Partially Observed Structural Causal Models

Turan Orujlu[1,2] **Jordan Matelsky**[3] **Martin V. Butz**[1] **Charley M. Wu**[4,5] **Konrad P. Kording**[3]

[1]University of Tübingen
[2]MPI for Biological Cybernetics
[3]University of Pennsylvania
[4]TU Darmstadt
[5]Hessian.AI

Abstract

Here we introduce Partially Observed Structural Causal Models (POSCMs) that formalize causal systems where latent contexts co-determine both the interaction structure and downstream mechanisms on observed variables. POSCMs provide an extension of structural causal models (SCMs), as a self-contained causal modeling framework for endogenous graphs, allowing for an intervention hierarchy spanning node- and edge-level context and endogenous variable interventions. To enable surgical edge interventions, we adopt a Kolmogorov-Arnold-Sprecher edge-functional decomposition, an existence theorem for representing each node mechanism as a sum of univariate functions of its parents, yielding an explicit parametrization of dyadic functional contributions. We provide an identifiability theory that clarifies which intervention families would suffice to disentangle structure formation from mechanisms. We empirically validate these predictions in a biophysically detailed virtual human retina simulator, constructing intervention protocols that (i) reproduce the non-identifiability predicted when context is latent and no context-level interventions are available, (ii) exhibit structure-mechanism confounding under latent edges when only node interventions are observed, and (iii) recover synaptic input-output relationships via targeted node interventions, consistent with our positive kernel identifiability result. Our work generalizes SCMs in a way that allows it to work in a world closer to the one we live in.

1 INTRODUCTION

Many biological, social, and engineered systems generate networks whose structure depends on latent contexts, which in turn, shape downstream dynamics on nodes. Structural Causal Models (SCMs) typically treat graphs as fixed/exogenous. However, in many real-world systems, whether one variable causally influences another is not a fixed background fact, but is itself determined by latent processes and can be altered by interventions. A genetic perturbation may abolish a synapse between two neurons; a policy change may sever a regulatory link between two companies. In such settings, we need causal semantics in which the existence of edges is itself an endogenous mechanism, allowing them to be reasoned about, intervened on, and identified, rather than a static scaffold on which mechanisms are defined. Furthermore, causal discovery methods based on SCMs typically assume we can directly observe system variables, potentially based on some knowledge of the network, and infer a single, context-independent mechanism for each variable. While powerful, the assumptions behind classical SCMs often fail to capture the complexity of real-world systems.

Consider neuroscience, where we aim to infer neural circuitry from indirect measurements. For example, calcium imaging provides a noisy view of neuronal activity (partial observability). A neuron's influence (mechanism, e.g. excitatory or inhibitory) and connectivity (structure) thus depend on its type/local environment (context), which are themselves often noisily observed. Crucially, this type is often shaped by the connectivity itself during development, creating an entanglement between context and structure. We thus need to solve a nontrivial inference about context, mechanism, and structure, where each is partially observed. Similarly, the transition from Markov decision processes (MDPs) to partially observed MDPs (POMDPs) fundamentally changes the problem structure, requiring new solution concepts once the agent no longer has direct access to the full state [Kaelbling et al., 1998]. Here, causal modeling faces an analogous gap when structure, context, and mechanisms are only partially observed. Standard SCMs correspond to the "fully observed" regime, assuming a fixed, known graph with directly measurable variables. Thus, we

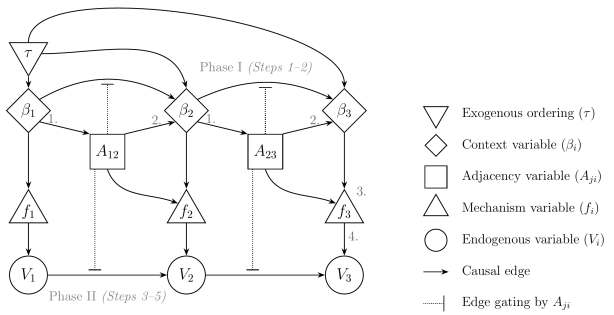

Figure 1: Illustration of the expanded DAG for a 3-node POSCM under ordered generation $\tau = (1, 2, 3)$. Phase I sequentially generates context and adjacency ($\beta_1 \to A_{12} \to \beta_2 \to \cdots$). Phase II assigns mechanisms f_i depending on both local context β_i and realized parents (e.g., $A_{12} \to f_2$), then generates the observed variables V_i. The full expanded graph remains acyclic.

require a partially observed counterpart for describing more real-world causal systems.

Here we introduce *Partially Observed Structural Causal Models (POSCMs)* as a formal framework to model such cases: latent contexts β supervise both (i) the stochastic mechanism that forms directed edges A and (ii) the value mechanisms that generate observed variables V along the realized graph. To our knowledge, no prior causal framework treats structure formation as an endogenous causal mechanism driven by latent contexts while also providing an explicit intervention vocabulary that separates changes to context, structure, and value mechanisms. Moreover, none embeds these capabilities within an ordered-generation semantics that accommodates partial observability of the generating context variables.

Contributions. We (i) formalize POSCMs via ordered generation, (ii) show SCMs arise as a special case, (iii) use a dyadic/message view to define a context/value node–edge intervention hierarchy, (iv) give identifiability barriers and kernel-level positive results under readout/coverage assumptions, and (v) validate predictions in a virtual retina simulator.

For related work, see Appendix A.

2 POSCM FORMALISM

Context affects causal structure (e.g., a cone photoreceptor's type determines whether it synapses onto an ON or OFF bipolar cell) , while causal structure in turn affects the generative model for context (the pattern of synaptic inputs a bipolar cell receives influences its functional identity as ON- or OFF-type). The relevant generative models are thus non-trivially entangled. To causally model such situations, we introduce a formalism based on *ordered generation* [Simon, 1977, Barabási and Albert, 1999], ensuring the overall generative process remains acyclic.

Definition 2.1 (POSCM with Ordered Generation)**.** *A POSCM is a tuple $M = (A, \beta, V, U, \alpha, \phi, f, \Gamma, \tau)$, characterized by an adjacency variable (A), context variables (β), endogenous variables (V), exogenous variables (U), a structure kernel (α), context variable mechanisms (ϕ), endogenous variable mechanisms (f), mechanism operator (Γ), and an exogenous total ordering τ over the nodes $V = \{V_1, \ldots, V_N\}$.*

The POSCM M specifies the data-generating causal process*. What can be* observed *is specified separately by a measurement model (Sec. 2.1), and what can be* manipulated *is specified by an intervention family (Sec. 4).*

The generative process proceeds sequentially according to the order τ. Without loss of generality, assume $\tau(i) = i$.

