Canonically Gluing Causal Models
Along Shared Substructures

Markus Englberger[1] **Devendra Singh Dhami**[1]

[1]Department of Mathematics and Computer Science, TU Eindhoven

Abstract

Causal Bayesian Networks (CBNs) provide a framework for reasoning about interventions, but analyzing complex systems often requires working with multiple overlapping models or hierarchical levels of granularity. We study how to canonically merge separate causal models that share a common substructure or abstract representation. We formalize this problem using category theory. We construct two distinct categories of causal models where morphisms are defined by maps that commute with interventions. Depending on whether these maps are injective or surjective, they represent causal submodels or causal abstractions, respectively. We demonstrate that the canonical merging of models sharing a common subsystem corresponds to computing a pushout in the category of causal submodels, while merging models sharing a common abstraction corresponds to a pullback in the category of causal abstractions. We establish sufficient topological and domain-level conditions under which these universal constructions exist.

1 INTRODUCTION

Causal Bayesian Networks (CBNs) [Pearl, 1995, Druzdzel and Simon, 1993] provide a formal framework for representing causal relationships and reasoning about the effects of interventions. Researchers may possess separate models that describe different parts of a larger phenomenon, or they may work with models of the same system at different levels of granularity. This raises a fundamental question: *How can we canonically merge two causal models that share a common structure?*

As a motivating example, consider a simple model with two binary variables, treatment and recovery, describing whether a patient receives a treatment and whether they recover. There are several possible refinements of this scenario. One might introduce a biological mediator between treatment and recovery, refine the binary recovery variable into a more fine-grained outcome space, or add a second effect capturing potential side effects of the treatment. It is a priori not clear whether such distinct refinements can be canonically unified into a single detailed causal model.

In this paper, we address questions like this by investigating three specific merging scenarios:

1. First, we consider the case where two models share a common submodel. We ask whether there exists a consistent *union* model that respects the common causal submodel.
2. Second, we consider the case where two causal models M_1, M_2 both abstract to the same high-level causal model M_3. Here, we ask whether they can be combined into a single, detailed low-level model that refines both M_1 and M_2 while preserving their shared abstract behavior.
3. Third, we consider the case where two high-level causal models, M_1 and M_2, are both causal abstractions of the same complex underlying model M_0. We ask whether we can combine these perspectives to study the system as a single, unified high-level abstraction.

Diagrammatically, writing $\hookrightarrow$ for the inclusion of a causal submodel and $\twoheadrightarrow$ for a causal abstraction (both made precise in Section 4), the three scenarios correspond to the following shapes:

1. $M_1 \hookleftarrow M_0 \hookrightarrow M_2$
2. $M_1 \twoheadrightarrow M_3 \twoheadleftarrow M_2$
3. $M_1 \twoheadleftarrow M_0 \twoheadrightarrow M_2$

In each case, we ask whether there is a canonical fourth model P completing the shape to a commuting square — as

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:1474–1497.

a joint super-model in the first scenario, a joint refinement in the second, and a joint abstraction in the third.

Our goal is to find merging constructions that are *canonical* in the sense that they introduce no arbitrary information and satisfy a universal property. To achieve this, we adopt the language of category theory. Building on existing literature studying causal abstractions [Rubenstein et al., 2017, Beckers and Halpern, 2019, Geiger et al., 2025, Rischel, 2020], we formalize a category of causal submodels $\mathbf{CM_{sub}}$ and a category of causal abstractions $\mathbf{CM_{abs}}$. We can then formally identify the three merging operations as, respectively, a *pushout in the category of causal submodels*, a *pullback in the category of causal abstractions*, and a *pushout in the category of causal abstractions*.

Our main technical contributions are theorems establishing sufficient conditions for these constructions to exist; for the third construction, existence is unconditional. We derive criteria regarding graph topology and domain compatibility that guarantee the existence of the resulting pushouts and pullbacks.

2 RELATED WORK

Causal abstraction has been formalized in several ways. The *exact transformations* of Rubenstein et al. [2017] relate two models by a single map between their joint value spaces; Beckers and Halpern [2019] refine this into a hierarchy of τ-, strong, and *constructive* abstractions, the last requiring the map to factor over a partition of the low-level variables; and Rischel [2020] and Rischel and Weichwald [2021] work with a surjective node map together with fiber-wise value maps, termed α*-abstractions* in the review of Zennaro [2022]. Beckers et al. [2020] and Massidda et al. [2023] extend abstraction to approximate consistency and to soft interventions. Our causal abstraction morphisms are close to the strong constructive abstractions of Beckers and Halpern [2019] and to the exact abstractions of Rischel [2020]. Schooltink and Zennaro [2025] work out the precise correspondence between the α-, constructive-τ-, and graph-level notions. On the injective side, relations akin to our causal submodels appear in Otsuka and Saigo [2022] and in the *causal embeddings* of Schooltink and Zennaro [2026]; the latter are also used to merge overlapping datasets, but without a universal property singling out the merged model. Geiger et al. [2025] develop causal abstraction as a foundation for mechanistic interpretability; in that setting, Pîslar et al. [2025] combine several abstractions of the same network into a more faithful one, similar to our third merging scenario, with a learned rather than canonical combination.

Categorical treatments of Bayesian networks and interventions go back to Fong [2012] and Jacobs et al. [2019]. Englberger and Dhami [2025] model individual causal models as functors between Markov categories and abstractions as natural transformations; merging is not addressed there. Here, instead, the CBNs themselves are the objects of a category, and merging becomes a (co)limit. D'Acunto and Battiloro [2025] arrange structural causal models into a functor category and use network (co)sheaves to transfer causal knowledge across interacting models.

Zhu et al. [2024] learn causal abstractions and observe that many consistent abstractions of a low-level model typically coexist; such families of abstractions motivate the kind of overlapping-abstraction setting addressed by our pushout construction. In causal discovery, structures learned over overlapping variable sets are combined by intersecting their constraints [Tillman et al., 2008, Triantafillou and Tsamardinos, 2015], which yields equivalence classes of graphs rather than a single merged model. The principle of independent causal mechanisms [Peters et al., 2017] states that the mechanisms of a system do not inform or influence one another; *independent mechanism analysis* [Gresele et al., 2021] formalizes this as an orthogonality condition on the generative mechanism. Our results reflect this principle: canonical merging succeeds when the refinements contributed by the two models do not overlap.

3 BACKGROUND

In this section, we provide some background on the required category-theoretical concepts as well as on CBNs.

3.1 CATEGORY-THEORETIC BACKGROUND

This subsection briefly recalls the two universal constructions used throughout the paper, pullbacks and pushouts, illustrated in the category **Set** of sets (whose objects are sets and whose morphisms are functions). The general definition of a category, together with the related notions of products and coproducts and a remark on how to read commuting diagrams, is recalled in Section 7.1. Readers familiar with category theory may safely skip this section.

Definition 1. *Given morphisms $f : A \to C$ and $g : B \to C$, their* ***pullback*** *consists of an object P and morphisms $p_A : P \to A$ and $p_B : P \to B$ such that $f \circ p_A = g \circ p_B$. Further, it satisfies a universal property: for any object X equipped with morphisms $q_A : X \to A, q_B : X \to B$ such that $f \circ q_A = g \circ q_B$, there exists a unique morphism $u : X \to P$ (dashed) such that $p_A \circ u = q_A$ and $p_B \circ u = q_B$:*

$$\begin{tikzcd}
X \arrow[drr, bend left, "q_B"] \arrow[dr, dashed, "u"] \arrow[ddr, bend right, "q_A"'] & & \\
& P \arrow[r, "p_B"] \arrow[d, "p_A"'] & B \arrow[d, "g"] \\
& A \arrow[r, "f"'] & C
\end{tikzcd}$$

Example 1. In **Set**, the pullback is the subset

$$A \times_C B = \{(a, b) \in A \times B \mid f(a) = g(b)\}.$$

Definition 2. *Dually, given morphisms $f : C \to A$ and $g : C \to B$, their **pushout** consists of an object Q and morphisms $q_A : A \to Q$ and $q_B : B \to Q$ such that $q_A \circ f = q_B \circ g$. Furthermore, it satisfies a universal property: for any object Y equipped with morphisms $r_A : A \to Y, r_B : B \to Y$ such that $r_A \circ f = r_B \circ g$, there exists a unique morphism $v : Q \to Y$ (dashed) such that $v \circ q_A = r_A$ and $v \circ q_B = r_B$:*

$$\begin{tikzcd}
C \arrow[r, "g"] \arrow[d, "f"'] & B \arrow[d, "q_B"'] \arrow[ddr, bend left, "r_B"] & \\
A \arrow[r, "q_A"] \arrow[drr, bend right, "r_A"'] & Q \arrow[dr, dashed, "v"] & \\
& & Y
\end{tikzcd}$$

Example 2. In **Set**, the pushout is obtained from the disjoint union $A \sqcup B$ by identifying $f(c) \sim g(c)$ for all $c \in C$. Intuitively, pushouts describe gluing sets along a shared part.

3.2 CAUSAL BAYESIAN NETWORKS

Causal Bayesian networks provide a formal framework for representing causal relationships between random variables. They extend ordinary Bayesian networks by equipping them with an explicit semantics for interventions. In this subsection, we briefly recall the standard definition and notation used throughout the paper.

Definition 3. *A* Bayesian network M *consists of:*

- *a finite directed acyclic graph (DAG) $G_M = (V_M, E_M)$ with node set V_M and edge set E_M, whose nodes represent random variables,*
- *for each node $v \in V_M$, a discrete random variable with a discrete domain $\mathrm{dom}_M(v)$,*
- *for each node $v \in V_M$, local mechanisms*

  $$p_M(v \mid \mathrm{pa}_M(v)),$$

  where $\mathrm{pa}_M(v)$ denotes the set of parents of v in G_M. Formally, a local mechanism is a function

  $$\mathrm{dom}_M(\mathrm{pa}_M(v)) \longrightarrow \Delta\big(\mathrm{dom}_M(v)\big),$$

  where $\Delta(X)$ denotes the probability simplex over the finite set X, i.e., the set of probability distributions on X.

Together, these data define a joint probability distribution over the variables $v \in V_M$ via the factorization

$$p_M(V_M) = \prod_{v \in V_M} p_M(v \mid \mathrm{pa}_M(v)).$$

Remark 3.1. We strictly limit our scope to discrete variables throughout this paper to avoid the measure-theoretic complications.

Throughout, we denote *sets* of variables by boldface letters (e.g., $\mathbf{v} \subseteq V_M$) to distinguish them from individual variables (e.g., $v \in V_M$). For $\mathbf{v} \subseteq V_M$ we write $\mathrm{dom}_M(\mathbf{v}) := \prod_{u \in \mathbf{v}} \mathrm{dom}_M(u)$, and we identify $\mathrm{dom}(\{v\}) \cong \mathrm{dom}(v)$ for a single variable v. When the model M is clear from context, we drop the subscript and simply write $G = (V, E)$, $\mathrm{dom}(v)$, $\mathrm{pa}(v)$, and p; we write p^M when we want to emphasize to which model a distribution belongs.

A *causal Bayesian network* (CBN) augments a Bayesian network with a causal interpretation of the directed edges. An edge $u \to v$ is interpreted as indicating that u is a direct cause of v. This causal interpretation allows one to reason not only about observational distributions, but also about the effects of interventions. In a CBN, the joint distribution is assumed to arise from an underlying data-generating process compatible with the graph structure, and the factorization is interpreted as reflecting causal mechanisms. For any CBN, we assume causal sufficiency, i.e., the absence of unobserved confounders.

An *intervention* replaces the causal mechanism of one or more variables by fixing their values. Formally, for a subset $\mathbf{v} \subseteq V_M$ and an assignment $a \in \mathrm{dom}_M(\mathbf{v})$, the intervention $\mathrm{do}(\mathbf{v} = a)$ modifies the network by removing all incoming edges into nodes in $\mathbf{v}$ and setting $u = a_u$ for each $u \in \mathbf{v}$ where a_u is the projection of a to the domain of u. The resulting intervened distribution, denoted $p_{\mathrm{do}(\mathbf{v}=a)}$, is given by the truncated factorization

$$p_{\mathrm{do}(\mathbf{v}=a)}(V_M) = \prod_{u \notin \mathbf{v}} p(u \mid \mathrm{pa}(u)) \cdot \prod_{u \in \mathbf{v}} \mathbb{I}[u = a_u],$$

where $\mathbb{I}[\cdot]$ denotes the indicator function.

A causal Bayesian network thus gives rise to a family of interventional distributions

$$\big\{p_{\mathrm{do}(\mathbf{v}=a)} \bigm| \mathbf{v} \subseteq V_M,\ a \in \mathrm{dom}_M(\mathbf{v})\big\}.$$

These distributions encode the causal behavior of the system under all possible interventions of interest. We will require minimality of the underlying DAG, i.e., no edge can be removed from G_M such that the resulting model still gives rise to the same family of interventional distributions.

4 THE CATEGORY OF CAUSAL BAYESIAN NETWORKS

We can relate two CBNs by defining a map on nodes and on domains that respects the causal structure and that commutes with interventions:

Definition 4 (Causal map)**.** *Consider two CBNs M_1, M_2. A **Causal map** between CBNs*

$$(\eta, \tau) : M_1 \longrightarrow M_2$$

consists of

1. *a node map $\eta : V_{M_1} \to V_{M_2}$*
2. *a domain map $\tau : \mathrm{dom}_{M_1}(V_{M_1}) \to \mathrm{dom}_{M_2}(\eta(V_{M_1}))$ compatible with η in the sense that it factors fiberwise as*
$$\tau \;=\; \prod_{v_2 \in \eta(V_{M_1})} \tau_{v_2}$$
where for each $v_2 \in \eta(V_{M_1})$ we have a component map $\tau_{v_2} : \mathrm{dom}_{M_1}(\eta^{-1}(v_2)) \to \mathrm{dom}_{M_2}(v_2)$.[1] *For a set $\mathbf{v} \subseteq \eta(V_{M_1})$ we will write*
$$\tau_{\mathbf{v}} := \prod_{u \in \mathbf{v}} \tau_u$$
3. *and such that τ commutes with interventions: for all $\mathbf{v} \subseteq \eta(V_{M_1})$ (including $\mathbf{v} = \emptyset$, i.e., the observational case) and $a \in \mathrm{dom}_{M_1}(\eta^{-1}(\mathbf{v}))$:*
$$\tau_* \left(p^{M_1}_{\mathrm{do}(\eta^{-1}(\mathbf{v})=a)}(V_{M_1}) \right) = p^{M_2}_{\mathrm{do}(\mathbf{v}=\tau_{\mathbf{v}}(a))}(\eta(V_{M_1})).$$
where τ_* is the pushforward of (discrete) probability distributions along τ, i.e., $(\tau_* p)(y) := \sum_{x \in \tau^{-1}(y)} p(x)$. Explicitly, the condition means that for any $y \in \mathrm{dom}_{M_2}(\eta(V_{M_1}))$:
$$\sum_{x \in \tau^{-1}(y)} p^{M_1}_{\mathrm{do}(\eta^{-1}(\mathbf{v})=a)}(x) = p^{M_2}_{\mathrm{do}(\mathbf{v}=\tau_{\mathbf{v}}(a))}(y)$$

Such causal maps between CBNs do, however, not equip the family of CBNs with the structure of a category, as causal maps are generally not composable: Given two causal maps $(\eta, \tau) : M_1 \to M_2, (\eta', \tau') : M_2 \to M_3$, the composition $(\eta' \circ \eta, \tau' \circ \tau)$ may not constitute a causal map, as $\tau' \circ \tau$ does not generally have to allow a fiberwise factorization. The following example illustrates this failure and the intuition behind it.

Example 3 (Failure of fiberwise factorization under composition). Consider modeling the incomes of two individuals: let M_1 contain a single variable I_A (the income of A), M_2 the variables I_A, I_B, and M_3 a single variable S for the *total* income, with fiberwise map $\tau'_S(i_A, i_B) = i_A + i_B$. The inclusion $M_1 \to M_2$ is an injective causal map (a submodel) and $M_2 \to M_3$ a surjective one (an abstraction), but the composite sends $I_A \mapsto S$ with fiber $\{I_A\}$, so a fiberwise factorization would require a map $\mathrm{dom}(I_A) \to \mathrm{dom}(S)$ computing the total income from A's income alone, which is impossible without knowing the income of B, which lies outside the image of M_1. Hence the composition is not a causal map.

Such compositions are composable in the two special cases of restricting either to injective or to surjective node and domain maps η, τ. In the former case, where both η and τ are injections, we can view M_1 as a **submodel** of M_2. In the latter case of η and τ being surjections, M_2 can be interpreted as describing the causal model M_1 on a higher-abstraction model, and we will refer to these as **causal abstractions**.

We can now define the categories of **causal submodels** and **causal abstractions**:

Definition 5 (Category of causal submodels)**.** *The **category of causal submodels**, denoted $\mathbf{CM}_{\mathrm{sub}}$, has as objects causal Bayesian networks. A **causal submodel morphism** (or simply **causal submodel**) is a Causal Map*

$$(\eta, \tau) : M_1 \longrightarrow M_2$$

such that η and τ are injective and such that $\eta : V_{M_1} \hookrightarrow V_{M_2}$ preserves edges, i.e., $\forall x, y \in V_{M_1}$:

$$x \to y \Rightarrow \eta(x) \to \eta(y)$$

Additionally, η has to respect causal sufficiency: for all $u, v \in V_{M_1}$ with $\eta(u) \neq \eta(v)$ and all $x \in V_{M_2}$ with $x \notin \{\eta(u), \eta(v)\}$, the existence of directed paths $x \rightsquigarrow \eta(u)$ and $x \rightsquigarrow \eta(v)$ that share no node other than x (i.e., x being a confounder of $\eta(u)$ and $\eta(v)$) implies that x is in the image of η, compatibly with the paths: there exists $x' \in V_{M_1}$ such that $\eta(x') = x$, together with paths $x' \rightsquigarrow u$ and $x' \rightsquigarrow v$ in M_1 sharing no node other than x'.

