boldsymbol{x}$. Thus we obtain:

$$\begin{aligned} m_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil}(\boldsymbol{y}) &= [\![g_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil}(\boldsymbol{y})]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} \\ &= [\![g_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil}(\boldsymbol{y})]\!]_{w} && (\text{by } \xi_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} = \xi_w) \\ &= [\![g_w(\boldsymbol{y})]\!]_w && (\text{by } g_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil}(\boldsymbol{y}) = g_w(\boldsymbol{y})) \\ &= m_w(\boldsymbol{y}). \end{aligned}$$

Thus, $w \models \boldsymbol{c}' = \boldsymbol{y}$ iff $w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \models \boldsymbol{c}' = \boldsymbol{y}$. Therefore, $w \models \boldsymbol{c}'{=}\boldsymbol{y} \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (\boldsymbol{c}'{=}\boldsymbol{y})$.
2. Let $c_0, c_1 \in \mathsf{Const}$. Assume that $w \models nanc(\boldsymbol{x}, \boldsymbol{y}) \wedge nanc(\boldsymbol{x}, \boldsymbol{z})$. Then $w \models nanc(\boldsymbol{x}, \boldsymbol{y} :: \boldsymbol{z})$. By Claim 1, $w \models c_0 = \boldsymbol{y} :: \boldsymbol{z} \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (c_0 = \boldsymbol{y} :: \boldsymbol{z})$ and $w \models c_1 = \boldsymbol{z} \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (c_1 = \boldsymbol{z})$. By $[\![\boldsymbol{y} :: \boldsymbol{z}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} = [\![\boldsymbol{y} :: \boldsymbol{z}]\!]_w$ and $[\![\boldsymbol{z}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} = [\![\boldsymbol{z}]\!]_w$, we have $[\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_{w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil} = [\![\boldsymbol{y}|_{\boldsymbol{z}}]\!]_w$. Therefore, $w \models f = \boldsymbol{y}|_{\boldsymbol{z}} \leftrightarrow \lceil \boldsymbol{c}/\boldsymbol{x} \rceil (f = \boldsymbol{y}|_{\boldsymbol{z}})$.
3. We show the direction from left to right as follows. Assume that $w \models nanc(\boldsymbol{x}, \boldsymbol{y})$. Then, in the diagram G, all variables in $\boldsymbol{x}$ are non-ancestors of the variables in $\boldsymbol{y}$; i.e., G has no *directed* path from $\boldsymbol{x}$ to $\boldsymbol{y}$. Since the eager intervention $\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ removes only arrows pointing to $\boldsymbol{x}$, $G\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ still has no *directed* path from $\boldsymbol{x}$ to $\boldsymbol{y}$. Therefore, $w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil\, nanc(\boldsymbol{x}, \boldsymbol{y})$.
The other direction is shown in a similar way, since the eager intervention $\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ only remove arrows pointing to $\boldsymbol{x}$.
4. Assume that $w \models nanc(\boldsymbol{x}, \boldsymbol{y})$. By Claim 3, $w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil\, nanc(\boldsymbol{x}, \boldsymbol{y})$, hence $w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \models nanc(\boldsymbol{x}, \boldsymbol{y})$. Then, in the diagram $G\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$, all variables in $\boldsymbol{x}$ are non-ancestors of the variables in $\boldsymbol{y}$; i.e., $G\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ has no *directed* path from $\boldsymbol{x}$ to $\boldsymbol{y}$.
Suppose that $G\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ has no *undirected* path between $\boldsymbol{x}$ and $\boldsymbol{y}$. By Definition 2, $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\emptyset$, namely, they are independent. Hence, $w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \models dsep(\boldsymbol{x}, \boldsymbol{y}, \emptyset)$. Therefore, $w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil\, dsep(\boldsymbol{x}, \boldsymbol{y}, \emptyset)$.
Suppose that $G\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ has some undirected path p between $\boldsymbol{x}$ and $\boldsymbol{y}$. By the definition of the eager intervention, $G\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ has no arrows pointing to $\boldsymbol{x}$, hence