Phase I: Sequential Generation of Structure and Context

For $i = 1$ to N:

1. ***Structure Formation (Incoming Edges):** For each potential parent $j < i$, the edge A_{ji} is drawn from a structure kernel whose conditional law depends on the (already generated) source context β_j: $A_{ji} = \alpha(\beta_j, U^A_{ji})$. Let $Pa_A(i) = \{j : A_{ji} = 1\}$ be the realized parent set.*
2. ***Context Generation:** The context β_i is generated based on the contexts of its realized parents: $\beta_i = \phi_i(\{\beta_j : j \in Pa_A(i)\}, U^\beta_i)$.*

This interleaved process establishes a valid Directed Acyclic Graph (DAG) over the expanded set of variables (e.g., $\beta_1 \to A_{12} \to \beta_2 \to \ldots$, see Fig. 1), resolving any circularity between A and β and capturing the co-evolution of structure and identity.

Phase II: Mechanisms, Variables, and Measurement

After (A, β) are fully generated:

3. ***Mechanism Assignment:** The mechanism f_i is drawn conditioned on the local context β_i* and *the realized parent set $Pa_A(i)$. The parent set defines the input signature (domain) of the function. $f_i = \Gamma(\beta_i, Pa_A(i), U^f_i)$.*
4. ***Endogenous Variable Generation (SCM):** $V_i = f_i(\{V_j : j \in Pa_A(i)\}, U^V_i)$.*
5. ***Measurement (fixed $\mathcal{O}$):** We observe (possibly noisy) versions of (A, β, V) through an external measurement model $\mathcal{O}$ (Def. 2.2), yielding an observable $\tilde{Z}$ that may include any subset of $(\tilde{A}, \tilde{\beta}, \tilde{V})$.*

2.1 MEASUREMENT MODEL

The causal model M does not, by itself, specify what is observed. Instead, we couple M to a (possibly noisy) mea-

surement model $\mathcal{O}$ that determines the available data.

Definition 2.2 (Measurement model)**.** *A measurement model is a tuple $\mathcal{O} = (\mathcal{O}^A, \mathcal{O}^\beta, \mathcal{O}^V, U^{\mathcal{O}})$ specifying observation channels*

$$\tilde{A} = \mathcal{O}^A(A, U_A^{\mathcal{O}}), \quad \tilde{\beta} = \mathcal{O}^\beta(\beta, U_\beta^{\mathcal{O}}), \quad \tilde{V} = \mathcal{O}^V(V, U_V^{\mathcal{O}}), \tag{1}$$

where $U^{\mathcal{O}} = (U_A^{\mathcal{O}}, U_\beta^{\mathcal{O}}, U_V^{\mathcal{O}})$ denotes measurement noise. The observable $\tilde{Z}$ may include any subset of $(\tilde{A}, \tilde{\beta}, \tilde{V})$ (e.g., adjacency observed but contexts latent).

Convention. Throughout, identifiability statements treat $\mathcal{O}$ as fixed and known (or identifiable within a specified parametric family), and ask which parts of the causal model M are determined from interventional distributions over $\tilde{Z}$. This formalism highlights a fundamental entanglement: the mechanism cannot be defined independently of the structure, since the structure determines the mechanism's input domain.

Remark 2.3 (Reduction to SCMs)**.** *A POSCM reduces to a standard SCM variant (e.g., heterogeneous mechanism SCMs) iff the adjacency matrix variable A and the mechanism variables f are fixed to A^* and f^* respectively, i.e., $A \sim \delta(A - A^*), f \sim \delta(f - f^*)$ where δ is the Dirac delta function.*

Remark 2.4 (Acyclicity and ordered generation)**.** *Ordered generation combines two established ideas: the topological-order forward sampling standard in acyclic SCMs, where each variable is generated as a function of earlier variables and exogenous noise [Simon, 1977], and the sequential node-arrival construction used in graph generative models [Barabási and Albert, 1999]. It extends both by interleaving edge formation with context propagation under causal intervention semantics; the former assumes a fixed graph, while the latter lacks an intervention calculus.*

The ordered-generation semantics assumes an exogenous total order τ over nodes, which rules out instantaneous feedback loops in the one-shot (non-temporal) model. This assumption is shared with DAG-SCMs *[Pearl, 2009] and most DAG-based causal discovery methods [Spirtes et al., 2000, Chickering, 2002, Peters et al., 2017], and it ensures that the induced augmented causal graph over (β, A, V) is acyclic (Fig. 1).*

Importantly, τ does not need to be interpreted as an arbitrary modeling artifact: in settings with an intrinsic temporal or developmental sequence, τ can be anchored to observed time. For example, in our virtual retina experiments (Sec. 6), the laminar circuit provides a biologically grounded ordering (PR→HZ→BC→AC→RGC), which we use as τ at the layer level.

When such an intrinsic order is unavailable, τ should be viewed as part of the POSCM specification; different choices of τ may correspond to different models and thus different counterfactual semantics, analogous to how different SCM mechanism specifications can agree observationally yet disagree counterfactually.

3 MESSAGE-AUGMENTED POSCM

With the help of the Kolmogorov-Arnold-Sprecher (KAS) theorem, we introduce *message-augmented* POSCM representation in which each directed dyad carries an explicit latent message variable. KAS guarantees that such message parameterizations are available for broad classes of mechanisms. This POSCM representation will help us introduce new types of causal interventions down the road.

Definition 3.1 (Message-augmented POSCM)**.** *Fix an ordering τ and, for each node i, let $m_i := |\{j : j < i\}| = i-1$ denote the number of* potential *parents under ordered generation. A POSCM is* message-augmented *if for each node i there exists a finite message dimension d_i (fixed for that node) and functions*

$$H^\beta_{i\leftarrow j} : \mathcal{B} \to \mathbb{R}^{d_i}, \qquad H^V_{i\leftarrow j} : \mathcal{V} \to \mathbb{R}^{d_i},$$

together with aggregation maps

$$\Phi_i : \mathbb{R}^{d_i \times m_i} \to \mathcal{B}, \qquad F_i : \mathbb{R}^{d_i \times m_i} \to \mathcal{V},$$

such that (with the convention $M_{i\leftarrow j} \equiv 0$ when $A_{ji} = 0$)

$$M^\beta_{i\leftarrow j} = A_{ji}\, H^\beta_{i\leftarrow j}(\beta_j), \quad \beta_i = \Phi_i\Big(\{M^\beta_{i\leftarrow j}\}_{j<i},\, U_i^\beta\Big), \tag{2}$$

$$M^V_{i\leftarrow j} = A_{ji}\, H^V_{i\leftarrow j}(V_j), \quad V_i = F_i\big(\{M^V_{i\leftarrow j}\}_{j<i},\, U_i^V\big). \tag{3}$$

Remark 3.2 (Fixed message dimension)**.** *The message codomain d_i is* fixed for node i *and does not vary with the realized parent set size $|\mathrm{pa}(i)|$. This resolves the non-operational "variable-dimension" issue that arises if one defines an $i \leftarrow j$ intervention using a representation whose dimension depends on which other parents happen to be present in a given world.*

KAS edge-functional decomposition. The Kolmogorov-Arnold representation theorem states that every continuous $f : [0,1]^n \to \mathbb{R}$ can be written as a finite sum of univariate "outer" functions of sums of univariate "inner" functions [Kolmogorov, 1957, Arnold, 1957]. Sprecher further showed one may take *a single inner function* (up to shifts in the argument) [Sprecher, 1996].