Definition 6 (Category of causal abstractions)**.** *An* ordered CBN *is a pair $(M, \leq_M)$ consisting of a CBN M and a total order $\leq_M$ on V_M that is topological for E_M (i.e., $u \to v \in E_M$ implies $u <_M v$); since every DAG admits a topological order, every CBN can be so equipped. The **category of causal abstractions**, denoted $\mathbf{CM}_{\mathrm{abs}}$, has as objects ordered CBNs. A **causal abstraction morphism** is a Causal Map*

$$(\eta, \tau) : M_1 \longrightarrow M_2$$

between the underlying CBNs with η and τ being surjective, with η monotone, i.e., $u \leq_1 v \Rightarrow \eta(u) \leq_2 \eta(v)$, and such that η reflects edges, i.e., $\forall x, y \in V_{M_1}$:

$$\eta(x) \to \eta(y) \Rightarrow \exists x' \in \eta^{-1}(\eta(x)), y' \in \eta^{-1}(\eta(y)) \text{ such that } x' \to y' \tag{4.1}$$

When no confusion can arise, we suppress the order from the notation and simply write M for $(M, \leq_M)$.

[1] The components τ_{v_2} are indexed only by nodes in the *image* $\eta(V_{M_1})$; nothing is assigned to nodes of M_2 outside it. The value τ assigns to v_2 depends only on the values on the fiber $\eta^{-1}(v_2)$, i.e., $\tau(x) = (\tau_{v_2}(x_{\eta^{-1}(v_2)}))_{v_2 \in \eta(V_{M_1})}$. This locality is what makes the intervention commutation in point 3 meaningful: intervening on v_2 constrains exactly the variables in its fiber.

In contrast to $\mathbf{CM}_{\mathrm{sub}}$, the objects of $\mathbf{CM}_{\mathrm{abs}}$ thus carry a total order; this is to avoid problems with cycles in the constructions of Theorem 2 and Theorem 3; Example 8 illustrates the problem, see also Remark 7.2.

The asymmetry between the two definitions — edges *preserved* for submodels, *reflected* for abstractions — is explained in Remark 7.3 (appendix).

To prove that the two defined objects actually are categories, we have to show compositionality:

Proposition 4.1. $\mathbf{CM}_{\mathrm{sub}}$ *and* $\mathbf{CM}_{\mathrm{abs}}$ *are well-defined categories. In particular, composition is well-defined.*

Proof. See Section 7.6. $\square$

In Remark 7.6 (appendix) we discuss in which sense these two categories are *maximal* among classes of composable causal maps.

5 MERGING CAUSAL MODELS SHARING A COMMON SUBSTRUCTURE

In this section, we study in which cases one can canonically merge two CBNs that either share a common causal submodel, abstract to the same high-level causal abstraction, or that are both abstractions from the same low-level CBN. These are three of the four possible pushout/pullback constructions across our two categories; the remaining one, pullbacks in $\mathbf{CM}_{\mathrm{sub}}$ (intersecting two submodels of a common model), is sketched in the appendix (Section 7.3).

5.1 MERGING MODELS SHARING A COMMON CAUSAL SUBMODEL

Assume we have two CBNs M_1, M_2 and we know that they share a common causal submodel M_0, that is they overlap on some part of the underlying graph. We want to know under which conditions we can canonically merge these two models such that the shared causal submodel is respected. More explicitly, we want to find a causal model P that has both M_1 and M_2 as causal submodels such that first including M_0 into M_1 and then M_1 into P commutes with first including M_0 into M_2 and then M_2 into P. Additionally, we want P to be canonical in the sense that every other CBN satisfying these conditions should itself have P as a causal submodel. This is exactly captured by the categorical notion of a pushout (see Definition 2) in the category $\mathbf{CM}_{sub}$. Diagrammatically, we seek a CBN P together with submodel morphisms g_1, g_2 (dashed) making the square

$$\begin{tikzcd} M_0 \arrow[r, hook, "f_2"] \arrow[d, hook, "f_1"'] & M_2 \arrow[d, hook, dashed, "g_2"] \\ M_1 \arrow[r, hook, dashed, "g_1"'] & P \end{tikzcd}$$

commute, such that P is universal among all CBNs with this property. The following theorem provides a sufficient condition for when such a canonical merging operation is possible:

Theorem 1 (Pushouts in $\mathbf{CM}_{sub}$)**.** *Assume two CBNs M_1 and M_2 share a common submodel M_0, witnessed by two morphisms in $\mathbf{CM}_{sub}$:*

$$M_0 \xrightarrow{f_1=(\eta_1,\tau_1)} M_1, \quad M_0 \xrightarrow{f_2=(\eta_2,\tau_2)} M_2$$

Consider the following graph $G_P = (V_P, E_P)$:

- ***Nodes** (V_P)**:** Define $V_P := (V_{M_1} \sqcup V_{M_2})/\sim$, identifying $\eta_1(v) \sim \eta_2(v)$ for all $v \in V_{M_0}$. Let $j_1 : V_{M_1} \to V_P$ and $j_2 : V_{M_2} \to V_P$ be the canonical injections.*
- ***Edges** (E_P)**:** An edge $u \to v$ exists in P if there exists an edge $u_1 \to v_1$ in M_1 such that $j_1(u_1) = u, j_1(v_1) = v$, or an edge $u_2 \to v_2$ in M_2 such that $j_2(u_2) = u, j_2(v_2) = v$.*

Assume the following conditions hold:

1. *The graph G_P is acyclic and j_1, j_2 preserve causal sufficiency as required in Definition 5.*
2. *For all $v \in V_{M_i}$, v has positive probability, i.e., supp$(p^{M_i})(v) = \mathrm{dom}_{M_i}(v)$ for both $i = 1$ and $i = 2$.*
3. *For any $v_1 \in V_{M_1} \setminus \eta_1(V_{M_0})$, $v_2 \in V_{M_2} \setminus \eta_2(V_{M_0})$:*
    - *at least one of the following holds:* (i) $\exists u_1, u_2 \in V_{M_0}$ *such that there exist directed paths $v_1 \rightsquigarrow \eta_1(u_1)$ and $v_2 \rightsquigarrow \eta_2(u_2)$; or* (ii) $\exists u \in V_{M_0}$ *such that there exist directed paths $v_1 \rightsquigarrow \eta_1(u), \eta_2(u) \rightsquigarrow v_2$ or directed paths $v_2 \rightsquigarrow \eta_2(u), \eta_1(u) \rightsquigarrow v_1$*
    - *$\nexists u \in V_{M_0}$ such that both $v_1 \to \eta_1(u) \in E_{M_1}$ **and** $v_2 \to \eta_2(u) \in E_{M_2}$*

Then, the pushout object P exists in $\mathbf{CM}_{sub}$ and is explicitly constructed as follows:

- ***Graph:** The underlying DAG is given by G_P*
- ***Domains** (dom_P)**:** Let $u \in V_P$.*
    - *If $u = j_1(u_1)$ for $u_1 \in V_{M_1} \setminus \eta_1(V_{M_0})$, set* $\mathrm{dom}_P(u) := \mathrm{dom}_{M_1}(u_1)$.
    - *If $u = j_2(u_2)$ for $u_2 \in V_{M_2} \setminus \eta_2(V_{M_0})$, set* $\mathrm{dom}_P(u) := \mathrm{dom}_{M_2}(u_2)$.

– *If u is an overlap node, i.e.,*

$$\exists v \in V_{M_0} : u = (j_1 \circ \eta_1)(v) = (j_2 \circ \eta_2)(v)$$

let $\mathrm{dom}_P(u)$ *be the pushout*

$$\mathrm{dom}_P(u) := \mathrm{dom}_{M_1}(\eta_1(v)) \sqcup \mathrm{dom}_{M_2}(\eta_2(v)) / \sim$$

identifying $\tau_{1,\eta_1(v)}(a) \sim \tau_{2,\eta_2(v)}(a)$ *for* $a \in \mathrm{dom}_{M_0}(v)$.

Let θ_1, θ_2 be the canonical domain injections.

- ***Local Mechanisms:*** *The conditional probability distributions are defined via the domain maps θ_1, θ_2. For any $u \in V_P$: If $u = j_1(u_1)$ for $u_1 \in V_{M_1}$, define for $a \in \mathrm{dom}_{M_1}(u_1)$, $b \in \mathrm{dom}_{M_1}(\mathrm{pa}_{M_1}(u_1))$:*

$$p^P\left(u = \theta_{1,u}(a) \mid j_1(\mathrm{pa}_{M_1}(u_1)) = \theta_{1,\mathrm{pa}_{M_1}(u_1)}(b)\right) := p^{M_1}\left(u_1 = a \mid \mathrm{pa}_{M_1}(u_1) = b\right)$$

Define analogously for $u = j_2(u_2)$ for $u_2 \in V_{M_2}$.

The model P, equipped with morphisms $g_1 = (j_1, \theta_1)$ and $g_2 = (j_2, \theta_2)$, is the pushout of f_1 and f_2.

Proof. See Appendix Section 7.6. □

Briefly, the two bullets of condition 3 rule out, respectively, disconnected fragments whose causal relationship the merge would have to invent (Example 6), and incompatible parent sets of a shared node (Example 5); see Remark 7.4 (appendix) for a discussion of necessity. Example 13 (appendix) shows why condition 2 cannot be dropped.

Example 4 (Merging Causal Chains via Pushout). We illustrate Theorem 1 with a simple scenario where we merge an upstream causal model with a downstream model via a shared mediator.

Let M_0 be a causal model consisting of a single node Y with no edges.

$$V_0 = \{Y\}, \quad E_0 = \emptyset$$

Let M_1 extend M_0 by adding a cause X.

$$V_{M_1} = \{X, Y\}, \quad E_{M_1} = \{X \to Y\}$$

Let M_2 extend M_0 by adding an effect Z.

$$V_2 = \{Y, Z\}, \quad E_{M_2} = \{Y \to Z\}$$

The morphisms $f_1 : M_0 \to M_1$, $f_2 : M_0 \to M_2$ are simply the inclusion mapping $Y \mapsto Y$.

The Pushout (P): The pushout object P is the union of the graphs glued at Y:

$$V_P = \{X, Y, Z\}, \quad E_P = \{X \to Y, Y \to Z\}$$

Visual Representation:

$$\begin{tikzcd}
 & M_0 : \boxed{Y} \arrow[dl, hook, "f_1"'] \arrow[dr, hook, "f_2"] & \\
M_1 : \boxed{X \to Y} \arrow[dr, hook, "g_1"'] & & M_2 : \boxed{Y \to Z} \arrow[dl, hook', "g_2"] \\
 & P : \boxed{X \to Y \to Z} &
\end{tikzcd}$$

The local mechanisms $p^P(Y|X)$, $p^P(Z|Y)$ are given by $p^{M_1}(Y|X)$, $p^{M_2}(Z|Y)$, respectively.

The resulting model P represents the full causal chain $X \to Y \to Z$, satisfying the universal property of the pushout in $\mathbf{CM}_{sub}$. Note that the graph of the pushout is canonical: for example, we could not add an edge $X \to Z$ as this would make X a confounder with regard to Y and Z and therefore violate the requirement that the inclusion $g_2 : M_2 \to P$ preserves causal sufficiency. This is the motivation behind the first bullet point of condition 3 in Theorem 1.

Condition 3 holds for the only relevant pair $(v_1, v_2) = (X, Z)$: the first bullet via case *(ii)* with $u = Y$ (directed paths $X \rightsquigarrow \eta_1(Y)$ and $\eta_2(Y) \rightsquigarrow Z$), and the second bullet trivially.

Example 5 (Counter-Example: Underdefined Mechanism). We illustrate a case where a canonical merging of two causal models is not possible because the local mechanism of the merged graph would be underdefined. This scenario specifically violates the third condition in Theorem 1.

Let the shared submodel M_0 be a simple causal model consisting of a single node C with no edges.

- $V_0 = \{C\}$, $E_0 = \emptyset$.

Let M_1 extends M_0 by introducing a cause A that directly influences C.

- $V_{M_1} = \{A, C\}$, $E_{M_1} = \{A \to C\}$.
- Morphism f_1: Inclusion mapping $C \mapsto C$.

Let M_2 extend M_0 by introducing a different cause B that also directly influences C.

- $V_{M_2} = \{B, C\}$, $E_{M_2} = \{B \to C\}$.
- Morphism f_2: Inclusion mapping $C \mapsto C$.

Failure of the Pushout Construction: If we were to naively merge these graphs by gluing them at node C, we would obtain a graph with edges $A \to C \leftarrow B$. To be a valid CBN, the merged model P would require a conditional probability distribution $p^P(C \mid A, B)$. However, M_1 only provides the marginal mechanism $p^{M_1}(C \mid A)$ and M_2 only provides $p^{M_2}(C \mid B)$. Without additional assumptions, there is no canonical way to combine these into a joint mechanism $p^P(C \mid A, B)$.

Example 6 (Counter-Example: Undetermined Dependence). Let the shared submodel M_0 consist of a single node A, and let M_1 and M_2 extend it by two different effects, $E_{M_1} = \{A \to B\}$ and $E_{M_2} = \{A \to C\}$, say, two clinical studies of one drug A's effect on blood pressure B and on heart rate C. The first bullet of condition 3 fails for the pair (B, C). The glued fork $B \leftarrow A \to C$ is a consistent merged model, but it declares B and C conditionally independent given A, information contained in neither study; a merged model keeping them dependent (e.g., via an edge $B \to C$) is equally consistent, and the fork admits no morphism into it; no candidate is universal, so the pushout does not exist.

5.2 MERGING CAUSAL MODELS THAT SHARE A CAUSAL ABSTRACTION

Assume we have two CBNs M_1, M_2 and we know that they abstract to the same higher-level causal model M_3, that is there exist causal abstractions $f_1 : M_1 \to M_3, f_2 : M_2 \to M_3$. We want to know under which conditions we can canonically merge these two models on the lower level, such that the shared causal abstraction is respected. More explicitly, we want to find a causal model P that abstracts to both M_1 and M_2, such that first abstracting from P to M_1 and then to M_3 commutes with first abstracting from P to M_2 and then to M_3. Additionally, we want P to be canonical in the sense that it should be the unique, smallest causal model that satisfies this condition.

This is exactly captured by the categorical notion of a pullback (see Definition 1) in the category $\mathbf{CM}_{abs}$. Diagrammatically, we seek a CBN P together with abstraction morphisms p_1, p_2 (dashed) making the square

$$\begin{tikzcd} P \arrow[r, dashed, two heads, "p_2"] \arrow[d, dashed, two heads, "p_1"'] & M_2 \arrow[d, two heads, "f_2"] \\ M_1 \arrow[r, two heads, "f_1"'] & M_3 \end{tikzcd}$$

commute, such that P is universal among all CBNs with this property. The following Theorem gives a sufficient condition under which such a canonical merging operation exists.

Theorem 2 (Pullbacks in $\mathbf{CM}_{abs}$)**.** *Let $f_1 = (\eta_1, \tau_1) : (M_1, \leq_1) \to (M_3, \leq_3)$ and $f_2 = (\eta_2, \tau_2) : (M_2, \leq_2) \to (M_3, \leq_3)$ be morphisms in $\mathbf{CM}_{abs}$, and assume that for every node $c \in V_{M_3}$ there exists an index $i(c) \in \{1, 2\}$ such that the fiber $\eta_{i(c)}^{-1}(c)$ is a singleton and the fiber domain map $\tau_{i(c),c} : \mathrm{dom}_{M_{i(c)}}(\eta_{i(c)}^{-1}(c)) \to \mathrm{dom}_{M_3}(c)$ is a bijection. Write $k(c) \neq i(c)$ for the* detailed *side at c. Define an ordered CBN $(P, \leq_P)$ as follows.*

- ***Nodes:*** *$V_P := V_{M_1} \times_{V_{M_3}} V_{M_2} = \{(v_1, v_2) : \eta_1(v_1) = \eta_2(v_2)\}$, with projections π_1, π_2 and $\pi := \eta_1\pi_1 = \eta_2\pi_2$. For $u \in V_P$ write $k(u) := k(\pi(u))$.*
- ***Order:*** *$u <_P w$ iff $\pi(u) <_3 \pi(w)$, or $\pi(u) = \pi(w)$ and $\pi_{k(u)}(u) <_{k(u)} \pi_{k(u)}(w)$. order.)*
- ***Initial edges:*** *$u \to w \in E_{pr}$ iff $\pi_{k(w)}(u) \to \pi_{k(w)}(w) \in E_{M_{k(w)}}$. In particular $\mathrm{pa}_{pr}(u) = \pi_{k(u)}^{-1}\big(\mathrm{pa}_{M_{k(u)}}(\pi_{k(u)}(u))\big)$: the preliminary parents of u are the* full *$\pi_{k(u)}$-fibers over the detailed-side parents.*
- ***Domains:*** *$\mathrm{dom}_P(u) := \mathrm{dom}_{M_{k(u)}}(\pi_{k(u)}(u))$. This realizes the canonical isomorphisms $\mathrm{dom}_P(\pi^{-1}(c)) \cong \mathrm{dom}_{M_1}(\eta_1^{-1}(c)) \times_{\mathrm{dom}_{M_3}(c)} \mathrm{dom}_{M_2}(\eta_2^{-1}(c))$ of the fiber product of the joint fiber maps $\tau_{1,c}, \tau_{2,c}$. Let θ_1, θ_2 be the pullback projections; concretely, θ_k restricts to the identity on detailed-side fibers, and over the singleton-side node of c (say $i(c) = 1$) it is the joint map $\theta_{1,c} := \tau_{1,c}^{-1} \circ \tau_{2,c} : \mathrm{dom}_{M_2}(\eta_2^{-1}(c)) \to \mathrm{dom}_{M_1}(\eta_1^{-1}(c))$ on the whole π_1-fiber. These maps factor fiberwise over π_1, π_2 in the sense of Definition 4.*
- ***Mechanisms:*** *for $u \in V_P$ with $k := k(u)$, $v_k := \pi_k(u)$, $a \in \mathrm{dom}_P(u)$ and $b \in \mathrm{dom}_P(\mathrm{pa}_{pr}(u))$,*

$$\begin{aligned} &p^P\big(u = a \mid \mathrm{pa}_{pr}(u) = b\big) := \\ &p^{M_k}\big(v_k = a \mid \mathrm{pa}_{M_k}(\pi_k(u)) = \theta_{k,\mathrm{pa}_{M_k(v_k)}}(b)\big), \end{aligned} \tag{5.1}$$

- ***Final edges:*** *$E_P \subseteq E_{pr}$ retains exactly the edges required by the minimality condition of Definition 3: an edge is removed if Eq.* (5.1) *is constant in the corresponding parent coordinate given the others.*

Then $(P, \leq_P)$ is a well-defined ordered CBN, the projections $p_1 = (\pi_1, \theta_1)$ and $p_2 = (\pi_2, \theta_2)$ are abstraction morphisms with $f_1 p_1 = f_2 p_2$, and $(P, \leq_P; p_1, p_2)$ is the pullback of f_1, f_2 in $\mathbf{CM}_{abs}$.