has arrows pointing from $\boldsymbol{x}$. On the other hand, since $G\lceil \boldsymbol{c}/\boldsymbol{x} \rceil$ has no *directed* path from $\boldsymbol{x}$ to $\boldsymbol{y}$, p is not directed. Thus, p has a collider node v; i.e., it is of the form $\boldsymbol{x} \to \cdots \to v \leftarrow \cdots \boldsymbol{y}$. Then, by (c) in Definition 2, p is d-separated by $\emptyset$; namely, $w\lceil \boldsymbol{c}/\boldsymbol{x} \rceil \models dsep(\boldsymbol{x}, \boldsymbol{y}, \emptyset)$. Therefore, $w \models \lceil \boldsymbol{c}/\boldsymbol{x} \rceil\, dsep(\boldsymbol{x}, \boldsymbol{y}, \emptyset)$.

5. We show the contraposition as follows. Assume that $w \models \neg nanc(\boldsymbol{x}, \boldsymbol{y}) \wedge dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. Then it is sufficient to prove $w \models \neg nanc(\boldsymbol{x}, \boldsymbol{z})$.
Recall the definition in Appendix A.4. By assumption, $\boldsymbol{x} \cap \mathtt{ANC}(\boldsymbol{y}) \neq \emptyset$. Then there are $x_0 \in \boldsymbol{x}$ and $y_0 \in \boldsymbol{y}$ s.t. x_0 is an ancestor of y_0, i.e., $x_0 \in \mathtt{ANC}(y_0)$. Then there exists a directed path from x_0 to y_0. Let p be a directed path from x_0 to y_0. By $w \models dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, p is d-separated by $\boldsymbol{z}$. By (a) of Definition 2, there is a variable $z_0 \in \boldsymbol{z}$ on p, hence $x_0 \in \mathtt{ANC}(z_0)$. Therefore, $\boldsymbol{x} \cap \mathtt{ANC}(\boldsymbol{z}) \neq \emptyset$, i.e., $w \models \neg nanc(\boldsymbol{x}, \boldsymbol{z})$. □

C.3 Validity of the Axioms with Other Causal Predicates

Here are the axioms of $\mathbf{AX}^{\mathbf{CP}}$ that replace *allnanc* with *nanc* and *pa* with *nanc* or *dsep*.

AllNanc	$\vdash_g$	$allnanc(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \to nanc(\boldsymbol{x}, \boldsymbol{z})$
PaNanc	$\vdash_g$	$pa(\boldsymbol{x}, \boldsymbol{y}) \to nanc(\boldsymbol{y}, \boldsymbol{x})$
PaDsep	$\vdash_g$	$pa(\boldsymbol{z}, \boldsymbol{x}) \to \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor\, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$

We prove the validity of these axioms as follows.

Proposition 19 (Validity of axioms with other causal predicates) *Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{CVar}^+$ be disjoint, and $\boldsymbol{c} \in \mathsf{dConst}^+$.*

1. AllNanc $\models allnanc(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \to nanc(\boldsymbol{x}, \boldsymbol{z})$.
2. PaNanc $\models pa(\boldsymbol{x}, \boldsymbol{y}) \to nanc(\boldsymbol{y}, \boldsymbol{x})$.
3. PaDsep $\models pa(\boldsymbol{z}, \boldsymbol{x}) \to \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor\, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$.

Proof. Let $w = (g_w, \xi_w, m_w)$ be a world such that $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{Var}(w)^+$.

1. This claim is straightforward from the definitions of the semantics of *allnanc* and *nanc*.
2. This claim is straightforward from Proposition 7.
3. Assume that $w \models pa(\boldsymbol{z}, \boldsymbol{x})$. Then, in the diagram G, $\boldsymbol{z}$ is the set of all variables pointing to $\boldsymbol{x}$.
If $G\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor$ has no *undirected* path between $\boldsymbol{x}$ and $\boldsymbol{y}$, then $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$, hence $w \models \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor\, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$.
Otherwise, let p be an undirected path between $\boldsymbol{x}$ and $\boldsymbol{y}$ in $G\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor$. By definition, $G\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor$ has no arrows emerging from $\boldsymbol{x}$. Since $\boldsymbol{z}$ is the set of all variables pointing to $\boldsymbol{x}$, p has:
   - either a chain $x_0 \leftarrow z_0 \leftarrow v$ s.t. $x_0 \in \boldsymbol{x}$, $z_0 \in \boldsymbol{z}$, and $v \notin \boldsymbol{x} \cup \boldsymbol{z}$,
   - or a fork $x_0 \leftarrow z_0 \to v$ s.t. $x_0 \in \boldsymbol{x}$, $z_0 \in \boldsymbol{z}$, and $v \notin \boldsymbol{x} \cup \boldsymbol{z}$.