Assume (after rescaling to $[0,1]$ or restricting to a high-probability compact set; Appendix B) that the mechanism is continuous. Applying KAS to the canonical extension of the mechanism as a function of all $m_i = i-1$ potential parents (ignoring coordinates corresponding to absent edges) yields

a representation with a fixed dimension $d_i = 2m_i + 1$. Concretely, there exist constants $\eta_i \in \mathbb{R}$, weights $\{\lambda_{i\leftarrow j}\}_{j<i}$, and univariate functions Ψ_i, ψ_i such that

$$V_i = \sum_{q=0}^{2m_i} \Psi_i \left(q + \sum_{j<i} A_{ji}\, \lambda_{i\leftarrow j}\, \psi_i(V_j + \eta_i q) + \lambda_U \psi_i(U_i^V + \eta_i q) \right), \tag{4}$$

and analogously for contexts (componentwise if β_i is vector-valued). Equation (4) induces message functions with coordinates $[H^V_{i\leftarrow j}(x)]^q := \lambda_{i\leftarrow j}\, \psi_i(x + \eta_i q)$ for $q = 0, \ldots, 2m_i$ (and similarly for $H^\beta_{i\leftarrow j}$). We stress that KAS is invoked here only to justify existence of a *dyadic message parameterization*; interventions in Sec. 4 are defined directly on the message primitives (H^β, H^V).

KAS is exact for continuous mechanisms on compact domains. We use it only as an *existence theorem* for dyadic message parameterizations, with extensions beyond this regime summarized in the following lemma:

Lemma 3.3 (An $L^p(\mu)$ density statement for KAS edge–functional decompositions)**.** *Let $X \sim \mu$ be an $\mathbb{R}^n$-valued random vector and let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function with $\mathbb{E}[|f(X)|^p] < \infty$ for some $1 \le p < \infty$. For any $\varepsilon > 0$, there exists a measurable function $g : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\mathbb{E}[\,|f(X) - g(X)|^p\,] < \varepsilon,$$

and such that g is representable in KAS edge-functional decomposition form (on a sufficiently large compact cube), in the following sense: there exists $R < \infty$ for which $g|_{[-R,R]^n}$ admits a KAS representation after affine rescaling of $[-R, R]^n$ to $[0, 1]^n$.

For proof of the above lemma, see Appendix B. All proofs in the Appendix have been formally verified in Lean 4 [Moura and Ullrich, 2021].

Remark 3.4 (Representational non-uniqueness as a gauge)**.** *KAS representations are generally non-unique: different choices of $(\Psi_i, \psi_i, \eta_i, \lambda)$ can induce the same overall mechanism. Accordingly, identifiability of* message coordinates *should be understood either (i) relative to a fixed message parameterization (Def. 3.1), or (ii) "up to" the induced representation gauge. A convenient way to formalize this is to view the gauge as the class of internal message reparameterizations that preserve the induced parent-to-child mechanism; for example, any bijection of the message space that is compensated inside F_i (and similarly inside Φ_i) yields an observationally equivalent representation.*

4 INTERVENTION HIERARCHY

POSCMs admit interventions that act on context/value (i) nodes and (ii) edges. To make edge interventions operational, we define them on the *message primitives* of the message-augmented representation (Def. 3.1).

4.1 PRIMITIVE INTERVENTION TYPES

Definition 4.1 (Node interventions)**.** *A β-node intervention $do(\beta_j = \widetilde{b})$ sets node j's context exogenously and then runs the remaining generative process forward (affecting both structure formation downstream of j and any mechanisms supervised by β_j). A V-*node intervention *$do(V_j = \widetilde{v})$ sets node j's value exogenously and runs the downstream SCM forward on the realized graph.*

Definition 4.2 (Edge-message interventions)**.** *In a message-augmented POSCM (Def. 3.1), a β-*edge intervention *on dyad $s \to t$ replaces the context message function on that channel,*

$$do_\beta(s \to t;\, \widetilde{H}^\beta_{t\leftarrow s}) : \quad H^\beta_{t\leftarrow s} \leftarrow \widetilde{H}^\beta_{t\leftarrow s},$$

*leaving all other components of the POSCM unchanged. Similarly, a V-*edge intervention *on dyad $j \to i$ replaces the value message function,*

$$do_V(j \to i;\, \widetilde{H}^V_{i\leftarrow j}) : \quad H^V_{i\leftarrow j} \leftarrow \widetilde{H}^V_{i\leftarrow j}.$$

If $A_{st} = 0$ (resp. $A_{ji} = 0$) in a realized world, then the corresponding message is identically zero by convention (Def. 3.1), so a message intervention is a no-op.

Relation to edge/path/soft interventions in SCMs. In standard SCMs, "edge" or "path" interventions can be viewed as modifying a single parent contribution while leaving the rest of the mechanism fixed Shpitser and Tchetgen [2014]. Message-augmented POSCMs make this idea explicit: an edge-message intervention replaces a single dyadic channel $H_{i\leftarrow j}$ while leaving all other mechanisms unchanged (Def. 4.2). This instantiates the usual notion of mechanism-level (soft) interventions, but in a setting where structure itself is stochastic and intervention-sensitive.

Toy example (distributive law): why edge interventions form a strictly stronger tier (see Fig. 2). Consider two deterministic causal models with inputs (x, y, z) and a single observed output W. Model $\mathcal{M}_{\mathrm{LHS}}$ computes $W = x{\cdot}(y{+}z)$ via an internal addition node and a single multiplication node, while model $\mathcal{M}_{\mathrm{RHS}}$ computes $W = (x \cdot y) + (x \cdot z)$ via two internal multiplications. Suppose the intermediate arithmetic nodes are unobserved and cannot be intervened upon, so the observation model reveals only W. Then for every node intervention $\mathrm{do}(x = a, y = b, z = c)$ the two models agree on the observed outcome W, hence they are $(\mathcal{I}_{\text{node}}, \mathcal{O})$-equivalent for this restricted interface: node interventions identify the input-output kernel, but not the internal decomposition. In contrast, an edge intervention that perturbs only *one* channel of x distinguishes them. For instance, replace only the input carried on the edge $x \to *_1$ by a value x' while leaving the other occurrence of x unchanged; then $\mathcal{M}_{\mathrm{LHS}}$ yields $W = x'(y{+}z)$ whereas $\mathcal{M}_{\mathrm{RHS}}$ yields $W = x'y + xz$, which differ for generic assignments

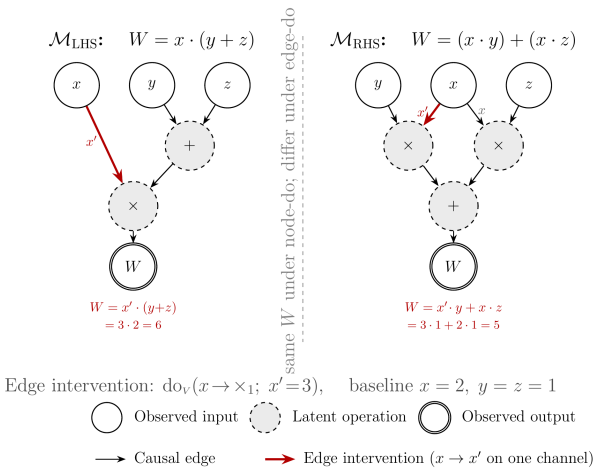

Figure 2: Distributive law toy example.