Proof. See appendix Section 7.6. ☐

Example 7. We consider a case where one model (M_1) refines the internal causal structure of a variable A into a cascade, while the other (M_2) refines the effect B into parallel outcomes. As a concrete reading, let M_3 record that *drug intake* (A) affects *patient outcome* (B): one researcher refines the *cause* into its biological stages, absorption A_1 followed by metabolism A_2, while another refines the *effect* into the parallel outcomes efficacy B_1 and side effects B_2. The pullback canonically combines the two refinements into a single model.

The abstract model M_3 is given by the following graph:

$$M_3 \quad A \longrightarrow B$$

Model M_1 decomposes A into A_1, A_2 via a surjective map $\tau_{1,A} : \mathrm{dom}(\{A_1, A_2\}) \to \mathrm{dom}(A)$:

$$M_1 \quad A_1 \longrightarrow A_2 \longrightarrow B$$

Model M_2 decomposes B into B_1, B_2 via a surjective map $\tau_{2,B} : \mathrm{dom}(\{B_1, B_2\}) \to \mathrm{dom}(B)$.

$$\begin{tikzcd}[row sep=tiny] & & B_1 \\ M_2 & A \arrow[ur] \arrow[dr] & \\ & & B_2 \end{tikzcd}$$

The conditions of Theorem 2 are met and the Pullback P factorizes over the following DAG:

Visual Representation:

$$\begin{tikzcd}[row sep=tiny] & & & B_1 \\ P & A_1 \arrow[urr] \arrow[r] \arrow[drr] & A_2 \arrow[ur] \arrow[dr] & \\ & & & B_2 \end{tikzcd}$$

The local mechanism at $B_1 \in V_P$ is given as follows: $\forall b_1 \in \mathrm{dom}(B_1), (a_1, a_2) \in \mathrm{dom}(\{A_1, A_2\})$

$$p^P(b_1|(a_1, a_2)) = p^{M_2}(b_1|\tau_{1,A}((a_1, a_2)))$$

It may not be intuitive at first why the edges $A_1 \to B_1, A_1 \to B_2$ are included in P as there is no edge $A_1 \to B$ in M_1. However, knowing that B is independent of A_1 given A_2 does not in general imply that a refinement of B given by $B_1 \times B_2$ is also independent of A_1 given A_2.

Example 14 (appendix) gives a concrete counterexample as to why the bijectivity requirement of Theorem 2 cannot be dropped: if both models refine the domain of the *same* node in genuinely different ways, the merged mechanism is underdetermined.

The following example shows that the total order on the objects of $\mathbf{CM}_{\mathrm{abs}}$ cannot be dropped:

Example 8 (Counter-Example: Opposite Causal Directions). Let M_3 have two nodes A, B and no edge, and let M_1 (edge $A \to B$) and M_2 (edge $B \to A$) refine the domain of B resp. A so that the refined effect is averaged out (Remark 7.3): e.g., $B \in \{0, 1, 2, 3\}$ in M_1 with $\tau_{1,B}(b) = b \bmod 2$, and $B \mid A{=}0$ uniform on $\{0, 1\}$, $B \mid A{=}1$ uniform on $\{2, 3\}$. All fibers are singletons and the bijectivity condition of Theorem 2 holds, but the fiber-product construction would lift $A \to B$ from M_1 and $B \to A$ from M_2, a directed cycle: the two refinements attribute *opposite causal directions* to (A, B). The order excludes such contradictory cospans: monotonicity of f_1 forces $A <_3 B$, that of f_2 forces $B <_3 A$, so f_1 and f_2 cannot both be morphisms into the same ordered M_3.

5.3 MERGING MODELS THAT BOTH ARE ABSTRACTED FROM A LARGER MODEL

Consider a potentially complex CBN M_0 which we study at a higher-abstraction level. To this end, we are given two high-level causal models M_1 and M_2 that both are causal abstractions of M_0, i.e., there exist morphisms $f_1 : M_0 \to M_1, f_2 : M_0 \to M_2$ in the category of causal abstractions. Instead of studying two separate high-level abstractions, we want to combine them and study the system as a single high-level abstraction: we seek a CBN P that is a causal abstraction of both M_1 and M_2, such that first abstracting from M_0 to M_1 and then to P commutes with first abstracting from M_0 to M_2 and then to P. Additionally, we want P to be canonical in the sense that it should be the unique, largest CBN that satisfies these two conditions. This is exactly captured by the categorical notion of a pushout (see Definition 2) in the category of causal abstractions CM_{abs}. Diagrammatically, we seek a CBN P together with abstraction morphisms h_1, h_2 (dashed) making the square

$$\begin{tikzcd} M_0 \arrow[r, two heads, "f_2"] \arrow[d, two heads, "f_1"'] & M_2 \arrow[d, two heads, dashed, "h_2"] \\ M_1 \arrow[r, two heads, dashed, "h_1"'] & P \end{tikzcd}$$

commute, such that P is universal among all CBNs with this property.

The following theorem shows that such a pushout always exists, and constructs it explicitly.

Theorem 3 (Pushouts in CM_{abs})**.** *Let $(M_0, \leq_0)$, $(M_1, \leq_1)$, $(M_2, \leq_2)$ be ordered CBNs and let $f_1 = (\eta_1, \tau_1) : M_0 \to M_1$ and $f_2 = (\eta_2, \tau_2) : M_0 \to M_2$ be morphisms in $\mathbf{CM}_{\mathrm{abs}}$ (causal abstractions). Then the pushout of f_1 and f_2 in $\mathbf{CM}_{\mathrm{abs}}$ always exists, and is explicitly constructed as follows:*

- ***Nodes** (V_P)**:** Let V_P be the set pushout of the node maps. Since η_1, η_2 are surjective, V_P is the quotient of V_0 by the equivalence relation $\sim$ generated by:*

  $$u \sim v \iff \eta_1(u) = \eta_1(v) \vee \eta_2(u) = \eta_2(v)$$

  Let $q : V_0 \to V_P$ be the canonical quotient map. Let $g_1 : V_{M_1} \to V_P$ and $g_2 : V_{M_2} \to V_P$ be the induced maps such that $g_1 \circ \eta_1 = q = g_2 \circ \eta_2$.
- ***Order** ($\leq_P$)**:** The total order on V_P is forced (see also Step 1 of the proof)*
- ***Initial Edges** (E_{pr})**:** Define a preliminary set of edges as follows: a directed edge $u \to v$ with $u \neq v$ exists in P if either there exists an edge $u_1 \to v_1$ in M_1 such that $g_1(u_1) = u$, $g_1(v_1) = v$, or there exists an edge $u_2 \to v_2$ in M_2 such that $g_2(u_2) = u$, $g_2(v_2) = v$ (edges whose endpoints are identified by the quotient, i.e., $u = v$, are discarded)*
- ***Domains** (dom_P)**:** For a node $u \in V_P$, let $\mathrm{dom}_P(u)$ be the set pushout of the joint fiber maps over the whole fiber $q^{-1}(u)$:*

  $$\mathrm{dom}_P(u) := \Big(\mathrm{dom}_{M_1}\big(g_1^{-1}(u)\big) \sqcup \mathrm{dom}_{M_2}\big(g_2^{-1}(u)\big)\Big) / \sim$$

where $\sim$ is generated by $\tau_1(b) \sim \tau_2(b)$ for $b \in \mathrm{dom}_{M_0}(q^{-1}(u))$, with τ_i denoting the joint fiber maps of f_i over the nodes of $g_i^{-1}(u)$ (recall that the fiber maps of Definition 4 act jointly on fibers). Let $\lambda_1 : \mathrm{dom}(V_{M_1}) \to \mathrm{dom}(V_P), \lambda_2 : \mathrm{dom}(V_{M_2}) \to \mathrm{dom}(V_P)$ be the induced domain maps, which factor fiberwise over g_1, g_2, and let $\tau_P := \lambda_1 \circ \tau_1 = \lambda_2 \circ \tau_2$. Since τ_1, τ_2 are fiberwise surjective, so are λ_1, λ_2 and τ_P.

- ***Mechanisms** (p^P): The local mechanisms are defined* interventionally*: for any $u \in V_P$, $a \in \mathrm{dom}_P(u)$, and $c \in \mathrm{dom}_P(\mathrm{pa}_{pr}(u))$,*

$$p^P\left(u = a \mid \mathrm{pa}_{pr}(u) = c\right) := p^{M_0}_{\mathrm{do}\left(q^{-1}(\mathrm{pa}_{pr}(u))=b\right)}\left(q^{-1}(u) \in \tau_{P,u}^{-1}(a)\right) \tag{5.2}$$

for an arbitrary $b \in \mathrm{dom}_{M_0}(q^{-1}(\mathrm{pa}_{pr}(u)))$ with $\tau_P(b) = c$ (such b exists by surjectivity of τ_P, and the proof shows the right-hand side does not depend on the choice).

- ***Final Edges** (E_P): The final edge set $E_P \subseteq E_{Pr}$ is defined by removing any edges from E_{pr} that are not strictly necessary for the factorization of p^P. That is, if $p^P(u \mid \mathrm{pa}_{pr}(u))$ is conditionally independent of a specific parent given the others, the edge from that parent is removed to satisfy the minimality requirement of CBNs (Definition 3).*

Then $(P, \leq_P)$, equipped with morphisms $h_1 = (g_1, \lambda_1)$ and $h_2 = (g_2, \lambda_2)$, is the pushout of f_1 and f_2.

Proof. See Appendix Section 7.6 □

Possible relaxations of the conditions of Theorems 1 to 3 are discussed in Remark 7.7 (appendix).

Example 9 (Combining Structural Abstractions via Pushout). We illustrate Theorem 3 with a scenario where we want to combine two different high-level abstractions of a complex underlying model into a single, unified abstraction.

Consider a detailed causal model M_0 with underlying DAG

$$M_0: \quad A_1 \longrightarrow A_2 \rightrightarrows \{B_1, B_2\}$$

The model M_1 is a causal abstraction of M_0 that lumps the first two variables A_1, A_2 into a single abstract node A using a surjective map $\tau_{1,A} : \mathrm{dom}_{M_0}(\{A_1, A_2\}) \to \mathrm{dom}_{M_1}(A)$, resulting in the following DAG:

$$M_1: \quad A \rightrightarrows \{B_1, B_2\}$$

M_2 is another causal abstraction of M_0 that groups B_1 and B_2 into a single abstract node B using a surjective map $\tau_{2,B} : \mathrm{dom}_{M_0}(\{B_1, B_2\}) \to \mathrm{dom}_{M_2}(B)$, resulting in the DAG:

$$M_2: \quad A_1 \longrightarrow A_2 \longrightarrow B$$

The Pushout (P): The pushout merges these two high-level models into the unique, largest causal network that serves as an abstraction for both M_1 and M_2 in CM_{abs}. Its underlying DAG is

$$P: \quad A \longrightarrow B$$

with the local mechanism $A \to B$:

$$\forall a \in \mathrm{dom}(A), b \in \mathrm{dom}(B) : p^P(b|a) = p^{M_1}(\lambda_{1,B}^{-1}(b)|a)$$

where λ_1 is the induced domain map of Theorem 3, whose component $\lambda_{1,B} : \mathrm{dom}_{M_1}(\{B_1, B_2\}) \to \mathrm{dom}_P(B)$ groups the values of B_1, B_2.

A further example for Theorem 3, in which the underlying graph remains fixed and the two abstractions instead group domain values of different variables, is given in Example 15 (appendix). Remark 7.8 (appendix) quantifies sizes and computational complexity.

6 CONCLUSION AND FUTURE WORK

We introduced a category-theoretic framework for modular causal inference: by defining the categories of causal submodels (CM_{sub}) and causal abstractions (CM_{abs}) and constructing pushouts and pullbacks in them, we established sufficient conditions under which CBNs sharing a common causal submodel, a common abstraction, or a common underlying model can be canonically merged.

Since our theorems gave sufficient conditions for canonical merging, a natural next step is to find necessary ones. A second limitation is our standing assumption of causal sufficiency. Relaxing it, e.g. via probabilistic dependency graphs [Richardson and Halpern, 2020] as merge targets, is an interesting direction. The categorical perspective suggests further directions: which other shapes of (co)limits exist in $\mathbf{CM}_{\mathrm{sub}}$ and $\mathbf{CM}_{\mathrm{abs}}$; whether the two kinds of morphisms can be reconciled in one framework, where devices such as double categories do not apply directly since general causal maps fail to compose (Example 3); and redeveloping our constructions for PDGs (Remark 7.5), which promises more permissive merging at the price of leaving the class of CBNs. Finally, turning our merging operations into practical algorithms is left to future work, e.g. for mechanistic interpretability, where neural networks are examined through the lens of causal models [Geiger et al., 2025].

Acknowledgements

The TU Eindhoven authors received support from their Department of Mathematics and Computer Science and the Eindhoven Artificial Intelligence Systems Institute.

References

Sander Beckers and Joseph Y Halpern. Abstracting causal models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2678–2685, 2019.

Sander Beckers, Frederick Eberhardt, and Joseph Y. Halpern. Approximate causal abstractions. In *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence*, volume 115 of *Proceedings of Machine Learning Research*, pages 606–615. PMLR, 2020.

Gabriele D'Acunto and Claudio Battiloro. The relativity of causal knowledge. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pages 863–881. PMLR, 2025. URL https://proceedings.mlr.press/v286/d-acunto25a.html.

Marek J Druzdzel and Herbert A Simon. Causality in bayesian belief networks. In *Uncertainty in Artificial Intelligence*, pages 3–11. Elsevier, 1993.

Markus Englberger and Devendra Singh Dhami. Causal abstractions, categorically unified, 2025. URL https://arxiv.org/abs/2510.05033.

Brendan Fong. Causal theories: A categorical perspective on Bayesian networks. Master's thesis, University of Oxford, 2012. arXiv:1301.6201.

Atticus Geiger, Duligur Ibeling, Amir Zur, Maheep Chaudhary, Sonakshi Chauhan, Jing Huang, Aryaman Arora, Zhengxuan Wu, Noah Goodman, Christopher Potts, and Thomas Icard. Causal abstraction: A theoretical foundation for mechanistic interpretability. *Journal of Machine Learning Research*, 26(83):1–64, 2025. URL http://jmlr.org/papers/v26/23-0058.html.

Luigi Gresele, Julius von Kügelgen, Vincent Stimper, Bernhard Schölkopf, and Michel Besserve. Independent mechanism analysis, a new concept? In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 28233–28248. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/edc27f139c3b4e4bb29d1cdbc45663f9-Paper.pdf.

Bart Jacobs, Aleks Kissinger, and Fabio Zanasi. Causal inference by string diagram surgery. In *Foundations of Software Science and Computation Structures (FoSSaCS)*, volume 11425 of *Lecture Notes in Computer Science*, pages 313–329. Springer, 2019.

Riccardo Massidda, Atticus Geiger, Thomas Icard, and Davide Bacciu. Causal abstraction with soft interventions. In *Proceedings of the Second Conference on Causal Learning and Reasoning*, volume 213 of *Proceedings of Machine Learning Research*, pages 68–87. PMLR, 2023. URL https://proceedings.mlr.press/v213/massidda23a.html.

Jun Otsuka and Hayato Saigo. On the equivalence of causal models: A category-theoretic approach. In Bernhard Schölkopf, Caroline Uhler, and Kun Zhang, editors, *Proceedings of the First Conference on Causal Learning and Reasoning*, volume 177 of *Proceedings of Machine Learning Research*, pages 634–646. PMLR, 11–13 Apr 2022. URL https://proceedings.mlr.press/v177/otsuka22a.html.

Judea Pearl. From bayesian networks to causal networks. In *Mathematical models for handling partial knowledge in artificial intelligence*, pages 157–182. Springer, 1995.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press, 2017.

Theodora-Mara Pîslar, Sara Magliacane, and Atticus Geiger. Combining causal models for more accurate abstractions of neural networks. In *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, volume 275 of *Proceedings of Machine Learning Research*, pages 114–138. PMLR, 2025. URL https://proceedings.mlr.press/v275/pislar25a.html.

Oliver Richardson and Joseph Y Halpern. Probabilistic dependency graphs, 2020. URL https://arxiv.org/abs/2012.10800.

Eigil F. Rischel and Sebastian Weichwald. Compositional abstraction error and a category of causal models. In Cassio de Campos and Marloes H. Maathuis, editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1013–1023. PMLR, 27–30 Jul 2021. URL https://proceedings.mlr.press/v161/rischel21a.html.

Eigil Fjeldgren Rischel. The category theory of causal models. Master's thesis, University of Copenhagen, 2020.

Paul K. Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M. Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In *Proceedings of the*

33rd Conference on Uncertainty in Artificial Intelligence (UAI), 2017.

Willem Schooltink and Fabio Massimo Zennaro. Aligning graphical and functional causal abstractions. In Biwei Huang and Mathias Drton, editors, *Proceedings of the Fourth Conference on Causal Learning and Reasoning*, volume 275 of *Proceedings of Machine Learning Research*, pages 704–730. PMLR, 07–09 May 2025. URL `https://proceedings.mlr.press/v275/schooltink25a.html`.

Willem Schooltink and Fabio Massimo Zennaro. Multi-level causal embeddings, 2026. URL `https://arxiv.org/abs/2602.22287`.

Robert E. Tillman, David Danks, and Clark Glymour. Integrating locally learned causal structures with overlapping variables. In *Advances in Neural Information Processing Systems*, volume 21, 2008.

Sofia Triantafillou and Ioannis Tsamardinos. Constraint-based causal discovery from multiple interventions over overlapping variable sets. *Journal of Machine Learning Research*, 16:2147–2205, 2015.

Fabio Massimo Zennaro. Abstraction between structural causal models: A review of definitions and properties, 2022. URL `https://arxiv.org/abs/2207.08603`.

Yuchen Zhu, Sergio Hernan Garrido Mejia, Bernhard Schölkopf, and Michel Besserve. Unsupervised causal abstraction. In *Causal Representation Learning Workshop at NeurIPS*, 2024. URL `https://openreview.net/forum?id=BP0e8RvFwd`.

7 APPENDIX

7.1 BASIC CATEGORY-THEORETIC NOTIONS

For completeness, we recall here the general definition of a category and the universal constructions of products and coproducts, complementing the pullbacks and pushouts of Section 3.1.