   Thus, p is d-separated by z_0. Hence, in $G\lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor$, $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$. Therefore, $w \models \lfloor \boldsymbol{c}/\boldsymbol{x} \rfloor\, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$. □

D Details of the Derivation of the Do-Calculus Rules Using $\mathbf{AX^{CP}}$

In this section, we formalize and derive the three rules of Pearl's do-calculus [29] using our statistical causal language (StaCL).

By Proposition 1, StaCL formulas correspond to the do-calculus notations as follows.

- $\lceil {}^{\boldsymbol{c}}\!/\!{}_{\boldsymbol{x}} \rceil(\boldsymbol{c}' = \boldsymbol{y})$ describes the post-intervention distribution $P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x} = \boldsymbol{c}))$ of $\boldsymbol{y}$. For instance, given a world w, $w \models \lceil {}^{\boldsymbol{c}}\!/\!{}_{\boldsymbol{x}} \rceil \boldsymbol{c}' = \boldsymbol{y}$ represents $P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x} = \boldsymbol{c})) = [\![\boldsymbol{c}']\!]_w$. Note that the $do(\boldsymbol{x} = \boldsymbol{c})$ operation is expressed as the eager intervention $\lceil {}^{\boldsymbol{c}}\!/\!{}_{\boldsymbol{x}} \rceil$ in our formulation.
- $\lceil {}^{\boldsymbol{c}}\!/\!{}_{\boldsymbol{x}} \rceil(f = \boldsymbol{y}|_{\boldsymbol{z}})$ describes the post-intervention conditional distribution $P_{G_w}(\boldsymbol{y} \mid do(\boldsymbol{x} = \boldsymbol{c}), \boldsymbol{z})$ of $\boldsymbol{y}$ given $\boldsymbol{z}$. Note that the conditioning on $\boldsymbol{z}$ takes place after the intervention $do(\boldsymbol{x} = \boldsymbol{c})$ is performed.

To formalize the rules of the do-calculus, we denote the set of all *conditioning variables* appearing in a formula φ by:

$$\mathsf{cdv}(\varphi) = \{\boldsymbol{z} \mid \boldsymbol{y}|_{\boldsymbol{z}} \in \mathsf{fv}(\varphi) \cap \mathsf{FVar}\} \cup \{(\boldsymbol{z} :: \boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{c}} \mid \boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c}} \in \mathsf{fv}(\varphi) \cap \mathsf{FVar}\}.$$

Now we formalize the rules of Pearl's do-calculus in Proposition 2. After that, we explain the meaning of these rules.

Proposition 2 (Do-calculus rules) *Let* $\boldsymbol{v}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathsf{CVar}^+$ *be disjoint,* $\boldsymbol{x_1}, \boldsymbol{x_2} \in \mathsf{CVar}^+$*, and* $\boldsymbol{c_0}, \boldsymbol{c_1}, \boldsymbol{c_2} \in \mathsf{dConst}^+$*. Let* $S = \mathsf{cdv}(\varphi_0) \cup \mathsf{cdv}(\varphi_1)$*.*

1. Do1. *Introduction/elimination of conditioning:*

$$\vdash_g \lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}} \rceil(dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \textstyle\bigwedge_{\boldsymbol{s} \in S} pos(\boldsymbol{s})) \rightarrow ((\lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}} \rceil \varphi_0) \leftrightarrow \lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}} \rceil \varphi_1)$$

where φ_1 *is obtained by replacing some occurrences of* $\boldsymbol{y}|_{\boldsymbol{z}}$ *in* φ_0 *with* $\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c_1}}$*;*