(e.g., $x = 2, x' = 3, y = z = 1$ gives 6 vs. 5). This illustrates why edge interventions are a genuinely stronger primitive: they isolate one causal channel without globally changing the source node. In POSCMs, the analogous separation underlies the gap between kernel identification (Theorem 5.4) and identifiability of dyadic message primitives or edge-intervention counterfactuals under partial observability (Lemma 5.3, Theorem C.3).

4.2 STRUCTURAL ENDOGENEITY

Definition 4.3 (Per-source supervising measure)**.** *Fix a regime (baseline* 0 *or intervention* int*). For a source* j*, define its* first-generation supervising measure *as the conditional law of its outgoing edge* vector

$$\mu_j \;:=\; \mathcal{L}\big(A_{j,\,j+1:N-1} \,\big|\, \mathcal{H}_j\big), \tag{5}$$

where $\mathcal{H}_j$ *collects all variables that "supervise" edge generation from* j *(at minimum* β_j*).*

The ordered generative process means the causal graph A is not fixed, but co-evolves with the context β via $\mathbb{P}(A|\beta)$.

Definition 4.4 (Intervention-Induced Structural Change; IISC)**.** *IISC occurs if an intervention alters the supervising measure of the structure* A*.*

μ_j factorizes when outgoing edges are conditionally independent Bernoullis given $\mathcal{H}_j$. Although, *dependence across outgoing edges* yields non-product μ_j, our IISC definitions apply to both.

Remark 4.5. *Extending POSCMs beyond ordered generation is an important direction. One option is* time-unrolling *(dynamic POSCMs) in which feedback is represented via lagged variables, yielding an acyclic graph on a finite horizon. Another option is* cyclic/equilibrium *semantics via fixed-point SCMs or* σ*-separation [Bongers et al., 2021, Forré and Mooij, 2018], which would require revisiting the intervention semantics and IISC diagnostics in a cyclic setting.*

5 IDENTIFIABILITY THEORY

POSCMs introduce multiple interacting mechanisms, so identifiability must be indexed by the available intervention primitives. We formalize this and establish both negative and positive results.

Definition 5.1 ($(\mathcal{I}, \mathcal{O})$-equivalence and identifiability)**.** *Fix a measurement model* $\mathcal{O}$ *(Def. 2.2) and ordering* τ*. An* intervention family $\mathcal{I}$ *is a collection of interventions from the primitive types in Sec. 4 (node, edge-message), possibly restricted to a class (e.g.,* $\{do(\beta_j = b) : b \in \mathcal{B}_j\}$*). Two POSCMs* M, M' *are* $(\mathcal{I}, \mathcal{O})$-equivalent *if, for every intervention* $\iota \in \mathcal{I}$*, they induce identical distributions over the* same *observables* $\tilde{Z}^\iota := \mathcal{O}(A^\iota, \beta^\iota, V^\iota)$ *generated by the fixed measurement model* $\mathcal{O}$*. A parameter* θ *is* $\mathcal{I}$-identifiable (under $\mathcal{O}$) *if* $(\mathcal{I}, \mathcal{O})$*-equivalence implies* $\theta(M) = \theta(M')$*, up to stated equivalence transformations.*

No "observer swapping." Equivalence is defined relative to a fixed $\mathcal{O}$: we do *not* allow changing the observation process to restore equivalence between mismatched causal models. The following equivalence transformations are unavoidable when contexts are latent and interventions are not anchored to an intrinsic coordinate system: (i) *context reparameterization* (diffeomorphisms $\gamma : \mathcal{B} \to \mathcal{B}$ with appropriately transformed kernels and context mechanisms), and (ii) *discrete label permutation* when β_i takes values in a finite set. In addition, if message coordinates are obtained via a non-unique functional representation (e.g., KAS), then identification of *message coordinates* is understood either relative to a fixed representation or "up to" the induced representation gauge (Remark 3.4).

Proposition 5.2 (Non-identifiability without β-interventions)**.** *Let* $\mathcal{I} = \mathcal{I}_{\text{obs}} \cup \mathcal{I}_{V\text{-node}} \cup \mathcal{I}_{V\text{-edge}}$ *(no* β*-level interventions). If* β *is unobserved, then:*

- *(i) The structure kernel* $\mathbb{P}(A \mid \beta)$ *(equivalently, the supervising measures* $\mu_j(\beta_j) := \mathcal{L}(A_{j,>j} \mid \beta_j)$*) is not* $\mathcal{I}$*-identifiable beyond context reparameterization.*
- *(ii) The context propagation mechanisms* $\{\phi_i\}$ *(equivalently, the conditional laws* $\mathcal{L}(\beta_i \mid \beta_{Pa_A(i)}, Pa_A(i))$*) are not* $\mathcal{I}$*-identifiable beyond context reparameterization.*
- *(iii) The mechanism-assignment operator* Γ *(equivalently, the conditional laws* $\mathbb{P}(f_i \in \mathcal{F} \mid \beta_i, Pa_A(i))$*, and hence the induced value kernels* $\mathcal{L}(V_i \mid V_{Pa_A(i)}, \beta_i, Pa_A(i))$*) is not* $\mathcal{I}$*-identifiable beyond context reparameterization.*

Full proof in Appendix D.1.

Proposition 5.2 establishes that V-level data alone cannot disentangle the structure kernel, context propagation, or mechanism assignment when contexts are latent. A second barrier arises when edges are also latent.

Lemma 5.3 (Structure-mechanism confounding under latent edges)**.** *When edges are latent (not observed via $\mathcal{O}$), node-level interventions $\mathrm{do}(V_j = \tilde{v})$ are insufficient to determine the distribution resulting from an V-edge intervention, even with degenerate (non-random) contexts.*

Full proof in Appendix D.2.

Together, Proposition 5.2 and Lemma 5.3 identify two distinct barriers to identification: (1) latent contexts create reparameterization symmetries, and (2) latent edges create structure-mechanism confounding. Positive identification requires addressing both.

Assumptions for the positive result. We assume access to (i) probing-based structure readout for latent adjacency, (ii) context and value readout, (iii) joint β-node and V-node intervention coverage, and (iv) positivity. Formal statements are in Appendix D.