Definition 7. *A **category** $\mathcal{C}$ consists of:*

- *a collection of* objects*,*
- *for every pair of objects A, B, a collection $\mathrm{Hom}_{\mathcal{C}}(A, B)$ of* morphisms *$f : A \to B$,*
- *for every object A, an* identity morphism *$\mathrm{id}_A : A \to A$,*
- *a rule for* composition *of morphisms, assigning to each pair of morphisms $f : A \to B$ and $g : B \to C$ a composite morphism $g \circ f : A \to C$.*

These data (the objects, morphisms, identities, and composition rule) are required to satisfy:

1. ***Associativity:*** *$(h \circ g) \circ f = h \circ (g \circ f)$,*
2. ***Identity laws:*** *$f \circ \mathrm{id}_A = f$ and $\mathrm{id}_B \circ f = f$.*

Example 10. The category **Set** has sets as objects, functions as morphisms, identity functions as identity morphisms, and ordinary function composition.

Remark 7.1 (Commuting diagrams). Throughout, we phrase universal properties using *commuting diagrams*: a diagram of objects and morphisms *commutes* if any two directed paths with the same start and end object compose to the same morphism. For instance, a triangle of morphisms $f : X \to A$, $g : A \to B$, $h : X \to B$ commutes precisely when $g \circ f = h$; commutativity is thus a compact way of stating systems of equations between composite morphisms.

Definition 8 (Products)**.** *Given objects A and B in a category $\mathcal{C}$, their* product *consists of an object $A \times B$ together with morphisms $\pi_A : A \times B \to A$ and $\pi_B : A \times B \to B$ such that, for any object X and morphisms $f : X \to A$, $g : X \to B$, there exists a unique morphism*

$$\langle f, g\rangle : X \to A \times B$$

making the diagram commute, i.e., $\pi_A \circ \langle f, g\rangle = f$ and $\pi_B \circ \langle f, g\rangle = g$.

Example 11 (Set). In **Set**, the product is the Cartesian product

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\},$$

with the usual projection maps.

Definition 9. *Dually, the **coproduct** of objects A and B consists of an object $A \sqcup B$ and morphisms $\iota_A : A \to A \sqcup B$ and $\iota_B : B \to A \sqcup B$ such that for any object X and morphisms $f : A \to X$, $g : B \to X$, there exists a unique morphism*

$$[f, g] : A \sqcup B \to X$$

making the diagram commute.

Example 12. In **Set**, the coproduct is the disjoint union $A \sqcup B$, which allows functions out of $A \sqcup B$ to be defined by cases.

7.2 A SECOND RUNNING EXAMPLE: PREORDERS

To complement the category $\mathbf{Set}$ used in the main text, we illustrate the concepts of Sections 3.1 and 7.1 in a second, structurally very different class of categories: preorders. A *preorder* $(Q, \leq)$ is a set Q equipped with a reflexive and transitive relation $\leq$. Every preorder can be viewed as a category: the objects are the elements of Q, and for $a, b \in Q$ the set $\mathrm{Hom}(a, b)$ contains exactly one morphism if $a \leq b$ and is empty otherwise. Reflexivity provides the identity morphisms, and transitivity provides composition; associativity and the identity laws hold trivially because between any two objects there is at most one morphism.

In a preorder, the universal constructions of Sections 3.1 and 7.1 take the following form. The product of a and b is their *meet* (greatest lower bound) $a \wedge b$, when it exists: it satisfies $a \wedge b \leq a$ and $a \wedge b \leq b$ (the projections), and any x with $x \leq a$ and $x \leq b$ satisfies $x \leq a \wedge b$ (the universal property). Dually, the coproduct is the *join* (least upper bound) $a \vee b$. Since all diagrams commute automatically, the pullback of $a \to c \leftarrow b$ is again simply the meet $a \wedge b$, and the pushout of $a \leftarrow c \to b$ is the join $a \vee b$; the "shared" object c plays no role beyond guaranteeing comparability. For a concrete example, consider the natural numbers ordered by divisibility, $(\mathbb{N}, |)$: products/pullbacks are greatest common divisors and coproducts/pushouts are least common multiples.

Two lessons from this example are relevant for the present paper. First, whether (co)limits exist depends on the ambient category: in the preorder $(\mathbb{R}, \leq)$, for instance, the pushout of $a \leftarrow c \to b$ always exists (it is $\max(a, b)$), whereas in a preorder without upper bounds pushouts may fail to exist. This mirrors the situation in $\mathbf{CM}_{\mathrm{sub}}$ and $\mathbf{CM}_{\mathrm{abs}}$, where our theorems require explicit sufficient conditions for the existence of pushouts and pullbacks. Second, the same diagram shape can have very different interpretations in different categories (gluing of sets in $\mathbf{Set}$, suprema in a preorder, and merging of causal models in our categories) which is precisely the abstraction benefit of the categorical formulation.

7.3 THE FOURTH CONSTRUCTION: PULLBACKS IN $\mathbf{CM}_{\mathrm{sub}}$

The main text studies pushouts in $\mathbf{CM}_{\mathrm{sub}}$ (Theorem 1) as well as pullbacks and pushouts in $\mathbf{CM}_{\mathrm{abs}}$ (Theorem 2 and Theorem 3). For completeness, we briefly sketch the remaining combination, pullbacks in $\mathbf{CM}_{\mathrm{sub}}$, which we omitted from the main text as it is the simplest and arguably the least interesting of the four constructions.

Consider two CBNs M_1, M_2 that are both causal submodels of a common CBN M_3, witnessed by morphisms $f_1 = (\eta_1, \tau_1) : M_1 \to M_3$ and $f_2 = (\eta_2, \tau_2) : M_2 \to M_3$ in $\mathbf{CM}_{\mathrm{sub}}$. Loosely speaking, we hypothesize that the pullback is simply the *intersection* of the two submodels inside M_3: since η_1 and η_2 are injective, the pullback node set $\{(v_1, v_2) \in V_{M_1} \times V_{M_2} \mid \eta_1(v_1) = \eta_2(v_2)\}$ can be identified with the intersection $\eta_1(V_{M_1}) \cap \eta_2(V_{M_2})$, and likewise the fiberwise pullbacks of the (injective) domain maps identify with the intersections of the embedded domains, with edges and local mechanisms restricted accordingly. For example, let M_1 be a two-node CBN with graph $A \to B$ and M_2 a two-node CBN with graph $A \to C$, both included in the obvious way into a three-node CBN M_3 with graph $C \leftarrow A \to B$. The pullback is then the CBN consisting of the single node A, equipped with the marginal distribution of A. As for the constructions in the main text, suitable conditions are required for this intersection to constitute a valid object of $\mathbf{CM}_{\mathrm{sub}}$, for instance, the restricted mechanisms must remain well-defined and the inclusion of the intersection must respect causal sufficiency, and we leave a full treatment, including precise existence conditions, to future work.

7.4 ADDITIONAL REMARKS

Remark 7.2 (Total order assumption). The total orders enter Theorem 2 and Theorem 3 in exactly two ways: they restrict which spans and cospans exist in CMabs (monotonicity excludes contradictory causal directions, which is what makes acyclicity of the constructed graphs automatic), and they induce the order $\leq P$. All other data (node sets, edges, domains, mechanisms, edge pruning, structure morphisms, and the mediating morphisms of the universal properties) are defined purely from the underlying unordered CBNs and causal maps; the orders appear in the proofs only to verify monotonicity, interval fibers, and acyclicity. Hence any two admissible choices of orders yield the same underlying merged CBN and morphisms, differing only in $\leq P$. The order thus injects no information into the merged model, in keeping with canonicity: it merely certifies that the causal directions of the inputs are globally compatible

Remark 7.3 (Edge preservation versus edge reflection). For causal submodel morphisms we demand that edges are preserved in the image, whereas, in the case of causal abstractions, we demand that edges in the image are preserved in the preimage. This corresponds to the intuition that a causal abstraction coarsens the domains and may thus abstract away a causal effect. By contrast, a causal submodel morphism refines the domains and hence may add an edge.

Remark 7.4 (Causal reading of condition 3). Each bullet of condition 3 of Theorem 1 rules out a specific source of *causal* ambiguity for nodes v_1, v_2 outside the shared submodel. The first ensures they are connected through the overlap so that their causal relationship is determined (Example 6), with M_0 empty and M_1, M_2 single-variable models, no canonical merge exists. The second reflects that parent sets of a shared node merge unambiguously only when one contains the other (Example 5). Note that no separate condition is needed to handle *accidental duplicates*, exclusive variables of M_1 and M_2 agreeing in domain and distribution but not identified within M_0: the connectivity required by the first bullet already forces any competing cocone to keep such nodes distinct (see the injectivity argument in the proof), so the pushout, which never identifies them, remains universal. How far these sufficient conditions are from necessary is open (Section 6); Remark 7.5 discusses relaxations via a more expressive target class.

Remark 7.5 (On the restrictiveness of canonical merging). Canonical merging via pushouts is restrictive. Example 5 shows that two models pointing into a shared node cannot be glued, as the collider mechanism $p^P(C \mid A, B)$ is underdefined. Dually, Example 6 shows that two models pointing *out of* a shared node cannot be glued either: the merged CBN over $B \leftarrow A \to C$ would implicitly declare B and C conditionally independent given A, information contained in neither model. In both cases the universality requirement prevents such arbitrary choices; this makes the construction restrictive, but ensures that, whenever the pushout exists, the merged model is unambiguous and arguably the only "right" one. A way to circumvent the restrictiveness may be to merge into a more expressive model class: *probabilistic dependency graphs* (PDGs) [Richardson and Halpern, 2020] attach possibly conflicting conditional distributions to *edges* rather than nodes, so the two mechanisms of Example 5 can simply coexist, and we conjecture that pushout merging is always possible in a suitable category of PDGs. This would allow dropping the second bullet of condition 3 in Theorem 1 (and potentially condition 2), and may allow dropping the first condition of Theorem 2, at the price that the result is no longer a CBN; we leave this to future work.

Remark 7.6 (On the maximality of the two categories). A natural question is whether $\mathbf{CM}_{\mathrm{sub}}$ and $\mathbf{CM}_{\mathrm{abs}}$ are *maximal* subclasses of causal maps that are closed under composition. Example 3 shows that the injectivity respectively surjectivity assumptions on the *node* maps cannot simply be dropped: mixing an injective with a surjective node map already destroys the fiberwise factorization. On the level of the *domain* maps, however, the injectivity respectively surjectivity requirements could be relaxed without losing closure under composition, so in this strict sense the two categories are not maximal. We nevertheless impose the conditions on the domain maps as well, since it is precisely this combination that carries the intuitive meaning of a submodel (an embedding that loses no information) and of an abstraction (a coarsening that forgets information), and that connects to the existing notions of abstraction discussed above. We also note that the two classes of morphisms cannot straightforwardly be reconciled within a single ambient categorical framework (such as a double category or a model structure with two distinguished classes of maps), since general causal maps fail to compose; we leave this question to future work.

Remark 7.7 (On relaxing the sufficient conditions). The conditions in Theorems 1 to 3 are sufficient but not claimed to be necessary, and it is natural to ask how restrictive they are in practice. Several of them can in principle be relaxed, but generally at the cost of substantially more intricate hypotheses on the underlying domains and distributions, and of a corresponding loss of readability. Theorem 3 is hypothesis-free: acyclicity of the constructed graph is automatic in the ordered category, and disagreements between the two abstractions are absorbed by the quotient defining the domain pushout, at the price that the merged domains may become very coarse. The situation is different for Theorem 2: its singleton-fiber/bijectivity conditions (which guarantee that one side determines the merged domain) cannot simply be dropped, since a pullback must *refine* both models, and reconciling two genuinely different refinements of the same variable reintroduces exactly the kind of ambiguity that the counterexamples of Examples 5 and 14 show has no canonical resolution. As discussed in Remark 7.5, a principled way to lift several of these restrictions at once is to enlarge the target class of models (e.g. to probabilistic dependency graphs), which can absorb the otherwise-underdetermined mechanisms. We have chosen to present clean sufficient conditions that keep the constructions transparent, and regard the precise characterization of necessary conditions as the most important open problem left by this work.

Remark 7.8 (Size and computational cost of the constructions). While this paper focuses on existence and canonicity, the constructions of Theorems 1 to 3 are explicit, and the resulting models are small. The pushout of Theorem 1 has exactly $|V_{M_1}| + |V_{M_2}| - |V_{M_0}|$ nodes and at most $|E_{M_1}| + |E_{M_2}| - |E_{M_0}|$ edges (every edge of M_0 appears in both images and is identified), and its domains and local mechanisms are copies of those of M_1 and M_2, so it can be computed in time linear in the size of the inputs, including their conditional probability tables. For the pullback of Theorem 2, the singleton-fiber condition implies that each fiber product collapses to the larger fiber, giving exactly $|V_P| = \sum_{c \in V_{M_3}} \max\bigl(|\eta_1^{-1}(c)|, |\eta_2^{-1}(c)|\bigr) = |V_{M_1}| + |V_{M_2}| - |V_{M_3}|$ nodes; every preliminary edge is the lift of an edge of M_1 or M_2 along fibers, with at most F lifts per source edge, so $|E_P| \leq |E_{pr}| \leq F \cdot (|E_{M_1}| + |E_{M_2}|)$, where $F := \max_{i \in \{1,2\}} \max_{c \in V_{M_3}} |\eta_i^{-1}(c)|$ denotes the maximal fiber size; the mechanisms are again copies (from the respective detailed side). For the pushout of Theorem 3, V_P is a quotient of both V_{M_1} and V_{M_2}, so $|V_P| \leq \min(|V_{M_1}|, |V_{M_2}|)$, and every edge is the image of an edge of M_1 or M_2, so $|E_P| \leq |E_{M_1}| + |E_{M_2}|$; here the local mechanisms are pushforwards of those of M_0, whose computation requires summations over the fibers of τ_P and hence scales with the domain sizes of M_0. In Theorems 2 and 3, the final minimality-driven edge removal amounts to conditional-independence checks on the constructed mechanisms. Turning these observations into practical merging algorithms, including for models specified implicitly, is part of the future work outlined in Section 6.

7.5 FURTHER EXAMPLES

Example 13 (Counter-Example: Failure without Positivity). Condition 2 of Theorem 1 cannot be dropped. Let the shared submodel M_0 consist of a single treatment variable T with $\mathrm{dom}_{M_0}(T) = \{\text{none}, \text{low}\}$. Let M_1 consist of T alone but with the enlarged domain $\mathrm{dom}_{M_1}(T) = \{\text{none}, \text{low}, \text{high}\}$, where the dose *high* is foreseen by the study protocol but was never administered, i.e., has probability zero; the inclusion f_1 is then a morphism in $\mathbf{CM}_{\mathrm{sub}}$, since Definition 4 requires commutation only for interventions at values in the image of the fiber map. Let M_2 add an effect: $V_{M_2} = \{T, R\}$ with $T \to R$, recording recovery under the original two doses. Conditions 1 and 3 of Theorem 1 hold (M_1 has no exclusive nodes); only condition 2 fails, at the shared node. A merged model would have to carry a mechanism $p(R \mid T{=}\text{high})$, the response to a dose never given, which is constrained by neither model: every choice yields a valid cocone, no morphisms between different choices exist (they must be the identity on nodes and values, yet commute with $\mathrm{do}(T{=}\text{high})$), and hence no cocone is universal. This is precisely where positivity enters the proof: it forces the fiber maps at shared nodes to be bijections (Step 1), so that no interventional behavior at new domain values must be invented. Since the proof uses condition 2 only at shared nodes, the condition could be weakened to positivity of the variables in $\eta_i(V_{M_0})$.

Example 14 (Counter-Example: Failure of Domain Map Bijectivity). We illustrate a case where we have two simple treatment-effect models M_1, M_2 that both abstract to the same higher-level model M_3 by abstracting the effect-domain, but where we cannot canonically merge the low-level causal models.

Abstract Model M_3 is a simple binary cause-effect model.

- $V_{M_3} = \{C, E\}$, $E_{M_3} = \{C \to E\}$.
- **Domains:** $\mathrm{dom}_{M_3}(C) = \{0, 1\}$, $\mathrm{dom}_{M_3}(E) = \{0, 1\}$.

Model M_1 measures the effect E on a 3-point scale by refining effect value $E = 1$ into finer values $1_a, 1_b$.

- $V_{M_1} = \{C, E\}$, $E_{M_1} = \{C \to E\}$.
- $\mathrm{dom}_{M_1}(C) = \{0, 1\}$, $\mathrm{dom}_{M_1}(E) = \{0, 1_a, 1_b\}$.

Model M_2 also measures the effect E on a 3-point scale by having a different refinement of the effect value $E = 1$ into finer values $1_c, 1_d$.

- $V_{M_2} = \{C, E\}$, $E_{M_2} = \{C \to E\}$.
- $\mathrm{dom}_{M_2}(C) = \{0, 1\}$, $\mathrm{dom}_{M_2}(E) = \{0, 1_c, 1_d\}$.

To define the mechanism $p^P(E_{\text{merged}} \mid C)$ in the pullback, Theorem 2 requires that for the abstract node E, at least one model must provide a bijective domain map on E. Here, both maps fail bijectivity. E_{merged} would have domain

$$((0,0), (1_a, 1_c), (1_a, 1_d), (1_b, 1_c), (1_b, 1_d))$$

We cannot canonically define the conditional distribution $p^P(E_{merged} \mid C)$. For example, while we know that $p^P(\{(1_a, 1_c)\} \cup \{(1_a, 1_d)\} | c \in \mathrm{dom}(C))$ needs to equal $p^{M_1}(1_a | c \in \mathrm{dom}(C))$ and similarly $p^P(\{(1_a, 1_c)\} \cup \{(1_b, 1_c)\} | c \in \mathrm{dom}(C))$ needs to equal $p^{M_2}(1_c | c \in \mathrm{dom}(C))$, there are many distributions on $p(E_{merged} | C)$ marginalizing in such a way.

Example 15. We illustrate Theorem 3 with a scenario where the underlying causal graph remains constant across all models, but the models differ in how they group the domain values of a treatment variable.

The Base Model (M_0): Consider a cause-effect model tracking a drug dosage (T) and its effect on recovery (R): $V_0 = \{T, R\}$ and $E_0 = \{T \to R\}$. The domain of the treatment variable is $\mathrm{dom}_{M_0}(T) = \{\text{no treatment, low, high}\}$ representing 3 different dosage levels and the domain of the recovery is $\mathrm{dom}_{M_0}(R) = \{\text{death, unchanged, recovery}\}$.

Model 1 (M_1 - First Abstraction): A researcher notices that the drug dosages *low* and *high* have the same effect on recovery, i.e., it does not matter whether the patient takes a low or high dosage. Hence, they instead consider the causal model where dosage levels {low, high} are abstracted into a single value "treatment" leading to an abstracted model M_1.