2. Do2. *Exchange between intervention and conditioning:*

$$\vdash_g \lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}} \rceil \lfloor {}^{\boldsymbol{c_1}}\!/\!{}_{\boldsymbol{x}} \rfloor(dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \textstyle\bigwedge_{\boldsymbol{s} \in S} pos(\boldsymbol{s})) \rightarrow ((\lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}}, {}^{\boldsymbol{c_1}}\!/\!{}_{\boldsymbol{x}} \rceil \varphi_0) \leftrightarrow \lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}} \rceil \varphi_1)$$

where φ_1 *is obtained by replacing every occurrence of* $\boldsymbol{y}|_{\boldsymbol{z}}$ *in* φ_0 *with* $\boldsymbol{y}|_{\boldsymbol{z},\boldsymbol{x}=\boldsymbol{c_1}}$*;*

3. Do3 *Introduction/elimination of intervention:*

$$\vdash_g \lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}} \rceil(allnanc(\boldsymbol{x_1}, \boldsymbol{x}, \boldsymbol{y}) \wedge \lceil {}^{\boldsymbol{c_1}}\!/\!{}_{\boldsymbol{x_1}} \rceil(dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z}))) \\ \rightarrow ((\lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}} \rceil \varphi) \leftrightarrow \lceil {}^{\boldsymbol{c_0}}\!/\!{}_{\boldsymbol{v}}, {}^{\boldsymbol{c_1}}\!/\!{}_{\boldsymbol{x_1}}, {}^{\boldsymbol{c_2}}\!/\!{}_{\boldsymbol{x_2}} \rceil \varphi)$$

where $\mathsf{fv}(\varphi) = \{\boldsymbol{y}|_{\boldsymbol{z}}\}$ *and* $\boldsymbol{x} \stackrel{def}{=} \boldsymbol{x_1} :: \boldsymbol{x_2}$*.*

We explain these three rules as follows.

1. The first rule allows for adding/removing the conditioning on $\boldsymbol{x}$ when $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$ (hence when they are conditionally independent given $\boldsymbol{z}$).
In the do-calculus, this is expressed by:
$$P(\boldsymbol{y} \mid do(\boldsymbol{v}), \boldsymbol{z}) = P(\boldsymbol{y} \mid do(\boldsymbol{v}), \boldsymbol{z}, \boldsymbol{x}) \quad \text{if } (\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{y} \mid \boldsymbol{v}, \boldsymbol{z})_{G_{\overline{\boldsymbol{v}}}}$$
where
   - $G_{\overline{\boldsymbol{v}}}$ is the diagram obtained by deleting all arrows pointing to nodes in $\boldsymbol{v}$;
   - $(\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{y} \mid \boldsymbol{v}, \boldsymbol{z})_{G_{\overline{\boldsymbol{v}}}}$ represents that $\boldsymbol{x}$ and $\boldsymbol{y}$ are d-separated by $\boldsymbol{v} \cup \boldsymbol{z}$ in the causal diagram $G_{\overline{\boldsymbol{v}}}$.