Theorem 5.4 (Kernel identifiability)**.** *Let $\mathcal{I}_{\mathrm{kern}} := \mathcal{I}_{\mathrm{obs}} \cup \mathcal{I}_{\beta\text{-node}} \cup \mathcal{I}_{V\text{-node}}$ and fix the measurement model $\mathcal{O}$ and ordering τ (Def. 5.1). Assume formal assumptions in Appendix D.1–D.7. Then, on the intervention-reachable support, the POSCM kernels α, $\{\phi_i\}$, and Γ are $(\mathcal{I}_{\mathrm{kern}}, \mathcal{O})$-identifiable in the following kernel sense:*

(i) ***Structure kernel** α. For each node j, the supervising measure*

$$\mu_j(b) \;:=\; \mathcal{L}\big(A_{j,>j} \mid do(\beta_j = b)\big)$$

is identifiable. In particular, each dyadwise marginal $b \mapsto \mathbb{P}(A_{ji} = 1 \mid do(\beta_j = b))$ for $i > j$ is identifiable.

(ii) ***Context propagation kernels induced by** $\{\phi_i\}$. For each node i and each realized parent set S in the support of $Pa_A(i)$ under $\mathcal{I}_{\mathrm{kern}}$, the conditional context kernel*

$$K^{\beta}_{i,S}(\cdot \mid b_S) \;:=\; \mathcal{L}\big(\beta_i \mid do(\beta_S = b_S),\ Pa_A(i) = S\big)$$

is identifiable as a function of b_S on the intervention-reachable support. Equivalently, the conditional law $\mathcal{L}(\beta_i \mid \beta_{Pa_A(i)}, Pa_A(i))$ is identifiable (up to the equivalences discussed above when interventions are not anchored).

(iii) ***Endogenous-variable mechanism kernels induced by** Γ. For each node i and each realized parent set S in the support of $Pa_A(i)$ under $\mathcal{I}_{\mathrm{kern}}$, the conditional value kernel*

$$K^{V}_{i,S}(\cdot \mid v_S, b_i) \;:=\; \mathcal{L}\big(V_i \mid do(V_S = v_S),\ \beta_i = b_i,\ Pa_A(i) = S\big)$$

is identifiable as a function of (v_S, b_i) on the intervention-reachable support. Equivalently, the conditional law $\mathcal{L}(V_i \mid V_{Pa_A(i)}, \beta_i, Pa_A(i))$ is identifiable.

Identification is purely distributional and uses repeated sampling under interventions; adjacency is accessed through the probing-based readout in Assumption D.1.

Full proof in Appendix D.3

Additional results. Appendix C contains further discussion of intervention minimality and dyadic message identifiability (including Theorem C.3 and Proposition C.4).

6 EXPERIMENTS

We validate our identifiability results (Proposition 5.2, Lemma 5.3, Theorem 5.4) in a biophysically detailed virtual human retina simulator (NEURON; ModelDB #2018247; [Ly et al., 2025]). The retina is a natural POSCM: (i) latent cell type co-determines synaptic connectivity (structure kernel) and synaptic transfer functions (mechanism operator), and (ii) the laminar circuit supports a biologically grounded ordered-generation semantics.

Crucially, real retinal experiments exhibit precisely the partial observability that motivates POSCMs. Electrophysiological recordings (e.g., patch clamp or extracellular arrays) reveal neural activity but not cell type [Meister et al., 1994]; conversely, anatomical reconstructions such as serial-section electron microscopy recover wiring and morphological identity but not activity [Briggman et al., 2011]. Most standard preparations offer no perturbation capability [Meister et al., 1994]. When perturbations are available, they span multiple tiers of our intervention hierarchy: axon transection or laser ablation severs individual connections (V-edge interventions) [Nemitz et al., 2019], genetic inactivation under cell-type-specific promoters silences entire cell classes (β-node interventions) [Montgomery et al., 2010], and activity-dependent manipulations during development can rewire connectivity itself (structure-kernel interventions) [Shen et al., 2020]. A virtual retina simulator lets us access all of these intervention types with full ground-truth knowledge, enabling controlled tests of each identifiability prediction.

Signal flow is predominantly feedforward through stereotyped layers: photoreceptors (PR) $\rightarrow$ horizontal cells (HZ) $\rightarrow$ bipolar cells (BC) $\rightarrow$ amacrine cells (AC) $\rightarrow$ retinal ganglion cells (RGC). We use the laminar/developmental order as τ at the layer level; local lateral/feedback couplings (e.g., HZ$\rightarrow$PR, AC$\rightarrow$BC) are treated as within-layer noise consistent with ordered generation at this granularity.

We interpret discrete cell type as β_i (rod vs. cone; ON/OFF bipolar; ON/OFF RGC, etc.). Type supervises both edge formation (*who connects*) and synaptic physiology (*how signals transmit*), matching the POSCM entanglement $\beta \rightarrow$

A and $\beta \to f$.

A small retinal patch contains a functionally complete microcircuit; we simulate independent random patches ("seeds") as independent POSCM instances.

Chemical synapses implement graded release often modeled with an approximately $\tanh$ activation:

$$s_\infty = \tanh\left(\frac{V_{\text{pre}} - V_{\text{thr}}}{V_{\text{slope}}}\right), \qquad I_{\text{syn}} = g_{\max}\, s\,(V_{\text{post}} - e_{\text{rev}}), \tag{6}$$

with first-order binding kinetics $\dot{s} = (s_\infty - s)/[\tau(1 - s_\infty)]$. Typical parameters are $V_{\text{thr}} \approx -45$ mV, $V_{\text{slope}} = 10$ mV, $\tau = 10$ ms, and $g_{\max} = 0.00256$ μmho.

6.1 EXPERIMENT 1: NON-IDENTIFIABILITY WITHOUT β-INTERVENTIONS

Proposition 5.2 predicts that if cell types β are unobserved and no β-level interventions are available, then $(\alpha, \{\phi_i\}, \Gamma)$ are not identifiable beyond context reparameterization. We instantiate this by constructing a *type-swapped* twin retina model that is observationally indistinguishable even under V-node interventions.

Setup (twin models). Model M uses standard parameters. ON bipolar cells (ON-BC) receive sign-inverting mGluR synapses with conductance $g_{\text{ON}} = +0.00256$ μmho and threshold $V_{\text{thr}}^{\text{ON}} = -40$ mV; OFF-BCs receive sign-preserving iGluR synapses with $g_{\text{OFF}} = -0.00256$ μmho and $V_{\text{thr}}^{\text{OFF}} = -42$ mV. Model M' swaps ON-BC and OFF-BC labels (and swaps the corresponding synaptic parameters). In both models, β is *not recorded*.

Protocol. For each of $B = 2$ seeds, we run 10 worlds: observational (M and M'), plus V-node interventions that voltage-clamp a single photoreceptor ($\text{PR}_{\text{upper}}[0]$) at $v \in \{-60, -50, -40, -30\}$ mV in each model (NEURON `SEClamp`, series resistance $0.001\ \Omega$).