A second researcher is only interested in whether a patient survives and accordingly abstracts the three recovery values by combining the values {unchanged, recovery} to the single value {survival} leading to a second abstracted model M_2. Explicitly, the domain map τ_2 for the recovery variable is given by

$$\begin{aligned} \tau_{2,R}(\text{death}) &= \text{death} \\ \tau_{2,R}(\text{unchanged}) &= \text{survival} \\ \tau_{2,R}(\text{recovery}) &= \text{survival} \end{aligned} \tag{7.1}$$

Combining these two abstractions by constructing the pushout as in Theorem 3, we obtain the following high-level Treatment-Effect model: The pushout P is given by the merged model with two variables T, R and domains

$$\mathrm{dom}_P(T) = \{\text{no treat., treat.}\}, \mathrm{dom}_P(R) = \{\text{death, survival}\}$$

$p^P(R|T)$ has the following distribution:

$$\begin{aligned} &p^P(\text{death}|\text{no treat.}) = p^{M_1}(\text{death}|\text{no treat.}) \\ &p^P(\text{surv.}|\text{no treat.}) = p^{M_1}(\{\text{unch., rec.}\}|\text{no treat.}) \\ &p^P(\text{death}|\text{treat.}) = p^{M_1}(\text{death}|\text{treat.}) \\ &p^P(\text{surv.}|\text{treat.}) = p^{M_1}(\{\text{unch., rec.}\}|\text{treat.}) \end{aligned}$$

7.6 PROOFS

Proof of Proposition 4.1.

Proof. **(A) The submodel case $\mathbf{CM}_{\text{sub}}$.** Let $(\eta, \tau) : M_1 \to M_2$ and $(\eta', \tau') : M_2 \to M_3$ be causal submodels. Then $\eta : V_{M_1} \hookrightarrow V_{M_2}$ and $\eta' : V_{M_2} \hookrightarrow V_{M_3}$ are injective, hence

$$\eta'' := \eta' \circ \eta : V_{M_1} \hookrightarrow V_{M_3}$$

is injective. Define $\tau'' := \tau' \circ \tau : \mathrm{dom}(V_{M_1}) \to \mathrm{dom}(V_{M_3})$. We must show that τ'' factors fiberwise over η''.

let $v_3 \in V_{M_3}$ such that $(\eta' \circ \eta)^{-1}(v_3) \neq \emptyset$. By injectivity of η and η', $(\eta' \circ \eta)^{-1}(v_3)$ and $\eta'^{-1}(v_3)$ are single nodes in V_{M_1}, V_{M_2}, respectively. Hence, $\tau_{\eta'^{-1}(v_3)}, \tau'_{v_3}$ have compatible domains to form $\tau''_{v_3} := \tau'_{v_3} \circ \tau_{(\eta')^{-1}(v_3)}$

Intervention commutation for (η'', τ'') follows from the functoriality of pushforward: for $\mathbf{v} \subseteq \eta''(V_{M_1})$ set $\mathbf{w} := \eta'^{-1}(\mathbf{v}) \subseteq \eta(V_{M_1})$; the commutation identity of (η, τ) pushes $p^{M_1}_{\mathrm{do}(\eta^{-1}(\mathbf{w})=a)}$ to the marginal of $p^{M_2}_{\mathrm{do}(\mathbf{w}=\tau(a))}$ on $\eta(V_{M_1})$, and the identity of (η', τ') pushes the latter to the marginal of $p^{M_3}_{\mathrm{do}(\mathbf{v}=\tau''(a))}$ on $\eta''(V_{M_1})$, using that marginalization to a subset of coordinates commutes with the coordinate-wise pushforward τ'_*.

η'' preserves edges since η and η' do. It also respects causal sufficiency: if $x \in V_{M_3}$ is a confounder of $\eta''(u), \eta''(v)$, then, since η' respects causal sufficiency (applied to $\eta(u), \eta(v) \in V_{M_2}$), there exists $x' \in V_{M_2}$ with $\eta'(x') = x$ and $x' \rightsquigarrow \eta(u)$, $x' \rightsquigarrow \eta(v)$; applying the condition for η to x' yields $x'' \in V_{M_1}$ with $\eta(x'') = x'$, hence $\eta''(x'') = x$, and $x'' \rightsquigarrow u$, $x'' \rightsquigarrow v$.

Finally, identities are given by $(\mathrm{id}_V, \mathrm{id}_{\mathrm{dom}(V)})$ (with $\mathrm{dom}(\mathrm{id}(V)) = \mathrm{dom}(V)$), so $\mathbf{CM}_{\text{sub}}$ is a category.

(B) The abstraction case $\mathbf{CM}_{\text{abs}}$. Let $(\eta, \tau) : M_1 \to M_2$ and $(\eta', \tau') : M_2 \to M_3$ be causal abstractions. Then η, η' are surjective, so

$$\eta'' := \eta' \circ \eta : V_{M_1} \twoheadrightarrow V_{M_3}$$

is surjective, and monotone as a composite of monotone maps (identities are likewise monotone).

Define $\tau'' := \tau' \circ \tau : \mathrm{dom}_{M_1}(V_{M_1}) \to \mathrm{dom}_{M_3}(V_{M_3})$. We must show that τ'' factors fiberwise over η'': Let $v_3 \in V_{M_3}$. Then there exists a map $\tau'_{v_3} : \mathrm{dom}_{M_2}(\eta'^{-1}(v_3)) \to \mathrm{dom}_{M_3}(v_3)$. Since η and η' are surjective, $\eta((\eta' \circ \eta)^{-1}(v_3)) = \eta'^{-1}(v_3)$ and hence $\prod_{v_2 \in \eta'^{-1}(v_3)} \tau_{v_2}$ has signature $\mathrm{dom}_{M_1}((\eta' \circ \eta)^{-1}(v_3)) \to \mathrm{dom}_{M_2}(\eta'^{-1}(v_3))$ making the composition $\tau''_{v_3} := \tau'_{v_3} \circ (\prod_{v_2 \in \eta'^{-1}(v_3)} \tau_{v_2})$ well-defined.

Intervention commutation again follows from $(\tau' \circ \tau)_* = \tau'_* \circ \tau_*$ and the commutation equalities for τ and τ'.

It remains to show that η'' reflects edges. Let $\eta''(x) \to \eta''(y) \in E_{M_3}$. Edge reflection of η' yields $x_2 \in \eta'^{-1}(\eta'(\eta(x)))$ and $y_2 \in \eta'^{-1}(\eta'(\eta(y)))$ with $x_2 \to y_2 \in E_{M_2}$. By surjectivity of η there are $x_1, y_1 \in V_{M_1}$ with $\eta(x_1) = x_2$, $\eta(y_1) = y_2$; note $\eta''(x_1) = \eta''(x)$ and $\eta''(y_1) = \eta''(y)$. Edge reflection of η applied to $\eta(x_1) \to \eta(y_1)$ then yields $x'_1 \in \eta^{-1}(x_2)$, $y'_1 \in \eta^{-1}(y_2)$ with $x'_1 \to y'_1 \in E_{M_1}$, and $x'_1 \in \eta''^{-1}(\eta''(x))$, $y'_1 \in \eta''^{-1}(\eta''(y))$, as required. Thus $\mathbf{CM}_{\mathrm{abs}}$ is a category. $\square$

Proof of Theorem 1

Proof. We proceed in four steps: first verifying that P is a valid causal model, second verifying that g_1, g_2 are valid morphisms, third verifying commutativity, and fourth verifying the universal property.

Step 1: Well-definedness of P.

We check well-definedness of the local mechanisms: First note that f_1, f_2 being causal maps, condition 3 of Definition 4 implies that τ_1, τ_2 are measure preserving: This follows by letting the interventional set $\mathbf{v}$ of Definition 4 be the empty set. Together with the fact that every node in M_1 and M_2 has a positive probability function, and the domains are discrete, this implies that $\tau_{i,\eta_i(u)}$ are surjections for every node $u \in V_{M_0}$ and for $i \in \{1,2\}$: Assume this is not the case; then there exists $a \in \mathrm{dom}_{M_i}(\eta_i(u))$ that has no preimage under $\tau_{i,\eta_i(u)}$ but that has positive probability, contradicting measure preservation. Since they are injections by assumption, they are bijections on every fiber. Hence, if $u \in V_{M_0}$ (i.e., u is in the overlap), then the pushout on sets $(\mathrm{dom}^{M_1}(\eta_1(u)) \sqcup \mathrm{dom}^{M_2}(\eta_2(u)))/\sim$, identifying $\tau_{1,v}(a) \sim \tau_{2,v}(a)$ for $a \in \mathrm{dom}_{M_0}(u)$, is canonically isomorphic to $\mathrm{dom}_{M_1}(\eta_1(u))$ and to $\mathrm{dom}_{M_2}(\eta_2(u))$. Hence, the maps θ_1, θ_2 are surjective. Let $\eta := j_1 \circ \eta_1 = j_2 \circ \eta_2$.

Say that side $i \in \{1,2\}$ *covers* a node $w \in j_i(V_{M_i})$ if $\mathrm{pa}_P(w) = j_i(\mathrm{pa}_{M_i}(j_i^{-1}(w)))$; at a covered node, the mechanism of P defined via M_i is an exact copy of the mechanism of M_i at $j_i^{-1}(w)$. We now show that every node of P is covered by at least one side. Consider first an overlap node $\eta(u)$, $u \in V_{M_0}$, and assume w.l.o.g. that $\mathrm{pa}_{M_1}(\eta_1(u))$ is not contained in the image of η_1. Then by the second bullet point of condition three in the theorem, $\mathrm{pa}_{M_2}(\eta_2(u))$ is contained in the image of η_2. Since $\tau_{2,\eta_2(u)}$ is a bijection, the local mechanism at $\eta_2(u)$ is fully determined by the local mechanism at u and therefore $\mathrm{pa}_{M_2}(\eta_2(u)) = \eta_2(\mathrm{pa}_{M_0}(u))$. Indeed, as $\mathrm{pa}_{M_2}(\eta_2(u)) \subseteq \eta_2(V_{M_0})$, intervening in M_2 on $\eta_2(V_{M_0} \setminus \{u\})$ pins all parents of $\eta_2(u)$, so by truncation and the intervention commutation of f_2 (using that all overlap fiber maps are bijections) the mechanism of M_2 at $\eta_2(u)$ is the $\tau_{2,\eta_2(u)}$-pushforward of the mechanism of M_0 at u; the latter depends only on the coordinates of $\mathrm{pa}_{M_0}(u)$, so minimality of M_2 forces $\mathrm{pa}_{M_2}(\eta_2(u)) \subseteq \eta_2(\mathrm{pa}_{M_0}(u))$, while the reverse inclusion holds since f_2 preserves edges; hence $\mathrm{pa}_{M_2}(\eta_2(u)) = \eta_2(\mathrm{pa}_{M_0}(u))$. Consequently, side 1 covers $\eta(u)$: since f_1 preserves edges, $j_2(\mathrm{pa}_{M_2}(\eta_2(u))) = j_1(\eta_1(\mathrm{pa}_{M_0}(u))) \subseteq j_1(\mathrm{pa}_{M_1}(\eta_1(u)))$, and $\mathrm{pa}_P(\eta(u))$ is by construction the union $j_1(\mathrm{pa}_{M_1}(\eta_1(u))) \cup j_2(\mathrm{pa}_{M_2}(\eta_2(u)))$. Now assume that $\mathrm{pa}_{M_1}(\eta_1(u))$ is contained in the image of η_1. Then by the same reasoning $\mathrm{pa}_{M_1}(\eta_1(u)) = \eta_1(\mathrm{pa}_{M_0}(u))$, and symmetrically (edge preservation of f_2) side 2 covers $\eta(u)$. Finally, every exclusive node is covered by its own side: if $u_1 \in V_{M_1} \setminus \eta_1(V_{M_0})$, then $\mathrm{pa}_P(j_1(u_1)) = j_1(\mathrm{pa}_{M_1}(u_1))$, since no edge of M_2 maps onto an edge into $j_1(u_1)$; likewise for M_2.

Therefore, every node of P is covered, and the defined mechanisms cover all local mechanisms. Together with the fact that θ_1, θ_2 are surjections, this implies that the local mechanisms on P are fully defined by our construction. Further, the two candidate definitions agree wherever both apply: if an overlap node $\eta(u)$ is covered by both sides, then neither parent set can contain an exclusive node (its image would lie in the other side's parent-set image), so $\mathrm{pa}_{M_i}(\eta_i(u)) = \eta_i(\mathrm{pa}_{M_0}(u))$ for both $i \in \{1,2\}$ by the argument above, and by the interventional argument above the mechanism of M_i at $\eta_i(u)$ is, for both i, the $\tau_{i,\eta_i(u)}$-pushforward of the mechanism of M_0 at u evaluated at the τ_i-transported parent values; under the identification of the overlap domains via θ_1, θ_2 these two transported copies coincide. Thus all local mechanisms are well-defined. Finally, P satisfies the minimality requirement of Definition 3: the mechanism of P at any node is an exact copy of a mechanism of M_1 or M_2 at a node whose parent set is mapped bijectively onto its P-parent set (shown above), and by minimality of M_1 resp. M_2 it genuinely depends on every parent coordinate; intervening on the parents and varying that coordinate shows that no edge of E_P can be removed without changing the interventional family.

Step 2: g_1 (and analogously g_2) is a valid morphism. By construction, j_1 is injective and preserves edges, and it respects causal sufficiency by assumption (condition 1). The domain map θ_1 factors fiberwise and is bijective on every fiber (Step 1). It remains to show that g_1 commutes with interventions as required in Definition 4: for every $\mathbf{v} \subseteq j_1(V_{M_1})$ and $a \in \mathrm{dom}_{M_1}(j_1^{-1}(\mathbf{v}))$,

$$\theta_{1,*}\, p^{M_1}_{\mathrm{do}(j_1^{-1}(\mathbf{v})=a)} = p^{P}_{\mathrm{do}(\mathbf{v}=\theta_1(a))}\big(j_1(V_{M_1})\big). \tag{7.2}$$

Since θ_1 is fiberwise bijective, it suffices to verify this pointwise, and we suppress θ_1 and its inverse where they act as relabelings.

By Step 1, the local mechanism of each $u \in V_P$ is a copy of the mechanism of M_1 or of M_2 at a node whose parent set is mapped onto $\mathrm{pa}_P(u)$; call u of *type* 1 or 2 accordingly, fixing type 1 for overlap nodes at which both sides qualify (their mechanisms coincide by Step 1). Every factor of the truncated factorization of $p^P_{\mathrm{do}(\mathbf{v}=\theta_1(a))}$ belonging to a type-1 node involves only coordinates of $j_1(V_{M_1})$, every factor of a type-2 node only coordinates of $j_2(V_{M_2})$, and every indicator factor of an intervened node only coordinates of $j_1(V_{M_1})$, since $\mathbf{v} \subseteq j_1(V_{M_1})$. Summing over the coordinates of $V_P \setminus j_1(V_{M_1})$, which occur only in type-2 factors, the marginal on $j_1(V_{M_1})$ equals the product of all type-1 factors and all indicators, times the function

$$G := \sum_{x_{V_P \setminus j_1(V_{M_1})}} \prod_{\substack{u \notin \mathbf{v} \\ u \text{ of type } 2}} p^P\big(x_u \mid x_{\mathrm{pa}_P(u)}\big),$$

which depends only on the coordinates of the overlap $j_1(\eta_1(V_{M_0}))$.

We compute G in three moves. *First*, the factors of G are the mechanisms of M_2 at all nodes of V_{M_2} except those nodes of $\eta_2(V_{M_0})$ whose image in P is intervened or of type 1. A product of the mechanisms of all nodes outside a given set equals the truncated factorization under intervention on that set; hence G is the marginal, on the non-intervened overlap nodes of type 2, of an interventional distribution of M_2 whose interventional set is contained in $\eta_2(V_{M_0})$, with interventional values given by the overlap coordinates of x. *Second*, since this set lies in the image of η_2 and the fiber maps between overlap domains are bijections (Step 1, using condition 2), the commutation identity of f_2 expresses this marginal as the pushforward of the corresponding interventional distribution of M_0, and the commutation identity of f_1, applied to the same intervention of M_0, expresses it in turn as

$$p^{M_1}_{\mathrm{do}(\mathbf{w}=t)}\big(x_R\big), \tag{7.3}$$

where $\mathbf{w} \subseteq \eta_1(V_{M_0})$ consists of the η_1-images of the intervened overlap nodes, t of the transported values, and $R \subseteq \eta_1(V_{M_0})$ of the η_1-images of the non-intervened type-2 overlap nodes. *Third*, we claim

$$p^{M_1}_{\mathrm{do}(\mathbf{w}=t)}\big(x_R\big) = \prod_{y \in R} p^{M_1}\big(x_y \mid x_{\mathrm{pa}_{M_1}(y)}\big). \tag{7.4}$$

Indeed, for $y \in R$ we have $\mathrm{pa}_{M_1}(y) \subseteq \eta_1(V_{M_0})$: by Step 1, the η_2-image of y has a parent in $V_{M_2} \setminus \eta_2(V_{M_0})$, so a parent of y in $V_{M_1} \setminus \eta_1(V_{M_0})$ would violate the second bullet of condition 3, and an overlap parent edge of y in M_1 without counterpart in M_2 would make $j_1(y)$ of type 1 by Step 1. Hence every parent of a node of R lies in $R \cup \mathbf{w}$, no factor on the right-hand side of Eq. (7.4) involves a coordinate of $V_{M_1} \setminus \eta_1(V_{M_0})$, and on the left-hand side these coordinates sum to one in reverse topological order, proving Eq. (7.4) pointwise. Note that both sides are interventional expressions, so no conditioning on parent configurations of probability zero is involved.

Substituting back, the marginal of $p^P_{\mathrm{do}(\mathbf{v}=\theta_1(a))}$ on $j_1(V_{M_1})$ is the product of the M_1-mechanisms of all type-1 nodes, the M_1-mechanisms of the nodes of R, and the indicators of $\mathbf{v}$, together exactly the truncated factorization of M_1 under $\mathrm{do}(j_1^{-1}(\mathbf{v}) = a)$. This proves Eq. (7.2).

Step 3: Commutativity. Since by construction, θ_1, θ_2 factor fiberwise over j_1, j_2, and the domain maps for every fiber are pushouts on sets, we have commutativity

$$\theta_1 \circ \tau_1 \;=\; \theta_2 \circ \tau_2 \quad \text{on } \mathrm{dom}(V_0).$$

Since the set of nodes V_P is also defined as a pushout on sets, we have

$$j_1 \circ \eta_1 = j_2 \circ \eta_2.$$

Step 4: Universal Property. Let X be any object in $\mathbf{CM}_{sub}$ with morphisms $h_1 = (\alpha_1, \beta_1) : M_1 \to X$ and $h_2 = (\alpha_2, \beta_2) : M_2 \to X$ such that $h_1 \circ f_1 = h_2 \circ f_2$.