   In our formulation, the deletion of arrows pointing to $\boldsymbol{v}$ is expressed by the eager intervention $\lceil c_0/\boldsymbol{v} \rceil$.
2. The second rule represents that the conditioning on $\boldsymbol{x}$ and the intervention to $\boldsymbol{x}$ result in the same conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{z}$ under the condition that all back-door paths from $\boldsymbol{x}$ to $\boldsymbol{y}$ (Definition 3) are d-separated by $\boldsymbol{v} \cup \boldsymbol{z}$ (Definition 2).[6]
In the do-calculus, this is expressed by:
$$P(\boldsymbol{y} \mid do(\boldsymbol{v}), \boldsymbol{x}, \boldsymbol{z}) = P(\boldsymbol{y} \mid do(\boldsymbol{v}), do(\boldsymbol{x}), \boldsymbol{z}) \quad \text{if } (\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{v})_{G_{\overline{\boldsymbol{v}}\underline{\boldsymbol{x}}}}$$
where $G_{\overline{\boldsymbol{v}}\underline{\boldsymbol{x}}}$ is the diagram obtained by deleting all arrows pointing to nodes in $\boldsymbol{v}$ and deleting all arrows emerging from nodes in $\boldsymbol{x}$.
In our formulation, the "upper manipulation" $\overline{\boldsymbol{v}}$ is expressed by the eager intervention $\lceil c_0/\boldsymbol{v} \rceil$ whereas the "lower-manipulation" $\underline{\boldsymbol{x}}$ is expressed by the lazy intervention $\lfloor c_1/\boldsymbol{x} \rfloor$.
Recall that the lazy intervention $\lfloor c_1/\boldsymbol{x} \rfloor$ removes all arrows emerging from $\boldsymbol{x}$, and hence preserves only back-door paths from x to y while removing all other undirected paths between x and y (Remark 1). Thus, $\lfloor c_1/\boldsymbol{x} \rfloor\, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ represents that all back-door paths from $\boldsymbol{x}$ to $\boldsymbol{y}$ are d-separated by $\boldsymbol{z}$.
3. The third rule allows for adding/removing the intervention to $\boldsymbol{x}$ without changing the conditional probability distribution of $\boldsymbol{y}$ given $\boldsymbol{z}$ under a certain condition.
In the do-calculus, this is expressed by:
$$P(\boldsymbol{y} \mid do(\boldsymbol{v}), \boldsymbol{z}) = P(\boldsymbol{y} \mid do(\boldsymbol{v}), do(\boldsymbol{x}), \boldsymbol{z}) \quad \text{if } (\boldsymbol{x} \perp\!\!\!\perp \boldsymbol{y} \mid \boldsymbol{z}, \boldsymbol{v})_{G_{\overline{\overline{\boldsymbol{v}}\boldsymbol{x} \setminus \mathtt{ANC}(\boldsymbol{z})}}}$$
where $G_{\overline{\overline{\boldsymbol{v}}\boldsymbol{x} \setminus \mathtt{ANC}(\boldsymbol{z})}}$ is the diagram obtained by deleting all arrows pointing to nodes in $\boldsymbol{v}$ and then deleting those in $\boldsymbol{x} \setminus \mathtt{ANC}(\boldsymbol{z})$.

Now, we derive these three rules using $\mathbf{AX}^{\mathbf{CP}}$ as follows.

[6] This condition is denoted by $\lfloor c_1/\boldsymbol{x} \rfloor\, dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cup \boldsymbol{v})$ and follows from Proposition 17 and $dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cup \boldsymbol{v})$.

Proof. Let w be a world such that $\boldsymbol{v}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x_1}, \boldsymbol{x_2} \in \mathsf{CVar}(w)^+$.

1. We prove the first claim as follows. Let $\psi_{\text{pre}} \stackrel{\text{def}}{=} \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \bigwedge_{\boldsymbol{s} \in S} \mathit{pos}(\boldsymbol{s})$. Then:

$$\begin{aligned}
&\text{By DsepCI,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil} (\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{pos}(\boldsymbol{z})) \to (\boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}} = \boldsymbol{y}|_{\boldsymbol{z}}) &(7)\\
&\text{By (7), Eq2, PT, MP,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil} \psi_{\text{pre}} \to (\varphi_0 \leftrightarrow \varphi_1) &(8)\\
&\text{By (8), DG}_{\text{EI}}\text{, MP,}\\
&\vdash_g \lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil(\psi_{\text{pre}} \to (\varphi_0 \leftrightarrow \varphi_1)) &(9)\\
&\text{By (9), Distr}_{\text{EI}}{}^{\wedge}\text{, Distr}_{\text{EI}}{}^{\to}\text{, PT, MP,}\\
&\vdash_g (\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil\psi_{\text{pre}}) \to ((\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil\varphi_0) \leftrightarrow \lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil\varphi_1).
\end{aligned}$$

Therefore, Claim (1) follows.