Analysis. From each cell trace we extract a firing rate via threshold crossings at -20 mV and compare the across-cell firing-rate distributions between M and M' with a two-sample Kolmogorov–Smirnov test:

$$H_0 : F_M = F_{M'}, \qquad D = \sup_r |F_M(r) - F_{M'}(r)|\,.$$

Results. See Table 1.

Interpretation. Swapping ON/OFF type labels together with the corresponding synaptic rules preserves the induced V-distributions when types are latent, empirically validating the reparameterization symmetry underlying Proposition 5.2. In this setting, V-only observations and V-node interventions cannot break the type symmetry; β-level access is required.

Condition	KS D	p-value
Observational	0.023	0.999
$\text{do}(V_{\text{PR}} = -60\text{ mV})$	–	0.964
$\text{do}(V_{\text{PR}} = -50\text{ mV})$	–	0.980
$\text{do}(V_{\text{PR}} = -40\text{ mV})$	–	0.964
$\text{do}(V_{\text{PR}} = -30\text{ mV})$	–	0.964

Table 1: Experiment 1: KS test comparing firing-rate distributions between M and the type-swapped twin M'. All p-values are $\gg 0.05$, consistent with Proposition 5.2.

6.2 EXPERIMENT 2: STRUCTURE–MECHANISM CONFOUNDING UNDER LATENT EDGES

Lemma 5.3 shows that when edges are latent, V-node interventions do not suffice to predict V-edge intervention effects. We construct two retina models with different PR→BC connectivity densities whose V-responses match under node-level clamps but diverge under edge-level conductance perturbations.

Setup (calibrated confounding pair). Model M uses the natural PR→BC synapse density p and baseline conductance $g = 0.00256$ μmho. Model M' randomly blocks 40% of PR→BC synapses (sets $g_{\max} = 0$), yielding $p' = 0.6p$, and scales the remaining conductances to $g' = g/0.6 = 0.00427$ μmho so that $p\,g = p'\,g'$ (matching mean synaptic drive in the linear regime).

Protocol. For each of $B = 2$ seeds we run (i) *node-do* conditions that clamp a single PR at $v \in \{-60, -50, -40, -30\}$ mV, and (ii) *edge-do* conditions that replace the conductance on all active PR→BC synapses with a test value $g_{\text{test}} \in \{0.001, 0.002, 0.004, 0.008\}$ μmho (preserving blocked synapses in M').

Analysis. For each cell we compute a steady-state effect $\Delta V_i = \bar{V}_i^{(\text{int})} - \bar{V}_i^{(\text{obs})}$, where $\bar{V}_i$ is the mean potential over the last 50% of the trace. We focus on the postsynaptic BC population ($\text{BIP}_{\text{upper}}$) and compare effect distributions between M and M' using maximum mean discrepancy (MMD) with an RBF kernel:

$$\text{MMD}^2(X, Y) = \mathbb{E}[k(X, X')] - 2\,\mathbb{E}[k(X, Y)] + \mathbb{E}[k(Y, Y')],$$
$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right).$$

Results. See Table 2.

Interpretation. Under node-level clamps, the calibrated identity $p\,g = p'\,g'$ makes M and M' hard to distinguish (small MMD), illustrating structure–mechanism confounding when edges are latent. Under edge-level interventions, setting all active synapses to a common g_{test} exposes the differing synapse counts, producing substantially larger discrepancies, consistent with Lemma 5.3.

Intervention	Condition	MMD
Node-do	$v = -60\,\mathrm{mV}$	0.0440
	$v = -50\,\mathrm{mV}$	0.0435
	$v = -40\,\mathrm{mV}$	0.0465
	$v = -30\,\mathrm{mV}$	0.0437
	Mean	**0.0444**
Edge-do	$g_{\text{test}} = 0.001$	0.1135
	$g_{\text{test}} = 0.002$	0.0950
	$g_{\text{test}} = 0.004$	0.1853
	$g_{\text{test}} = 0.008$	0.1309
	Mean	**0.1312**

Table 2: Experiment 2: node-do interventions yield similar effect distributions (low MMD) for the calibrated pair (M, M'), while edge-do interventions reveal the structural difference (higher MMD).

Eccentricity (mm)	Total cells
-1.2 (baseline)	289
-1.5	306
-2.0	330
-2.5	347
-3.0	350
-3.5	346

Table 3: Experiment 3 (Phase I): sweeping eccentricity changes the network composition, providing context-level variation.

6.3 EXPERIMENT 3: KERNEL IDENTIFIABILITY VIA β-NODE AND V-NODE INTERVENTIONS

Theorem 5.4 states that with β-node and V-node interventions (plus readout/coverage assumptions) the POSCM kernels become identifiable. We demonstrate two corresponding identification "routes" in the retina: a context sweep (proxy β-node intervention) that changes cell-type composition, and a voltage sweep (V-node interventions) that recovers a synaptic transfer curve.

Phase I (context sweep). We vary the eccentricity parameter (a global context proxy) over $\{-1.2, -1.5, -2.0, -2.5, -3.0, -3.5\}$ mm before network construction, producing different type compositions and densities.

Phase II (voltage sweep). We simultaneously clamp all bipolar cells ($\text{BIP}_{\text{upper}}$) to $v \in \{-70, -60, -50, -40, -30, -20\}\,\mathrm{mV}$ (bulk `SEClamp`) and measure the induced effect on ganglion cells.

Results: context-dependent composition. See Table 3.

BC clamp v (mV)	Mean $\Delta\bar{V}_{\mathrm{RGC}}$ (mV)	Std (mV)
-70	-0.284	0.224
-60	-0.287	0.212
-50	-0.296	0.199
-40	-0.095	0.049
-30	$+0.412$	0.034
-20	$+0.460$	0.031

Table 4: Experiment 3 (Phase II): voltage sweep recovers a sigmoidal BC→RGC transfer curve consistent with the synaptic threshold $V_{\mathrm{thr}} \approx -45\,\mathrm{mV}$ in Eq. (6).

Results: BC→RGC transfer curve. We compute the mean ganglion-cell steady-state effect $\Delta\bar{V}_{\mathrm{RGC}} = \bar{V}_{\mathrm{RGC}}^{\mathrm{do}(V_{\mathrm{BC}}=v)} - \bar{V}_{\mathrm{RGC}}^{(\mathrm{obs})}$. See Table 4.

Interpretation. The eccentricity sweep demonstrates that changing context changes population composition (a proxy for identifying context-dependent kernels), while the voltage sweep traces a sigmoidal input–output relation across the BC→RGC pathway with a transition near $-45\,\mathrm{mV}$, consistent with the ground-truth $\tanh$ synapse. Together these experiments illustrate the sufficiency of β-node and V-node interventions for kernel-level identification in Theorem 5.4.