Since V_P is the pushout in **Set**, there exists a unique function $\alpha : V_P \to V_X$ such that $\alpha \circ j_i = \alpha_i$ for $i \in \{1, 2\}$. Similarly, since the domains of P are fiberwise pushouts, there exists a unique domain map $\beta : \mathrm{dom}(V_P) \to \mathrm{dom}(V_X)$ that factorizes fiberwise. We define the candidate morphism $k = (\alpha, \beta) : P \to X$.

α is injective: Let $u, v \in V_P$ and $u \neq v$. Then $\exists i_1, i_2 \in \{1, 2\}$ s.t. $\exists u_{i_1} \in V_{M_{i_1}}, v_{i_2} \in V_{M_{i_2}}$ and $j_{i_1}(u_{i_1}) = u, j_{i_2}(v_{i_2}) = v$. If $i_1 = i_2$, then $\alpha_{i_1}(u_{i_1}) \neq \alpha_{i_1}(v_{i_1})$ as α_{i_1} is injective and since $\alpha \circ j_{i_1} = \alpha_{i_1}$, $\alpha(u) \neq \alpha(v)$. Now assume $i_1 \neq i_2$; then $u_{i_1} \in V_{M_{i_1}} \setminus \eta_{i_1}(V_{M_0})$, $v_{i_2} \in V_{M_{i_2}} \setminus \eta_{i_2}(V_{M_0})$, w.l.o.g. $i_1 = 1, i_2 = 2$. Suppose, for contradiction, that $z := \alpha_1(u_1) = \alpha_2(v_2)$. Apply the first bullet of condition 3 to the exclusive pair (u_1, v_2). In case *(ii)*, say with directed paths $u_1 \rightsquigarrow \eta_1(w)$ and $\eta_2(w) \rightsquigarrow v_2$, the images under the edge-preserving maps α_1, α_2 yield directed paths $z \rightsquigarrow \alpha(\eta(w))$ and $\alpha(\eta(w)) \rightsquigarrow z$ in X; both are nontrivial since u_1, v_2 are exclusive, so X contains a directed cycle, contradicting acyclicity of X. In case *(i)*, let a be the first node of $\eta_1(V_{M_0})$ on the given M_1-path from u_1, and b the first node of $\eta_2(V_{M_0})$ on the given M_2-path from v_2. If $a = \eta_1(w_0)$ and $b = \eta_2(w_0)$ for the same $w_0 \in V_{M_0}$, these two edges violate the second bullet of condition 3. Hence a, b are images of distinct $w_a \neq w_b \in V_{M_0}$, and $\alpha(\eta(w_a)) \neq \alpha(\eta(w_b))$ by injectivity of $\alpha_1 \circ \eta_1$. The images of the two segments are directed paths $z \rightsquigarrow \alpha(\eta(w_a))$ with interior in $\alpha_1(V_{M_1} \setminus \eta_1(V_{M_0}))$ and $z \rightsquigarrow \alpha(\eta(w_b))$ with interior in $\alpha_2(V_{M_2} \setminus \eta_2(V_{M_0}))$; in particular, each avoids the other's endpoint (an interior node equal to an overlap image would force, by injectivity of α_1 resp. α_2, an exclusive node to coincide with an overlap image). Thus z is a confounder of the M_0-image nodes $\alpha(\eta(w_a))$ and $\alpha(\eta(w_b))$ with respect to the composite $h_1 \circ f_1 = h_2 \circ f_2 : M_0 \to X$ contradicting causal sufficiency. Therefore $\alpha(u) \neq \alpha(v)$.

β is injective: Let $u \in V_P$. Then there exists $i \in \{1, 2\}$ and $u_i \in V_{M_i}$ s.t. $j_i(u_i) = u$. The map $\beta_{i,\alpha(u)}$ is injective and $\theta_{i,u}$ is bijective. Since $\beta_i = \beta \circ \theta_i$, $\beta_{\alpha(u)}$ has to be injective. Hence, β is injective.

α preserves edges: By construction, every edge in E_P can be written as $j_i(x) \to j_i(y)$ for nodes $x, y \in V_i$ for either $i = 1$ or $i = 2$. Since α_1, α_2 are edge preserving, there needs to be an edge $\alpha_i(x) \to \alpha_i(y)$ which by construction is the image of the edge in E_P.

We show that k preserves causal sufficiency as required in Definition 5. Let $a, b \in V_P$ and let $x \in V_X$ be a confounder of $\alpha(a), \alpha(b)$, i.e., $x \notin \{\alpha(a), \alpha(b)\}$ with directed paths $x \rightsquigarrow \alpha(a)$ and $x \rightsquigarrow \alpha(b)$; we must produce $x_P \in V_P$ with $\alpha(x_P) = x$ and $x_P \rightsquigarrow a$, $x_P \rightsquigarrow b$ in P with the two paths disjoint. If a and b lie in the image of the same j_i (in particular whenever at least one of them is an overlap node), causal sufficiency of h_i applied to the preimage pair yields $x_i \in V_{M_i}$ with $\alpha_i(x_i) = x$ and directed paths to the preimages in M_i; since j_i preserves edges, $x_P := j_i(x_i)$ works. So assume w.l.o.g. $a \in j_1(V_{M_1} \setminus \eta_1(V_{M_0}))$ and $b \in j_2(V_{M_2} \setminus \eta_2(V_{M_0}))$, and write $a = j_1(a_1)$, $b = j_2(b_2)$. By the first bullet of condition 3, one of two cases applies. In case *(i)* there exist $c', c'' \in V_{M_0}$ with directed paths $a_1 \rightsquigarrow \eta_1(c')$ in M_1 and $b_2 \rightsquigarrow \eta_2(c'')$ in M_2. In this case x would be a confounder between $\alpha(\eta(c'))$ and $\alpha(\eta(c''))$. Since $\alpha \circ \eta$ respects causal sufficiency, there is already a confounder x' between c', c'' in V_{M_0} and we obtain the desired $x_P := \eta(x')$; a_1 (and respectively b_2) have to be on the corresponding paths in M_2 (and respectively M_3) as otherwise α_1 (and respectively α_2) would add a counfounder between a and c' (and respectively b and c''). In case *(ii)* there exists a single $u \in V_{M_0}$ with (w.l.o.g.) directed paths $a_1 \rightsquigarrow \eta_1(u)$ in M_1 and $\eta_2(u) \rightsquigarrow b_2$ in M_2; in particular $\eta(u) \rightsquigarrow b$ in P. x is a confounder of $\alpha_2(\eta_2(u))$ and $\alpha_2(b_2)$, and causal sufficiency of h_2 yields $x_2 \in V_{M_2}$ with $\alpha_2(x_2) = x$, $x_2 \rightsquigarrow \eta_2(u)$, and $x_2 \rightsquigarrow b_2$ in M_2; then $x_P := j_2(x_2)$ satisfies $x_P \rightsquigarrow j_2(\eta_2(u))$ and $x_P \rightsquigarrow j_2(u)$ in P. The relevant path $x \rightsquigarrow \alpha_2(\eta_2(u))$ has to visit $\alpha_1(a_1)$ as otherwise α_1 would introduce a confounder between $\alpha_1(a)$ and $\alpha_1(\eta_1(u))$; this implies also that the relevant path $x_P \rightsquigarrow \eta(u)$ has to visit a, showing the claim.

Interventions. We must show, for every $\mathbf{v} \subseteq V_P$ and $a \in \mathrm{dom}_P(\mathbf{v})$,

$$\beta_* \, p^P_{\mathrm{do}(\mathbf{v}=a)} = p^X_{\mathrm{do}(\alpha(\mathbf{v})=\beta(a))}\big(\alpha(V_P)\big). \qquad (7.5)$$

Write $A := \alpha(V_P)$, $A_O := \alpha(\eta(V_{M_0}))$, $A_i := \alpha_i(V_{M_i})$; then $A = A_1 \cup A_2$, $A_1 \cap A_2 = A_O$ by injectivity of α. All graphical statements below are stable under interventions, which only delete edges. We use two devices. *Entry points*: if exclusive nodes x_1, x_2 of the two models have M_i-paths into their overlaps whose first overlap nodes are $\eta_1(w), \eta_2(w)$ for the same $w \in V_{M_0}$, the last edges of the truncated paths violate the second bullet of condition 3. *Preimages*: if $z \in V_X$ has directed paths, each avoiding the other's endpoint, to two distinct nodes of A_i, causal sufficiency of h_i forces $z \in A_i$; to two distinct nodes of A_O, causal sufficiency of $h_1 \circ f_1 = h_2 \circ f_2 \in \mathbf{CM}_{sub}$ (Proposition 4.1) forces $z \in A_O$.

(U) *No node of V_X outside A_O has, for both $i = 1, 2$, a directed path with interior avoiding A into a node of $A_i \setminus A_O$* (a node of $A_i \setminus A_O$ counts as having a trivial such path to itself). Suppose $s \notin A_O$ has such paths π_1, π_2 into $\alpha_1(x_1), \alpha_2(x_2)$; note $\alpha_1(x_1) \neq \alpha_2(x_2)$, as $A_1 \cap A_2 = A_O$. If $s \notin A$, the last node s^* of π_1 lying on π_2 is a confounder of $\alpha_1(x_1)$ and $\alpha_2(x_2)$ lying outside A (the suffixes from s^* share only s^*, since the sole A-node of each π_i is its endpoint), contradicting causal sufficiency of k (established above). If $s \in A$, then s is exclusive, say $s = \alpha_1(x_1)$ with π_2 nontrivial, and the case analysis of the injectivity proof of α applies to (x_1, x_2), with π_2 in place of the equality $s = \alpha_2(x_2)$: in case *(i)*, and in case *(ii)* with orientation $x_2 \rightsquigarrow \eta_2(u)$, $\eta_1(u) \rightsquigarrow x_1$, the arguments run verbatim after prefixing the M_2-side image path with π_2, yielding $s \in A_O$ resp. a directed cycle. In the remaining orientation $x_1 \rightsquigarrow \eta_1(u)$, $\eta_2(u) \rightsquigarrow x_2$, s reaches the distinct A_2-nodes $\alpha(\eta(u))$ (image of the M_1-witness, nodes in A_1) and $\alpha_2(x_2)$ (via π_2), each path avoiding the other's endpoint, so the preimage device via h_2 gives $s \in A_1 \cap A_2 = A_O$, contradicting exclusivity.

(G) *Every $n \in V_X \setminus A$ with a descendant in A has exactly one gate*, where a gate is the first A-node on some directed path out of n, the path truncated there (so the gate is the sole A-node of its gate path). If n had gates $g \neq g'$ with gate paths π, π', the last node n^* of π lying on π' would be a confounder of g and g' lying outside A (the suffixes from n^* share only n^*, and neither gate lies on the other path, being the sole A-node of its own), contradicting causal sufficiency of k.

(F) *No $o \in A_O$ is fed by both sides*, where side i feeds o if a directed path avoiding A_O, except at its endpoint o, runs from a node of $A_i \setminus A_O$ to o. Let sides $1, 2$ feed o from $\alpha_1(x_1), \alpha_2(x_2)$; by (U) the feeding paths stay in their own side's exclusive images and outside nodes. If x_i has an M_i-path into the overlap, truncated at $\eta_i(w_i)$, then $\alpha(\eta(w_i)) = o$: otherwise the feeding path and the image path reach two distinct A_O-nodes, each avoiding the other's endpoint, and the preimage device contradicts exclusivity of x_i. If both x_1, x_2 have such paths, both truncations enter the class of o and the entry-points device applies. If neither does, the first bullet of condition 3 fails for (x_1, x_2). So exactly one, say x_1, has one, and the first bullet for (x_1, x_2) holds only in the orientation $x_1 \rightsquigarrow \eta_1(u^*)$, $\eta_2(u^*) \rightsquigarrow x_2$. If $\eta_1(u^*)$ lies in the class of o, the image of the witness and the feeding path close the cycle $o \rightsquigarrow \alpha_2(x_2) \rightsquigarrow o$. If the witness path passes through the class of o, its suffix gives an M_1-path from that class to $\eta_1(u^*)$, closing the cycle $o \rightsquigarrow \alpha(\eta(u^*)) \rightsquigarrow \alpha_2(x_2) \rightsquigarrow o$. If it avoids that class, x_1 has M_1-paths, each avoiding the other's endpoint, to two distinct overlap nodes, so causal sufficiency of the input morphism f_1 forces an M_0-preimage of the exclusive node x_1, a contradiction.

Chain rule. Fix a topological order of V_X, write $X' := X_{\mathrm{do}(\alpha(\mathbf{v})=\beta(a))}$, and enumerate A in the induced order (a linear extension of G_P, as α preserves edges). We claim, for non-intervened $z = \alpha(u)$ and conditioning values in the image of β,

$$p^{X'}\big(x_z \mid x_{D(z)}\big) = p^P\big(x_z \mid x_{\alpha(\mathrm{pa}_P(u))}\big), \tag{7.6}$$

where $D(z) := \{\text{nodes of } A \text{ preceding } z\}$.

(Equation (7.6) is asserted at conditioning configurations of positive $p^{X'}$-mass; this suffices for the induction below, since the inductive statement (equality of the marginals on each initial segment of A) forces both sides of Eq. (7.5) to vanish simultaneously at the remaining configurations.) Granting this: intervened factors match by construction, and a simultaneous induction along the order shows the marginal of $p^{X'}$ on A is concentrated on the image of β (each factor, by Eq. (7.6), is a transported distribution and charges only the image; the first off-image coordinate receives zero mass, so both sides of Eq. (7.5) vanish off the image). Multiplying the factors yields the transported truncated factorization of P, proving Eq. (7.5).

Proof of Eq. (7.6). Let H be the set of outside ancestors of z in the mutilated graph with gate z (unique gates by (G)). *Input purity: there is a side σ, covering $\mathrm{pa}_P(u)$ in the sense of Step 1, such that all parents of z and of nodes of H lie in $H \cup A_\sigma$.* An outside parent of a node of $H \cup \{z\}$ is an ancestor of z whose unique gate is z, hence lies in H. If $z \in A_i \setminus A_O$: a parent in the other exclusive region, of z or of a node of H (whose gate path has interior outside A), yields a path contradicting (U); take $\sigma := i$, the side defining the mechanism at the exclusive node u. If $z \in A_O$: a parent of $z \cup H$ in $A_j \setminus A_O$ makes side j feed z, so by (F) at most one side σ has such parents. If side $k \neq \sigma$ were augmented at u (Step 1), its exclusive parent's image edge would feed z from side k, contradicting (F); so either σ is the augmented side, which covers, or no side is augmented and both cover with coinciding mechanisms (Step 1).

Let An be the ancestral closure of $D(z) \cup \{z\}$ in the mutilated graph, $G_z := \{z\} \cup (H \cap \mathrm{An})$, and $G_r := \mathrm{An} \setminus G_z$. No node of G_z parents a node of G_r: a child of z in An would make z an ancestor of a node preceding it; a child of an H-node in G_r gives that H-node a path to a node of A preceding z, whose first A-node is a second gate, contradicting (G). Hence, in the marginal of $p^{X'}$ on $D(z) \cup \{z\}$ (the sum over the outside coordinates of An of the product of its factors) the product splits as $K\big(x_z; x_{D(z)\cap A_\sigma}\big) \cdot M\big(x_{D(z)}\big)$, where K sums the G_z-factors over the H-coordinates (by input purity these involve only G_z- and A_σ-coordinates) and M sums the G_r-factors. Since $\sum_{x_z} K = 1$ (proper kernels, reverse topological order), the conditional equals K.

Finally, in the model $X_{\mathrm{do}(S=x_S)}$ with $S := D(z) \cap A_\sigma$, the unpinned ancestors of z are exactly the nodes of $H \cap \mathrm{An}$ with unchanged mechanisms (every other ancestral path enters through a pinned node, by input purity), so the marginal of z there is this same K. As $S \subseteq A_\sigma$ and x_S lies in the image of β, the commutation identity of h_σ identifies K with the transported marginal of $\pi_\sigma(u)$ in M_σ under $\mathrm{do}(\alpha_\sigma^{-1}(S))$; since σ covers $\mathrm{pa}_P(u)$, we have $\alpha_\sigma(\mathrm{pa}_{M_\sigma}(\pi_\sigma(u))) = \alpha(\mathrm{pa}_P(u)) \subseteq S$ and $z \notin S$, so truncation in M_σ identifies this marginal with the transported mechanism of M_σ at $\pi_\sigma(u)$, i.e., with the mechanism of P at u. □

Uniqueness: The uniqueness of k follows directly from the uniqueness of the maps in the set-pushout and domain-pushout constructions. Thus, P is the pushout. □

The following observation is used repeatedly in the proofs of Theorems 2 and 3.

Remark 7.9 (Interval fibers). The node map η of a causal abstraction morphism (Definition 6) has *interval* fibers: if $u \leq_M v \leq_M w$ and $\eta(u) = \eta(w)$, then monotonicity gives $\eta(u) \leq_N \eta(v) \leq_N \eta(w)$, hence $\eta(v) = \eta(u)$. Thus the fibers are contiguous blocks of $\leq_M$, and two distinct fibers are wholly ordered against one another. Moreover $\eta(x) <_N \eta(y)$ implies $x <_M y$, since $y \leq_M x$ would give $\eta(y) \leq_N \eta(x)$.

Throughout the remaining proofs we freely use *truncation*: in any CBN M, if $S \supseteq \mathrm{pa}_M(x)$ and $x \notin S$, then $p^M_{\mathrm{do}(S=c)}(x) = p^M(x \mid \mathrm{pa}_M(x) = c|_{\mathrm{pa}_M(x)})$, depending only on the $\mathrm{pa}_M(x)$-coordinates of c; this is immediate from the truncated factorization. The following lemma isolates the downward induction that underlies the intervention-commutation arguments below; note that 7.9 uses only monotonicity, so it applies in the setting of the lemma.