2. We prove the second claim as follows. Let $\psi_{\text{pre}} \stackrel{\text{def}}{=} \mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \bigwedge_{\boldsymbol{s} \in S} \mathit{pos}(\boldsymbol{s})$. Then:

$$\begin{aligned}
&\text{By DsepCI,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor} (\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{pos}(\boldsymbol{z})) \to (\boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}} = \boldsymbol{y}|_{\boldsymbol{z}}) &(10)\\
&\text{By (10), DG}_{\text{EI}}\text{, MP,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil} \lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor\big((\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{pos}(\boldsymbol{z})) \to (\boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}} = \boldsymbol{y}|_{\boldsymbol{z}})\big) &(11)\\
&\text{By (11), Distr}_{\text{EI}}{}^{\to}\text{, MP,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil} (\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor(\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{pos}(\boldsymbol{z})))\\
&\qquad \to (\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor\boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}} = \boldsymbol{y}|_{\boldsymbol{z}}) &(12)\\
&\text{By (12), Eq}_{\text{F}}\text{, PT, MP,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil} (\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor(\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{pos}(\boldsymbol{z})))\\
&\qquad \to (\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor f = \boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}} \leftrightarrow \lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor f = \boldsymbol{y}|_{\boldsymbol{z}}) &(13)\\
&\text{By (13), Cond}_{\text{LI}}\text{, PT, MP,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil} (\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor(\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{pos}(\boldsymbol{z})))\\
&\qquad \to (f = \boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}} \leftrightarrow \lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor f = \boldsymbol{y}|_{\boldsymbol{z}}) &(14)\\
&\text{By (14), Excd}_{\text{EILI}}\text{, PT, MP,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil} (\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor(\mathit{dsep}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge \mathit{pos}(\boldsymbol{z})))\\
&\qquad \to (f = \boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x}=\boldsymbol{c_1}} \leftrightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x}\rceil f = \boldsymbol{y}|_{\boldsymbol{z}}) &(15)\\
&\text{By (15), Eq2, PT, MP,}\\
&\vdash_{g\lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil} (\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor\psi_{\text{pre}}) \to ((\lceil \boldsymbol{c_1}/\boldsymbol{x}\rceil\varphi_0) \leftrightarrow \varphi_1) &(16)\\
&\text{By (16), DG}_{\text{EI}}\text{, MP,}\\
&\vdash_g \lceil \boldsymbol{c_0}/\boldsymbol{v}\rceil\big((\lfloor \boldsymbol{c_1}/\boldsymbol{x}\rfloor\psi_{\text{pre}}) \to ((\lceil \boldsymbol{c_1}/\boldsymbol{x}\rceil\varphi_0) \leftrightarrow \varphi_1)\big) &(17)\\
&\text{By (17), Distr}_{\text{EI}}{}^{\to}\text{, PT, MP,}
\end{aligned}$$

$$\vdash_g (\lceil c_0/v \rceil \lfloor c_1/x \rfloor \psi_{\text{pre}}) \rightarrow ((\lceil c_0/v \rceil \lceil c_1/x \rceil \varphi_0) \leftrightarrow \lceil c_0/v \rceil \varphi_1) \quad (18)$$

By (18), $\textsc{Simul}_{\text{EI}}$, MP,

$$\vdash_g (\lceil c_0/v \rceil \lfloor c_1/x \rfloor \psi_{\text{pre}}) \rightarrow ((\lceil c_0/v, c_1/x \rceil \varphi_0) \leftrightarrow \lceil c_0/v \rceil \varphi_1).$$

Therefore, Claim (2) follows.

3. We prove the third claim as follows. Let $f \in \mathsf{Fsym}$, $\psi_{\text{dp}} \stackrel{\text{def}}{=} dsep(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z})$, $\psi_{\text{pre}} \stackrel{\text{def}}{=} allnanc(\boldsymbol{x_1}, \boldsymbol{x}, \boldsymbol{y}) \wedge \lceil c_1/x_1 \rceil \psi_{\text{dp}}$, and $\psi_{\text{do2}} \stackrel{\text{def}}{=} \lfloor c_2/x_2 \rfloor (dsep(\boldsymbol{x_2}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z}))$. Let $g_0 \stackrel{\text{def}}{=} g\lceil c_0/v \rceil$. Then:

By PT, MP,

$$\vdash_{g_0} \psi_{\text{pre}} \rightarrow \lceil c_1/x_1 \rceil \psi_{\text{dp}} \quad (19)$$

By $\textsc{DsepDc}$, PT, MP,

$$\vdash_{g_0\lceil c_1/x_1 \rceil} \psi_{\text{dp}} \rightarrow (dsep(\boldsymbol{x_1}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z})) \quad (20)$$

$$\vdash_{g_0\lceil c_1/x_1 \rceil} \psi_{\text{dp}} \rightarrow (dsep(\boldsymbol{x_2}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z})) \quad (21)$$

By $\textsc{DsepCI}$, MP,

$$\vdash_{g_0\lceil c_1/x_1 \rceil} (dsep(\boldsymbol{x_2}, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z})) \rightarrow (\boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x_2}=\boldsymbol{c_2}} = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (22)$$

By $\textsc{Eq}_{\text{F}}$, PT, MP,

$$\vdash_{g_0\lceil c_1/x_1 \rceil} (f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (23)$$

By $\textsc{DG}_{\text{EI}}$, MP,

$$\vdash_{g_0} \lceil c_1/x_1 \rceil (f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (24)$$

By (21), $\textsc{Dsep}_{\text{LI2}}$,PT,MP,

$$\vdash_{g_0\lceil c_1/x_1 \rceil} \psi_{\text{dp}} \rightarrow \lfloor c_2/x_2 \rfloor (dsep(\boldsymbol{x}_2, \boldsymbol{y}, \boldsymbol{z}) \wedge pos(\boldsymbol{z})) \quad (25)$$

By $\textsc{AllNanc}$, PT, MP,

$$\vdash_{g_0} \psi_{\text{pre}} \rightarrow nanc(\boldsymbol{x}_1, \boldsymbol{z}) \quad (26)$$

By (26), $\textsc{Nanc2}$, PT, MP,

$$\vdash_{g_0} \psi_{\text{pre}} \rightarrow \lceil c_1/x_1 \rceil nanc(\boldsymbol{x}_1, \boldsymbol{z}) \quad (27)$$

By $\textsc{Nanc4}$, PT, MP,

$$\vdash_{g_0\lceil c_1/x_1 \rceil} (dsep(\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{z}) \wedge nanc(\boldsymbol{x}_1, \boldsymbol{z})) \rightarrow (nanc(\boldsymbol{x}_1, \boldsymbol{y}) \wedge nanc(\boldsymbol{x}_1, \boldsymbol{z})) \quad (28)$$

By (28), $\textsc{DG}_{\text{EI}}$, MP,

$$\vdash_{g_0} \lceil c_1/x_1 \rceil ((dsep(\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{z}) \wedge nanc(\boldsymbol{x}_1, \boldsymbol{z})) \rightarrow (nanc(\boldsymbol{x}_1, \boldsymbol{y}) \wedge nanc(\boldsymbol{x}_1, \boldsymbol{z}))) \quad (29)$$

By (29), $\textsc{Distr}_{\text{EI}}{}^{\rightarrow}$, $\textsc{Distr}_{\text{EI}}{}^{\wedge}$, MP,

$$\vdash_{g_0} ((\lceil c_1/x_1 \rceil dsep(\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{z}) \wedge \lceil c_1/x_1 \rceil nanc(\boldsymbol{x}_1, \boldsymbol{z}))) \rightarrow (\lceil c_1/x_1 \rceil nanc(\boldsymbol{x}_1, \boldsymbol{y}) \wedge \lceil c_1/x_1 \rceil nanc(\boldsymbol{x}_1, \boldsymbol{z})) \quad (30)$$

By (30), $\textsc{Nanc2}$, PT, MP,

$$\vdash_{g_0} ((\lceil c_1/x_1 \rceil dsep(\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{z}) \wedge \lceil c_1/x_1 \rceil nanc(\boldsymbol{x}_1, \boldsymbol{z}))) \rightarrow (nanc(\boldsymbol{x}_1, \boldsymbol{y}) \wedge nanc(\boldsymbol{x}_1, \boldsymbol{z})) \quad (31)$$