7 DISCUSSION.

POSCMs extend SCMs to a world where interaction structure is itself generated by latent context and can be intervened on, while keeping an acyclic semantics via ordered generation. This yields a practical intervention vocabulary, spanning node- and edge-level operations on both context and values, and a message-augmented parameterization that supports genuinely "surgical" edge manipulations. A key contribution is that the identifiability results do not just provide positive guarantees: they also diagnose failure modes (latent-context symmetries and latent-edge structure–mechanism confounding) and thereby guide experiment design toward interfaces that can actually disentangle these factors. The retina experiments then serve as a concrete proof-of-concept that these predictions appear in a realistic biophysical simulator, and that kernel-level recovery is achievable under the proposed readout/coverage assumptions. Going forward, it will be valuable to weaken the required access (e.g., partial/noisy readouts, limited interventions), extend the semantics to cyclic or time-unrolled settings, and better characterize which questions require dyadic message identification versus those adequately answered by identifiable kernel targets.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – Project number 390727645. C.M.W is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (C4: 101164709), the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) under Germany's Excellence Strategy (EXC 3066/1 "The Adaptive Mind", Project No. 533717223), and the Excellence Cluster "Reasonable AI" by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) under Germany's Excellence Strategy – EXC-3057. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Turan Orujlu. The authors also thank Ilya Shpitser, Eric J Tchetgen Tchetgen, and Sacha Sokoloski for insightful discussions.

References

V I Arnold. On functions of three variables. *Dokl. Akad. Nauk SSSR*, 114:679–681, 1957.

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999. doi: 10.1126/science.286.5439.509. URL https://www.science.org/doi/abs/10.1126/science.286.5439.509.

Aida Bareghamyan, Changfeng Deng, Sarah Daoudi, Shubhash C Yadav, Xiaocen Lu, Wei Zhang, Robert E Campbell, Richard H Kramer, David M Chenoweth, and Don B Arnold. A toolbox for ablating excitatory and inhibitory synapses. *eLife*, 13:RP103757, apr 2025. ISSN 2050-084X. doi: 10.7554/eLife.103757. URL https://doi.org/10.7554/eLife.103757.

Elias Bareinboim and Judea Pearl. Meta-transportability of causal effects: A formal approach. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2013, Scottsdale, AZ, USA, April 29 - May 1, 2013*, volume 31 of *JMLR Workshop and Conference Proceedings*, pages 135–143. JMLR.org, 2013. URL http://proceedings.mlr.press/v31/bareinboim13a.html.

Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M. Mooij. Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5): pp. 2885–2915, 2021. ISSN 00905364, 21688966. URL https://www.jstor.org/stable/27170673.

Craig Boutilier, Nir Friedman, Moisés Goldszmidt, and Daphne Koller. Context-specific independence in bayesian networks. In Eric Horvitz and Finn Verner Jensen, editors, *UAI '96: Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence, Reed College, Portland, Oregon, USA, August 1-4, 1996*, pages 115–123. Morgan Kaufmann, 1996. URL https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=359&proceeding_id=12.

Kevin L. Briggman, Moritz Helmstaedter, and Winfried Denk. Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, 471(7337):183–188, March 2011. ISSN 1476-4687. doi: 10.1038/nature09818. URL http://dx.doi.org/10.1038/nature09818.

Peter Bühlmann, Jonas Peters, and Jan Ernest. CAM: causal additive models, high-dimensional order search and penalized regression. *CoRR*, abs/1310.1533, 2013. URL http://arxiv.org/abs/1310.1533.

Raymond J. Carroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. *Measurement Error in Nonlinear Models*. Chapman and Hall/CRC, June 2006. ISBN 9781420010138. doi: 10.1201/9781420010138. URL https://www.taylorfrancis.com/books/mono/10.1201/9781420010138/measurement-error-nonlinear-models-ciprian-cra

David Maxwell Chickering. Optimal structure identification with greedy search. *J. Mach. Learn. Res.*, 3:507–554, 2002. URL https://jmlr.org/papers/v3/chickering02b.html.

Rodrigo A Collazo, Christiane Goergen, and Jim Q Smith. *Chain event graphs*. CRC Press, London, England, June 2020.

Juan D. Correa and Elias Bareinboim. A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 10093–10100. AAAI Press, 2020. doi: 10.1609/AAAI.V34I06.6567. URL https://doi.org/10.1609/aaai.v34i06.6567.

Alexander Dawid. Decision-theoretic foundations for statistical causality. *Journal of Causal Inference*, 9:39–77, 05 2021. doi: 10.1515/jci-2020-0008.

Ricardo Bezerra de Andrade e Silva, Richard Scheines, Clark Glymour, and Peter Spirtes. Learning the structure of linear latent variable models. *J. Mach. Learn. Res.*, 7:191–246, 2006. URL https://jmlr.org/papers/v7/silva06a.html.

Eliana Duarte and Liam Solus. Representation of context-specific causal models with observational and interventional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkaf059, 10 2025. ISSN 1369-7412. doi: 10.1093/jrsssb/qkaf059. URL https://doi.org/10.1093/jrsssb/qkaf059.

Patrick Forré and Joris M. Mooij. Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders. In Amir Globerson and Ricardo Silva, editors, *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018*, pages 269–278. AUAI Press, 2018. URL http://auai.org/uai2018/proceedings/papers/117.pdf.

Amin Jaber, Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Causal discovery from soft interventions with unknown targets: Characterization and learning. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6cd9313ed34ef58bad3fdd504355e72c-Abstract.html.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1):99–134, 1998. ISSN 0004-3702. doi: https://doi.org/10.1016/S0004-3702(98)00023-X. URL https://www.sciencedirect.com/science/article/pii/S000437029800023X.

A N Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk SSSR*, 114(5):953–956, 1957.

Henri Lebesgue. Sur l'intégration des fonctions discontinues. *Annales scientifiques de l'École Normale Supérieure*, 27:361–450, 1910.

Jaron Jia Rong Lee, AmirEmad Ghassami, and Ilya Shpitser. A general identification algorithm for data fusion problems under systematic selection. In Negar Kiyavash and Joris M. Mooij, editors, *Uncertainty in Artificial Intelligence, 15-19 July 2024, Universitat Pompeu Fabra, Barcelona, Spain*, volume 244 of *Proceedings of Machine Learning Research*, pages 2188–2204. PMLR, 2024. URL https://proceedings.mlr.press/v244/lee24b.html.

Nikolai Lusin. Sur les propriétés des fonctions mesurables. *Comptes Rendus de l'Académie des Sciences de Paris*, 154:1688–1690, 1912.

Keith Ly, Michael L. Italiano, Mohit N. Shivdasani, David Tsai, Jia-Yi Zhang, Chunhui Jiang, Nigel H. Lovell, Socrates Dokos, and Tianruo Guo. Virtual human retina: Simulating neural signalling, degeneration, and responses to electrical stimulation. *Brain Stimulation*, 18(1):144–163, 2025. ISSN 1935-861X. doi: https://doi.org/10.1016/j.brs.2025.01.013. URL https://www.sciencedirect.com/science/article/pii/S1935861X25000154.