Lemma 7.10 (Co-singleton criterion)**.** *Let $(M, \leq_M)$ be an ordered CBN, $(V_N, \leq_N)$ a finite totally ordered set, $\eta : V_M \to V_N$ a surjective monotone map, and, for each $w \in V_N$, a finite domain $\mathrm{dom}_N(w)$ together with a surjective component map $\tau_w : \mathrm{dom}_M(\eta^{-1}(w)) \to \mathrm{dom}_N(w)$; write $\tau := \prod_w \tau_w$. Let E_N be a set of directed edges on V_N for which $\leq_N$ is topological, and suppose that for every $w \in V_N$ there is a kernel $p^N(w \mid \mathrm{pa}_{E_N}(w) = \cdot)$ such that, for every $\xi \in \mathrm{dom}_M(\eta^{-1}(V_N \setminus \{w\}))$,*

$$\begin{aligned}\tau_{w,*}\, p^M_{\mathrm{do}(\eta^{-1}(V_N\setminus\{w\})=\xi)}\big(\eta^{-1}(w)\big) \\ = p^N\big(w \mid \mathrm{pa}_{E_N}(w) = \tau(\xi)|_{\mathrm{pa}_{E_N}(w)}\big),\end{aligned} \tag{7.7}$$

i.e., the transported pinned response of each fiber depends only on the transported E_N-parent values. Then:

1. *For every $T \subseteq V_N$ and $b \in \mathrm{dom}_M(\eta^{-1}(T))$,*

$$\tau_* \, p^M_{\mathrm{do}(\eta^{-1}(T)=b)} \;=\; p^N_{\mathrm{do}(T=\tau(b))}, \qquad (7.8)$$

the right-hand side denoting the truncated factorization of the kernels p^N over (V_N, E_N).

2. *Let $E'_N \subseteq E_N$ be obtained by removing every edge $x \to w$ such that $p^N(w \mid \cdot)$ is constant in the x-coordinate given the remaining coordinates. This pruning is unambiguous (a kernel constant in each of several coordinates separately is constant in them jointly, the domains being finite), $N := (V_N, E'_N, \mathrm{dom}_N, p^N)$ is a well-defined CBN, and Eq.* (7.8) *holds verbatim with E'_N in place of E_N.*

3. *$(\eta, \tau) : (M, \leq_M) \to (N, \leq_N)$ is a morphism in $\mathbf{CM}_{abs}$; in particular η reflects the edges of E'_N. If, moreover, the data $(V_N, E_N, \mathrm{dom}_N, p^N)$ are those of a CBN given in advance, then $E'_N = E_N$ by its minimality, and (η, τ) is a morphism into that CBN.*

Proof. By 7.9 the η-fibers are $\leq_M$-intervals, wholly ordered against one another, and every M-edge between distinct fibers ascends in $\leq_N$: $x \to y \in E_M$ with $\eta(x) \neq \eta(y)$ gives $x <_M y$, so the fiber of x wholly precedes that of y. Hence, whenever u is the $\leq_N$-maximum of a set of unpinned nodes, no unpinned M-mechanism outside $F_u := \eta^{-1}(u)$ reads a coordinate of F_u (its node would lie in a later, pinned fiber), and F_u has no unpinned E_N-children.

(i) Downward induction on $|V_N \setminus T|$; for $T = V_N$ both sides are the point mass at $\tau(b)$. Let u be the $\leq_N$-maximum of $V_N \setminus T$ and $\mu_T := \tau_* p^M_{\mathrm{do}(\eta^{-1}(T)=b)}$. *Marginal:* summing the F_u-factors of the truncated factorization in reverse $\leq_M$-order shows the fine joint of the remaining fibers is unchanged by additionally pinning F_u at any ξ; hence $\mu_T(x_{V_N \setminus u}) = \mu_{T \cup \{u\}}(x_{V_N \setminus u})$, which by the induction hypothesis equals $p^N_{\mathrm{do}(T \cup \{u\} = (\tau(b), \tau_u(\xi)))}(x_{V_N \setminus u}) = p^N_{\mathrm{do}(T=\tau(b))}(x_{V_N \setminus u})$, the last step because all E_N-children of u lie in T. *Conditional:* fix x with $\mu_T(x_{V_N \setminus u}) > 0$. Since no unpinned factor outside F_u reads F_u, the truncated factorization splits, and the conditional of x_{F_u} given the fine coordinates ζ of all other fibers (at positive-mass ζ) is, by truncation, the response $p^M_{\mathrm{do}(\eta^{-1}(V_N \setminus \{u\})=(b,\zeta))}(x_{F_u})$. By Eq. (7.7) its τ_u-pushforward equals $p^N(\cdot \mid \tau(b,\zeta)|_{\mathrm{pa}_{E_N}(u)})$, a function of ζ only through $\tau(\zeta)$; being constant on τ-classes, it equals the conditional of the coarse u-coordinate given the coarse values, i.e., $\mu_T(x_u \mid x_{V_N \setminus u}) = p^N(x_u \mid x_{\mathrm{pa}_{E_N}(u)})$. Multiplying the two parts yields Eq. (7.8) at such x; where $\mu_T(x_{V_N \setminus u}) = 0$, both sides vanish by the marginal identity.

(ii) If $p^N(w \mid \cdot)$ is constant in each coordinate of a set D of parents separately, changing one coordinate at a time shows it is constant in the D-coordinates jointly; hence the pruned kernel is a well-defined function of the surviving parents, the truncated factorizations over E_N and E'_N coincide, and Eq. (7.8) persists. Each kernel is a probability distribution (a pushforward of one, by Eq. (7.7)), (V_N, E'_N) is a DAG, and N is minimal: for a surviving edge $x \to w$, intervening on $\mathrm{pa}_{E'_N}(w)$ and varying the x-value changes the interventional marginal of w, which a model with the edge removed cannot reproduce.

(iii) η is surjective and monotone and τ fiberwise surjective by hypothesis, and Eq. (7.8), for T ranging over all subsets of $V_N = \eta(V_M)$, is precisely the intervention commutation of Definition 4. For edge reflection, let $x \to w \in E'_N$ and suppose no M-edge runs from $\eta^{-1}(x)$ into F_w. Then the response of F_w in $p^M_{\mathrm{do}(\eta^{-1}(V_N \setminus \{w\})=\xi)}$ is a product of mechanisms none of which reads $\eta^{-1}(x)$, hence constant in the $\eta^{-1}(x)$-coordinates of ξ; but by Eq. (7.7) its τ_w-pushforward is the kernel at $\tau(\xi)|_{\mathrm{pa}_{E'_N}(w)}$, which genuinely varies with the x-coordinate (minimality, by (ii)) as $\xi|_{\eta^{-1}(x)}$ ranges over preimages of distinct values (surjectivity of τ_x), a contradiction. The final claim holds since a CBN given in advance is minimal by Definition 3, so no edge is pruned. $\square$

Proof of Theorem 2.

Proof. Throughout we freely use truncation, the interval-fiber property (7.9), and the co-singleton criterion (7.10). We also note that when both fibers over some $c \in V_{M_3}$ are singletons with bijective fiber domain maps, the construction of $(P, \leq_P)$ depends on the choice of $i(c)$; the universal property established below implies that any two choices yield canonically isomorphic pullbacks, so this ambiguity is harmless.

Step 1: $(P, \leq_P)$ is a well-defined ordered CBN.

(1a) $\leq_P$ is topological for E_{pr}; acyclicity is automatic. Let $u \to w \in E_{pr}$, so $\pi_{k(w)}(u) \to \pi_{k(w)}(w) \in E_{M_{k(w)}}$ and hence $\pi_{k(w)}(u) <_{k(w)} \pi_{k(w)}(w)$. Monotonicity of $f_{k(w)}$ gives $\pi(u) \leq_3 \pi(w)$. If the latter is strict, $u <_P w$ by definition; if $\pi(u) = \pi(w)$, then $k(u) = k(w)$ and the tie-break gives $u <_P w$ directly. Hence E_{pr}, and a fortiori E_P, is acyclic.

(1b) The mechanisms Eq. (5.1) *are well-defined.* By construction, $\mathrm{pa}_{pr}(u)$ consists of the full π_k-fibers over $\mathrm{pa}_{M_k}(v_k)$, so for each $x \in \mathrm{pa}_{M_k}(v_k)$ the joint map $\theta_{k,x}$ has all its arguments among the coordinates of b, and the extraction $\theta_k(b)$ is defined without choices. Since $\tau_{i(c),c}$ is bijective, $\mathrm{dom}_P(u) = \mathrm{dom}_{M_k}(v_k)$, and the mechanism of P at u is an exact copy of the mechanism of M_k at v_k. The mechanism need not depend on every coordinate of a parent fiber (a joint map $\theta_{k,x}$ may ignore some of them), which is why the pruning defining E_P can remove edges within lifted fibers.

(1c) Structure of E_P and minimality. Since $E_P \subseteq E_{pr}$, every edge $w \to u \in E_P$ projects under the detailed side of its head to an edge of $E_{M_{k(u)}}$. Conversely, for every $x \in \mathrm{pa}_{M_k}(v_k)$, minimality of M_k guarantees that Eq. (5.1) depends on the extracted x-coordinate, hence on at least one coordinate of the fiber over x, so at least one edge from that fiber into u survives the pruning. (The pruning is unambiguous: a mechanism constant in each of several parent coordinates separately, the others held fixed, is constant in them jointly, the domains being finite; cf. 7.10(ii).) By construction of E_P, P satisfies the minimality condition of Definition 3, so P is a well-defined CBN and, with (1a), a well-defined ordered CBN.

Step 2: p_1 is an abstraction morphism (and symmetrically p_2).

(2a) Surjectivity and fiberwise factorization. π_1 is the pullback in $\mathbf{Set}$ of surjections, hence surjective; θ_1 factors fiberwise over π_1 by construction, with fiberwise surjective components (identities, or composites of a surjection with a bijection).

(2b) Monotonicity of π_1. Let $u <_P w$. If $\pi(u) <_3 \pi(w)$, then $\pi_1(u) <_1 \pi_1(w)$ by the last sentence of 7.9 applied to f_1 (whose node map satisfies $\eta_1(\pi_1(u)) = \pi(u) <_3 \pi(w) = \eta_1(\pi_1(w))$). If $\pi(u) = \pi(w) =: c$: for $k(c) = 1$ the tie-break is $\leq_1$ itself; for $k(c) = 2$ we have $\pi_1(u) = \pi_1(w)$ (the singleton fiber).

(2c) A mechanism identity at bijective-side nodes. Let $x \in V_{M_1}$ with $c := \eta_1(x)$ and $i(c) = 1$, so $\eta_1^{-1}(c) = \{x\}$ and $\tau_{1,c}$ is a bijection. Then

$$\begin{aligned} &\eta_1\big(\mathrm{pa}_{M_1}(x)\big) = \mathrm{pa}_{M_3}(c), \\ &p^{M_1}\big(x = a \mid \mathrm{pa}_{M_1}(x) = m\big) \\ &\quad = p^{M_3}\big(c = \tau_{1,c}(a) \mid \mathrm{pa}_{M_3}(c) = t(m)\big), \end{aligned} \tag{7.9}$$

where $t(m)$ is the restriction to $\mathrm{pa}_{M_3}(c)$ of $\tau_1(m')$ for any extension m' of m to the η_1-saturation of $\mathrm{pa}_{M_1}(x)$; the proof shows the right-hand side does not depend on the chosen extension. *Proof.* We prove the inclusion $\supseteq$, then the identity using only $\supseteq$, then the reverse inclusion as a corollary; the order matters. $\supseteq$: an edge $d \to c \in E_{M_3}$ is reflected (Definition 6) to an edge $y \to x' \in E_{M_1}$ with $\eta_1(x') = c$; the singleton fiber forces $x' = x$, so $\eta_1(y) = d$ with $y \in \mathrm{pa}_{M_1}(x)$. Identity: let $S := \eta_1^{-1}(\eta_1(\mathrm{pa}_{M_1}(x))) \supseteq \mathrm{pa}_{M_1}(x)$; note $x \notin S$ (a parent with image c would equal x, a self-loop). For $m' \in \mathrm{dom}_{M_1}(S)$, truncation in M_1 gives $p^{M_1}_{\mathrm{do}(S=m')}(x) = p^{M_1}(x \mid \mathrm{pa}_{M_1}(x) = m'|_{\mathrm{pa}})$, while intervention commutation of f_1 (Definition 4) and bijectivity of $\tau_{1,c}$ give $p^{M_1}_{\mathrm{do}(S=m')}(x = a) = p^{M_3}_{\mathrm{do}(\eta_1(\mathrm{pa}_{M_1}(x))=\tau_1(m'))}(c = \tau_{1,c}(a))$. The left-hand side depends on m' only through $m'|_{\mathrm{pa}}$, the right-hand side only through $\tau_1(m')$; truncation in M_3 (using $\supseteq$ and $c \notin \eta_1(\mathrm{pa}_{M_1}(x))$) turns the right-hand side into $p^{M_3}(c = \tau_{1,c}(a) \mid \mathrm{pa}_{M_3}(c) = \cdot)$ at the corresponding restriction, proving the identity and the extension-independence of t. $\subseteq$: if some $y \in \mathrm{pa}_{M_1}(x)$ had $\eta_1(y) \notin \mathrm{pa}_{M_3}(c)$, varying the y-coordinate of m' perturbs $\tau_1(m')$ only in a component that t discards, so by the identity the mechanism of M_1 at x is constant in the y-coordinate, contradicting minimality. □

The same argument applied to f_2 shows, for the detailed fiber $F := \eta_2^{-1}(c)$: pinning the η_2-saturation of all external M_2-parents of F at fine values and pushing the joint response of F forward along $\tau_{2,c}$ yields $p^{M_3}(c \mid \mathrm{pa}_{M_3}(c) = \cdot)$ at the transported values (the truncated factorization of the sub-network F is pushed through the commutation identity of f_2 and truncated in M_3, using $\eta_2(\text{external parents of } F) \supseteq \mathrm{pa}_{M_3}(c)$ by edge reflection of f_2).

(2d) The co-singleton hypothesis holds for $(\pi_1, \theta_1) : P \to M_1$. Let $x \in V_{M_1}$, $c := \eta_1(x)$, and pin all fibers over $V_{M_1} \setminus \{x\}$ at a fine configuration ξ. If $i(c) = 2$, the fiber $\pi_1^{-1}(x)$ is a single node with detailed side 1 and $\theta_{1,x}$ is the identity; by truncation in P its response is the mechanism Eq. (5.1), an exact copy of the mechanism of M_1 at x evaluated at $\theta_1(\xi)|_{\mathrm{pa}_{M_1}(x)}$. If $i(c) = 1$, the fiber carries the M_2-copies of the detailed fiber $F = \eta_2^{-1}(c)$; its response is the product of these mechanisms at the extracted values, and by the second half of (2c) its $\theta_{1,c}$-pushforward (i.e., $\tau_{1,c}^{-1} \circ \tau_{2,c}$) equals $p^{M_3}(c \mid \mathrm{pa}_{M_3}(c) = \cdot)$ transported, which by Eq. (7.9) is the mechanism of M_1 at x at the extracted $\mathrm{pa}_{M_1}(x)$-values, using $\mathrm{pa}_{M_3}(c) = \eta_1(\mathrm{pa}_{M_1}(x))$.

(2e) Conclusion. π_1 is surjective and monotone with fiberwise surjective θ_1 (2a–2b), and M_1 is a CBN; by (2d) and 7.10, p_1 commutes with all interventions as in Definition 4 and reflects the edges of E_{M_1}, so p_1 is an abstraction morphism.

Step 3: Commutativity. On nodes, $\eta_1\pi_1 = \pi = \eta_2\pi_2$ by the definition of V_P. On domains, $\tau_1 \circ \theta_1 = \tau_2 \circ \theta_2$ fiberwise: on the fiber product presentation of $\mathrm{dom}_P(\pi^{-1}(c))$, both composites are the structure map to $\mathrm{dom}_{M_3}(c)$. Hence $f_1 p_1 = f_2 p_2$.

Step 4: Universal property. Let $(X, \leq_X)$ be an ordered CBN with abstraction morphisms $g_j = (\alpha_j, \beta_j) : X \to M_j$ satisfying $f_1 g_1 = f_2 g_2$.

(4a) The candidate is forced. Any (α, β) with $p_j \circ (\alpha, \beta) = g_j$ must satisfy $\alpha(z) = (\alpha_1(z), \alpha_2(z))$ (the unique solution of $\pi_j \alpha = \alpha_j$ in the set pullback), and, for $w \in V_P$ with $k := k(w)$, since the π_k-fiber of w is a singleton, $\alpha^{-1}(w) = \alpha_k^{-1}(\pi_k(w))$ and $\theta_k \circ \beta = \beta_k$ forces $\beta_w = \beta_{k,\pi_k(w)}$, the g_k-fiber map. In particular α is surjective (α_k is), β factors fiberwise with surjective fiber maps, the compatibility $\theta_j \circ \beta = \beta_j$ for the other index holds by $f_1 g_1 = f_2 g_2$, and uniqueness is settled. α is monotone: for $z <_X z'$, monotonicity of $\eta_j \alpha_j$ gives $\pi(\alpha(z)) \leq_3 \pi(\alpha(z'))$; if strict,

$\alpha(z) <_P \alpha(z')$; if equal, monotonicity of g_k for the detailed side k of the common coarse node gives $\alpha_k(z) \leq_k \alpha_k(z')$, which is the tie-break.

(4b) Pointwise rigidity. For every $w \in V_P$ and every fine configuration ξ of the fibers over $V_P \setminus \{w\}$: the set $V_P \setminus \{w\}$ is a union of $\pi_{k(w)}$-fibers (the removed node is a singleton $\pi_{k(w)}$-fiber), so its α-preimage equals $\alpha_{k(w)}^{-1}(\pi_{k(w)}(V_P \setminus \{w\}))$ and the commutation identity of $g_{k(w)}$ applies to pinning it at ξ. Pushing the response of $\alpha^{-1}(w)$ forward along β_w and truncating in $M_{k(w)}$ (everything else pinned) yields the mechanism of $M_{k(w)}$ at $\pi_{k(w)}(w)$ (by Step 1 the mechanism of P at w) evaluated at the values of the pins extracted through $\beta_{k(w)} = \theta_{k(w)} \circ \beta$, and depending on ξ only through the extractions at the E_P-parents of w. This holds pointwise for every ξ.