By (31), Nanc1, PT, MP,

$$\vdash_{g_0} ((\lceil \boldsymbol{c_1}/\boldsymbol{x}_1 \rceil dsep(\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{z}) \wedge \lceil \boldsymbol{c_1}/\boldsymbol{x}_1 \rceil nanc(\boldsymbol{x}_1, \boldsymbol{z}))) \\ \rightarrow ((f_1 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x}_1 \rceil (f_1 = \boldsymbol{y}|_{\boldsymbol{z}})) \quad (32)$$

By (32), (20), PT, MP,

$$\vdash_{g_0} ((\lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil \psi_{\mathrm{dp}}) \wedge \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil nanc(\boldsymbol{x}_1, \boldsymbol{z})) \\ \rightarrow ((f_1 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x}_1 \rceil (f_1 = \boldsymbol{y}|_{\boldsymbol{z}})) \quad (33)$$

By (33), (27), (19), PT, MP,

$$\vdash_{g_0} \psi_{\mathrm{pre}} \rightarrow ((f_1 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x}_1 \rceil (f_1 = \boldsymbol{y}|_{\boldsymbol{z}})) \quad (34)$$

By (34), (24), Eq2, PT, MP,

$$\vdash_{g_0} \psi_{\mathrm{pre}} \rightarrow ((f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x}_1 \rceil (f_0 = \boldsymbol{y}|_{\boldsymbol{z}})) \quad (35)$$

By Do2, PT,MP,

$$\vdash_{g_0 \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil} \psi_{\mathrm{do2}} \rightarrow (f_2 = \boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x_2} = \boldsymbol{c_2}} \leftrightarrow \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil f_2 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (36)$$

By (36), (25), PT, MP,

$$\vdash_{g_0 \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil} \psi_{\mathrm{dp}} \rightarrow (f_2 = \boldsymbol{y}|_{\boldsymbol{z}, \boldsymbol{x_2} = \boldsymbol{c_2}} \leftrightarrow \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil f_2 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (37)$$

By (37), (23), Eq2, PT, MP

$$\vdash_{g_0 \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil} \psi_{\mathrm{dp}} \rightarrow (f_0 = \boldsymbol{y}|_{\boldsymbol{z}} \leftrightarrow \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (38)$$

By (38), $\mathrm{DG_{EI}}$, $\mathrm{Distr_{EI}}^{\rightarrow}$, $\mathrm{Distr_{EI}}^{\wedge}$, PT, MP,

$$\vdash_{g_0} \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil \psi_{\mathrm{dp}} \rightarrow ((\lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (39)$$

By (39), (19), PT, MP,

$$\vdash_{g_0} \psi_{\mathrm{pre}} \rightarrow ((\lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (40)$$

By (35), (40), Eq2, PT, MP,

$$\vdash_{g_0} \psi_{\mathrm{pre}} \rightarrow ((f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (41)$$

By (41), $\mathrm{DG_{EI}}$, PT, MP,

$$\vdash_{g} (\lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil \psi_{\mathrm{pre}}) \\ \rightarrow ((\lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil \lceil \boldsymbol{c_1}/\boldsymbol{x_1} \rceil \lceil \boldsymbol{c_2}/\boldsymbol{x_2} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (42)$$

By (42), $\mathrm{Simul_{EI}}$, MP,

$$\vdash_{g} (\lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil \psi_{\mathrm{pre}}) \\ \rightarrow ((\lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \leftrightarrow \lceil \boldsymbol{c_0}/\boldsymbol{v}, \boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2} \rceil f_0 = \boldsymbol{y}|_{\boldsymbol{z}}) \quad (43)$$

By (43), Eq2, PT, MP,

$$\vdash_{g} (\lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil \psi_{\mathrm{pre}}) \\ \rightarrow ((\lceil \boldsymbol{c_0}/\boldsymbol{v} \rceil \varphi) \leftrightarrow \lceil \boldsymbol{c_0}/\boldsymbol{v}, \boldsymbol{c_1}/\boldsymbol{x_1}, \boldsymbol{c_2}/\boldsymbol{x_2} \rceil \varphi)$$

Therefore, Claim (3) follows. □