(4c) Conclusion via the co-singleton criterion. By (4a), α is surjective and monotone and β factors fiberwise with surjective fiber maps; (4b) states exactly the co-singleton hypothesis Eq. (7.7) for $(\alpha, \beta) : X \to P$ with target the CBN $(P, \leq_P)$ of Step 1, whose edge set E_P is minimal. 7.10 therefore yields the intervention commutation required by Definition 4 and reflection of the edges of E_P. Together with (4a), (α, β) is an abstraction morphism, unique with $p_j \circ (\alpha, \beta) = g_j$; hence $(P, \leq_P)$ is the pullback. □

Proof of Theorem 3

Proof. **Step 1: Well-definedness of P.** *(1o) Order, interval fibers, and acyclicity.* Every q-fiber is a union of η_1- and η_2-fibers of V_{M_0}: the equivalence classes of $\sim$ are the connected components of the hypergraph on V_{M_0} whose hyperedges are these fibers. By 7.9, each such fiber is a $\leq_0$-interval, and since a union of intervals of a total order whose intersection graph is connected is again an interval, every q-fiber is a $\leq_0$-interval. Distinct fibers are disjoint intervals, hence wholly ordered against one another, so the induced order $\leq_P$ is a well-defined total order on V_P; in particular q is monotone for $\leq_P$. Every E_{pr}-edge ascends in $\leq_P$: an edge $u \to v \in E_{pr}$ with $u \neq v$ is the g_i-image of some edge $x_i \to y_i \in E_{M_i}$, so $x_i <_i y_i$; choosing preimages $a \in \eta_i^{-1}(x_i)$ and $b \in \eta_i^{-1}(y_i)$ (surjectivity), the last sentence of 7.9 gives $a <_0 b$, and since $q(a) = u \neq v = q(b)$ lie in distinct, wholly ordered fibers, $u <_P v$. Hence (V_P, E_{pr}) is acyclic and $\leq_P$ is topological for E_{pr}, and a fortiori for $E_P \subseteq E_{pr}$; so $(P, \leq_P)$ is an ordered CBN once its mechanisms are shown to be well-defined, which occupies the remainder of this step. We establish three facts that carry the remainder of the proof. Throughout, for $S \subseteq V_P$ we write $q^{-1}(S) \subseteq V_{M_0}$ for the union of fibers and note $q^{-1}(S) = \eta_i^{-1}(g_i^{-1}(S))$, so that interventions on $q^{-1}(S)$ are η_i-saturated for both i; hence the commutation identity of f_i (Definition 4) applies to them.

(1a) Homogeneity. For every $S \subseteq V_P$ and $b, b' \in \mathrm{dom}_{M_0}(q^{-1}(S))$ with $\tau_P(b) = \tau_P(b')$:

$$\tau_{P,*}\, p^{M_0}_{\mathrm{do}(q^{-1}(S)=b)} \;=\; \tau_{P,*}\, p^{M_0}_{\mathrm{do}(q^{-1}(S)=b')}.$$

Proof. If $\tau_1(b) = \tau_1(b')$, then by commutation of f_1 both sides equal $\lambda_{1,*}\, p^{M_1}_{\mathrm{do}(g_1^{-1}(S)=\tau_1(b))}$, using $\tau_P = \lambda_1 \circ \tau_1$; likewise for τ_2. In general, $\tau_P(b) = \tau_P(b')$ means b and b' are connected, componentwise within the fibers of the intervened blocks, by a finite zigzag of identifications each of which is certified by equality of the full tuple under τ_1 or under τ_2 (changing one block's fine value at a time while keeping all others fixed preserves the full-tuple τ_i-image on the unchanged blocks). Chaining the per-step equalities proves the claim. □

In particular, the mechanisms Eq. (5.2) do not depend on the chosen representative b, and each is a probability distribution on $\mathrm{dom}_P(u)$ (the pushforward of a probability measure).

(1b) Locality. For every $u \in V_P$, every $S \supseteq \mathrm{pa}_{pr}(u)$ with $u \notin S$, and every b: the u-marginal $\tau_{P,u,*}\, p^{M_0}_{\mathrm{do}(q^{-1}(S)=b)}\big(q^{-1}(u)\big)$ depends only on the restriction of $\tau_P(b)$ to $\mathrm{pa}_{pr}(u)$ and equals the mechanism Eq. (5.2) at that restriction. *Proof.* By commutation of f_1, the quantity equals $\lambda_{1,u,*}$ of the marginal of $p^{M_1}_{\mathrm{do}(g_1^{-1}(S)=\tau_1(b))}$ on the block $B_1(u) := \eta_1(q^{-1}(u)) = g_1^{-1}(u)$. Every M_1-edge from another block into $B_1(u)$ induces, by the definition of E_{pr}, an edge of E_{pr} into u; hence all external M_1-parents of $B_1(u)$ lie in blocks of $\mathrm{pa}_{pr}(u) \subseteq S$ and are pinned. By the truncated factorization of the intervened M_1, the marginal of $B_1(u)$ is the product of the mechanisms of its nodes with external parents evaluated at the pinned values; in particular it depends only on the coordinates of $\tau_1(b)$ on the blocks of $\mathrm{pa}_{pr}(u)$, and coincides with the corresponding marginal for $S = \mathrm{pa}_{pr}(u)$. Finally, dependence only on $\tau_1(b)|_{\mathrm{pa}_{pr}(u)}$ upgrades to dependence only on $\tau_P(b)|_{\mathrm{pa}_{pr}(u)}$: given b, b' with equal τ_P-restrictions to $\mathrm{pa}_{pr}(u)$, pass from b to the hybrid $\tilde{b}$ agreeing with b on $q^{-1}(\mathrm{pa}_{pr}(u))$ and with b' elsewhere (no change by the τ_1-locality just proven), then from $\tilde{b}$ to b' by a zigzag within the $\mathrm{pa}_{pr}(u)$-fibers (no change by homogeneity (1a) applied to the full pushforward, hence to the u-marginal). □

(1c) Family identity. For every $T \subseteq V_P$ and $b \in \mathrm{dom}_{M_0}(q^{-1}(T))$:

$$\tau_{P,*}\, p^{M_0}_{\mathrm{do}(q^{-1}(T)=b)} \;=\; p^{P}_{\mathrm{do}(T=\tau_P(b))}, \tag{7.10}$$

where the right-hand side is the truncated factorization of the mechanisms Eq. (5.2). In particular ($T = \emptyset$) the mechanisms of P factorize the pushforward $\tau_{P,*}p^{M_0}$, so P is a well-defined CBN (minimality holds by the pruning defining E_P, which is unambiguous by 7.10(ii)).

Proof. This is 7.10(i) applied to M_0 with $\eta := q$, $\tau := \tau_P$, target order $(V_P, \leq_P)$, and candidate edge set E_{pr}: q is surjective and monotone with fiberwise surjective τ_P,

$\leq_P$ is topological for E_{pr} (Step (1o)), and the co-singleton hypothesis Eq. (7.7) is precisely the case $S = V_P \setminus \{u\}$ of Locality (1b), with kernels the mechanisms Eq. (5.2), whose independence of the chosen representative is Homogeneity (1a). Moreover, 7.10(ii) shows that the pruning defining E_P is unambiguous and that Eq. (7.10) holds verbatim with E_P in place of E_{pr}, and 7.10(iii) exhibits $(q, \tau_P) : M_0 \to P$ as a morphism in $\mathbf{CM}_{abs}$. $\square$

Step 2: Commutativity. We must show $h_1 \circ f_1 = h_2 \circ f_2$.

- **Nodes:** $g_1 \circ \eta_1 = q$ and $g_2 \circ \eta_2 = q$ by the definition of the set-pushout quotient.
- **Domains:** $\lambda_1 \circ \tau_1 = \tau_P$ and $\lambda_2 \circ \tau_2 = \tau_P$ by the definition of the domain pushout.

Thus, the diagram commutes.

Step 3: Morphism Properties. We verify $h_1 : M_1 \to P$ is a morphism in CM_{abs} (the case for h_2 is symmetric).

1. h_1 is a causal map: By construction λ_1 factorizes fiberwise (via the joint fiber maps over the g_1-fibers). For intervention commutation, let $S \subseteq V_P$ and $m \in \mathrm{dom}_{M_1}(g_1^{-1}(S))$. Choose $b \in \mathrm{dom}_{M_0}(q^{-1}(S))$ with $\tau_1(b) = m$ (τ_1 is surjective). Then
$$\lambda_{1,*}\, p^{M_1}_{\mathrm{do}(g_1^{-1}(S)=m)} \overset{(f_1)}{=} \lambda_{1,*}\, \tau_{1,*}\, p^{M_0}_{\mathrm{do}(q^{-1}(S)=b)} = \tau_{P,*}\, p^{M_0}_{\mathrm{do}(q^{-1}(S)=b)} \overset{(1c)}{=} p^P_{\mathrm{do}(S=\lambda_1(m))},$$
using the commutation identity of f_1 (the set $q^{-1}(S) = \eta_1^{-1}(g_1^{-1}(S))$ is η_1-saturated), $\tau_P = \lambda_1 \circ \tau_1$, the family identity Eq. (7.10), and $\tau_P(b) = \lambda_1(\tau_1(b)) = \lambda_1(m)$.
2. h_1 is a causal abstraction: Since η_1, η_2 are surjective and V_P is the pushout with regard to the underlying set of nodes with set maps g_1 and g_2, g_1 and g_2 are also surjective. Since the domains are defined as pushouts of fiberwise surjective maps, λ_1, λ_2 are also fiberwise surjective. Moreover, g_1 is monotone: let $x \leq_1 x'$ in V_{M_1}; if $g_1(x) = g_1(x')$ there is nothing to show, and if $x <_1 x'$ with $g_1(x) \neq g_1(x')$, choose preimages $a \in \eta_1^{-1}(x)$ and $a' \in \eta_1^{-1}(x')$; the last sentence of 7.9 gives $a <_0 a'$, and since the fibers $q^{-1}(g_1(x))$ and $q^{-1}(g_1(x'))$ are distinct and wholly ordered (Step 1), $g_1(x) <_P g_1(x')$. It remains to show that g_1 reflects the edges of E_P (note that this requires an argument: an edge of E_{pr} may originate from M_2 only, and only the edges surviving the pruning need to be reflected). Let $x \to u \in E_P$. By minimality (the pruning defining E_P), the mechanism Eq. (5.2) at u genuinely depends on the x-coordinate of its argument. By the proof of (1b), this mechanism is the $\lambda_{1,u}$-pushforward of the product of the M_1-mechanisms of the block $B_1(u)$, which depends only on the values of the M_1-*parent blocks* of $B_1(u)$. Hence x is an M_1-parent block of $B_1(u)$: there is an edge $x_1 \to u_1 \in E_{M_1}$ with $g_1(x_1) = x$ and $g_1(u_1) = u$, as required. (The symmetric argument through M_2 shows that every E_P-edge is likewise reflected by g_2; in particular, surviving edges are witnessed by edges on *both* sides.)

Step 4: Universal Property. Let $(X, \leq_X)$ be any ordered CBN equipped with morphisms $k_1 = (\alpha_1, \beta_1) : M_1 \to X$ and $k_2 = (\alpha_2, \beta_2) : M_2 \to X$ in $\mathbf{CM}_{abs}$ such that $k_1 \circ f_1 = k_2 \circ f_2$. Because $f_1 = (\eta_1, \tau_1)$ and $f_2 = (\eta_2, \tau_2)$, this commutativity explicitly means $\alpha_1 \circ \eta_1 = \alpha_2 \circ \eta_2$ for the node maps and $\beta_1 \circ \tau_1 = \beta_2 \circ \tau_2$ for the domain maps. We must show there exists a unique morphism $u = (u_V, u_{dom}) : P \to X$ in $\mathbf{CM}_{abs}$ such that $u \circ h_1 = k_1$ and $u \circ h_2 = k_2$, where $h_i = (g_i, \lambda_i)$.

1. Candidate: Because V_P is defined as the pushout in the category of Sets and we have $\alpha_1 \circ \eta_1 = \alpha_2 \circ \eta_2$ as maps on V_0, there exists a unique function $u_V : V_P \to V_X$ such that $u_V \circ g_1 = \alpha_1$ and $u_V \circ g_2 = \alpha_2$.
   For the domain map we argue by descent through λ_1: we claim β_1 is constant on the fibers of λ_1. Indeed, define φ on $\mathrm{dom}_{M_1}(V_{M_1}) \sqcup \mathrm{dom}_{M_2}(V_{M_2})$ by $\varphi|_{\mathrm{dom}_{M_1}} := \beta_1$ and $\varphi|_{\mathrm{dom}_{M_2}} := \beta_2$; exactly as in the proof of Homogeneity (1a), two tuples with the same image in $\prod_u \mathrm{dom}_P(u)$ are connected, blockwise and changing one block at a time, by a finite zigzag of identifications, each certified by a pair $\tau_1(c), \tau_2(c)$ with $c \in \mathrm{dom}_{M_0}(V_{M_0})$ (extended to full tuples by choosing common M_0-preimages on the unchanged blocks, using surjectivity of the τ_i); along each such link the cocone identity $\beta_1 \circ \tau_1 = \beta_2 \circ \tau_2$ makes φ constant. Hence there is a unique map $u_{\mathrm{dom}} : \mathrm{dom}_P(V_P) \to \mathrm{dom}_X(V_X)$ with $u_{\mathrm{dom}} \circ \lambda_1 = \beta_1$ (λ_1 being surjective), and $u_{\mathrm{dom}} \circ \lambda_2 = \beta_2$ follows: for $m_2 \in \mathrm{dom}_{M_2}(V_{M_2})$ choose c with $\tau_2(c) = m_2$, so $u_{\mathrm{dom}}(\lambda_2(m_2)) = u_{\mathrm{dom}}(\tau_P(c)) = u_{\mathrm{dom}}(\lambda_1(\tau_1(c))) = \beta_1(\tau_1(c)) = \beta_2(m_2)$. The same descent applied to the joint fiber map $\beta_{1,w}$ of k_1 over each $\alpha_1^{-1}(w) = g_1^{-1}(u_V^{-1}(w))$, $w \in V_X$, produces components $u_{\mathrm{dom},w} : \mathrm{dom}_P(u_V^{-1}(w)) \to \mathrm{dom}_X(w)$ that assemble to u_{dom}; hence u_{dom} factors fiberwise over u_V, with surjective components since the $\beta_{1,w}$ are surjective. We define our candidate morphism as $u = (u_V, u_{dom})$. It is unique: any morphism u' with $u' \circ h_1 = k_1$ satisfies $u'_V \circ g_1 = \alpha_1$ and $u'_{\mathrm{dom}} \circ \lambda_1 = \beta_1$ with g_1, λ_1 surjective, so u is unique if it is a valid morphism in $\mathbf{CM}_{abs}$.
2. u is a morphism in $\mathbf{CM}_{abs}$:
   - u commutes with interventions: By Definition 4 we must show, for every $\mathbf{v} \subseteq V_X$ and $t \in \mathrm{dom}_P(u_V^{-1}(\mathbf{v}))$, that $u_{\mathrm{dom},*}\, p^P_{\mathrm{do}(u_V^{-1}(\mathbf{v})=t)} = p^X_{\mathrm{do}(\mathbf{v}=u_{\mathrm{dom}}(t))}$. Set $S := u_V^{-1}(\mathbf{v})$ and note $q^{-1}(S) = (\alpha_1 \circ \eta_1)^{-1}(\mathbf{v})$, a saturated interventional set for the composite $k_1 \circ f_1 : M_0 \to X$,

which is a morphism in $\mathbf{CM}_{abs}$ by Proposition 4.1, with node map $\alpha_1 \circ \eta_1$ and domain map $\beta_1 \circ \tau_1$. Choose $b \in \mathrm{dom}_{M_0}(q^{-1}(S))$ with $\tau_P(b) = t$ (τ_P is surjective). Using the family identity Eq. (7.10), $u_{\mathrm{dom}} \circ \tau_P = u_{\mathrm{dom}} \circ \lambda_1 \circ \tau_1 = \beta_1 \circ \tau_1$, and the commutation identity of $k_1 \circ f_1$:

$$\begin{aligned} u_{\mathrm{dom},*}\, p^{P}_{\mathrm{do}(S=t)} &= u_{\mathrm{dom},*}\, \tau_{P,*}\, p^{M_0}_{\mathrm{do}(q^{-1}(S)=b)} \\ &= (\beta_1 \circ \tau_1)_*\, p^{M_0}_{\mathrm{do}((\alpha_1\circ\eta_1)^{-1}(\mathbf{v})=b)} \\ &= p^{X}_{\mathrm{do}(\mathbf{v}=\beta_1(\tau_1(b)))}, \end{aligned}$$

and $\beta_1(\tau_1(b)) = u_{\mathrm{dom}}(\tau_P(b)) = u_{\mathrm{dom}}(t)$, as required. (For non-saturated S the set $g_1^{-1}(S)$ need not be α_1-saturated, so the commutation identity of k_1 would not apply; Definition 4 only requires the saturated case.)

- u_V reflects edges: Let $x \to y \in E_X$ and suppose, for contradiction, that there is no E_P-edge from $u_V^{-1}(x)$ to $u_V^{-1}(y)$. Consider the interventional set $V_P \setminus u_V^{-1}(y)$ in P, pinned at arbitrary values, and the response of the fiber $u_V^{-1}(y)$. On the P-side, every node of $u_V^{-1}(y)$ has all its parents outside the fiber pinned, so the joint response is the product of the mechanisms of the fiber's nodes; since none of these has a parent in $u_V^{-1}(x)$, the response, and hence its u_{dom}-pushforward, is constant in the pinned $u_V^{-1}(x)$-values. On the X-side, by the commutation just established (applied to $\mathbf{v} = V_X \setminus \{y\}$, so that $S = u_V^{-1}(\mathbf{v}) = V_P \setminus u_V^{-1}(y)$ is saturated), this pushforward equals the response of y in X under $\mathrm{do}(V_X \setminus \{y\})$, which by truncation is the mechanism $p^X(y \mid \mathrm{pa}_X(y))$, and this genuinely depends on the x-coordinate by minimality of X (Definition 3), since $x \in \mathrm{pa}_X(y)$. As the pinned values of $u_V^{-1}(x)$ range over preimages of distinct x-values (surjectivity of u_{dom}), this is a contradiction. Hence u_V reflects edges.
- u_V is monotone: let $u \leq_P v$ in V_P; for $u = v$ this is trivial, and for $u <_P v$ any $a \in q^{-1}(u)$ and $b \in q^{-1}(v)$ satisfy $a <_0 b$ (the fibers are wholly ordered, Step 1), whence $u_V(u) = \alpha_1(\eta_1(a)) \leq_X \alpha_1(\eta_1(b)) = u_V(v)$ by monotonicity of the composite $\alpha_1 \circ \eta_1$.
- u_V, u_{dom} are surjective: Because k_1 is a causal abstraction, its node map α_1 is surjective. Since $\alpha_1 = u_V \circ g_1$, it follows necessarily that u_V must also be surjective. An identical argument via $\beta_1 = u_{dom} \circ \lambda_1$ proves u_{dom} is surjective.
- Fiberwise Factorization: Established in the construction of the candidate above (the descent through λ_1 was performed per u_V-fiber, with surjective components).

□