Marc Maier, Katerina Marazopoulou, David Arbour, and David Jensen. A sound and complete algorithm for learning causal models from relational data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, UAI'13, page 371–380, Arlington, Virginia, USA, 2013. AUAI Press.

Markus Meister, Jerome Pine, and Denis A. Baylor. Multi-neuronal signals from the retina: acquisition and analysis. *Journal of Neuroscience Methods*, 51(1):95–106, 1994. ISSN 0165-0270. doi: https://doi.org/10.1016/0165-0270(94)90030-2. URL https://www.sciencedirect.com/science/article/pii/0165027094900302.

Hermann Minkowski. *Geometrie der Zahlen*. Teubner, Leipzig, 1910.

Jacob E. Montgomery, Michael J. Parsons, and David R. Hyde. A novel model of retinal ablation demonstrates that the extent of rod cell death regulates the origin of the regenerated zebrafish rod photoreceptors. *Journal of Comparative Neurology*, 518(6):800–814, January 2010. ISSN 1096-9861. doi: 10.1002/cne.22243. URL http://dx.doi.org/10.1002/cne.22243.

Joris M. Mooij, Sara Magliacane, and Tom Claassen. Joint causal inference from multiple contexts. *J. Mach. Learn. Res.*, 21:99:1–99:108, 2020. URL https://jmlr.org/papers/v21/17-123.html.

Leonardo de Moura and Sebastian Ullrich. The lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28: 28th International Conference on Automated Deduction, Virtual Event, July 12–15, 2021, Proceedings*, page 625–635, Berlin, Heidelberg, 2021. Springer-Verlag. ISBN 978-3-030-79875-8. doi: 10.1007/978-3-030-79876-5_37. URL https://doi.org/10.1007/978-3-030-79876-5_37.

Lena Nemitz, Karin Dedek, and Ulrike Janssen-Bienhold. Rod bipolar cells require horizontal cells for invagination into the terminals of rod photoreceptors. *Frontiers in Cellular Neuroscience*, Volume 13 - 2019, 2019.

ISSN 1662-5102. doi: 10.3389/fncel.2019.00423. URL https://www.frontiersin.org/journals/cellular-neuroscience/articles/10.3389/fncel.2019.00423.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition, 2009.

Judea Pearl and Elias Bareinboim. External validity: From *Do*-calculus to transportability across populations. In Hector Geffner, Rina Dechter, and Joseph Y. Halpern, editors, *Probabilistic and Causal Inference: The Works of Judea Pearl*, volume 36 of *ACM Books*, pages 451–482. ACM, 2022. doi: 10.1145/3501714.3501741. URL https://doi.org/10.1145/3501714.3501741.

Ronan Perry, Julius von Kügelgen, and Bernhard Schölkopf. Causal discovery in heterogeneous environments under the sparse mechanism shift hypothesis. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/46a126492ea6fb87410e55a58df2e189-Abstract-Conference.html.

Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. Causal inference by using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(5):947–1012, 10 2016. ISSN 1369-7412. doi: 10.1111/rssb.12167. URL https://doi.org/10.1111/rssb.12167.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2017. ISBN 978-0-262-03731-0. URL https://mitpress.mit.edu/books/elements-causal-inference.

Eva Riccomagno and Jim Q. Smith. The causal manipulation of chain event graphs, 2007. URL https://arxiv.org/abs/0709.3380.

Susanne M. Schennach. Recent advances in the measurement error literature. *Annual Review of Economics*, 8:341–377, 2016. ISSN 19411383, 19411391. URL https://www.jstor.org/stable/26774373.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proc. IEEE*, 109(5):612–634, 2021. doi: 10.1109/JPROC.2021.3058954. URL https://doi.org/10.1109/JPROC.2021.3058954.

Ning Shen, Bing Wang, Florentina Soto, and Daniel Kerschensteiner. Homeostatic plasticity shapes the retinal response to photoreceptor degeneration. *Current Biology*, 30(10):1916–1926.e3, May 2020. ISSN 0960-9822. doi: 10.1016/j.cub.2020.03.033. URL http://dx.doi.org/10.1016/j.cub.2020.03.033.

Ilya Shpitser and Eric Tchetgen. Causal inference with a graphical hierarchy of interventions. *The Annals of Statistics*, 44, 11 2014. doi: 10.1214/15-AOS1411.

Herbert A. Simon. *Causal Ordering and Identifiability*, pages 53–80. Springer Netherlands, Dordrecht, 1977. ISBN 978-94-010-9521-1. doi: 10.1007/978-94-010-9521-1_5. URL https://doi.org/10.1007/978-94-010-9521-1_5.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search, Second Edition*. Adaptive computation and machine learning. MIT Press, 2000. ISBN 978-0-262-19440-2.

David A. Sprecher. A numerical implementation of kolmogorov's superpositions. *Neural Networks*, 9(5):765–772, 1996. doi: 10.1016/0893-6080(95)00081-X. URL https://doi.org/10.1016/0893-6080(95)00081-X.

Heinrich Tietze. Über funktionen, die auf einer abgeschlossenen menge stetig sind. *Mathematische Annalen*, 77(3):301–303, 1915.

Partially Observed Structural Causal Models (Supplementary Material)

Turan Orujlu[1,2] **Jordan Matelsky**[3] **Martin V. Butz**[1] **Charley M. Wu**[4,5] **Konrad P. Kording**[3]

[1]University of Tübingen
[2]MPI for Biological Cybernetics
[3]University of Pennsylvania
[4]TU Darmstadt
[5]Hessian.AI

A RELATED WORK

Causal discovery over fixed graphs. Classic causal discovery methods, constraint-based (e.g., Peter-Clark, fast causal inference) and score-based (e.g., greedy equivalence search), typically aim to recover a *fixed* causal graph from samples, with extensions handling latent confounding, interventions, and partial observability [Spirtes et al., 2000, Chickering, 2002, de Andrade e Silva et al., 2006]. These methods assume the causal graph is a fixed object to be recovered. This precludes modeling settings where the interaction structure is itself a stochastic outcome of latent processes where interventions may alter not just the mechanisms on a given graph, but the law that generates the graph. POSCMs address this limitation by treating structure formation as an endogenous causal mechanism subject to its own intervention semantics (Sec. 4).

Context-dependent mechanisms. A central idea in modern causal inference is that mechanisms may shift across environments, and that such shifts aid identification. Invariance-based frameworks exploit this: invariant causal prediction [Peters et al., 2016] identifies causal parents as those whose conditional holds invariant across environments, joint causal inference [Mooij et al., 2020] pools observational and interventional datasets by modeling regime explicitly, and transportability theory [Pearl and Bareinboim, 2022] characterizes when causal queries transfer across domains. A related line of work makes the regime indicator a first-class variable in the graphical model: Dawid [2021]'s decision-theoretic framework introduces a regime node to unify observational and interventional laws, while the data-fusion literature uses selection variables to flag which node-level mechanisms differ between settings [Bareinboim and Pearl, 2013, Perry et al., 2022, Lee et al., 2024]. Separately, context-specific independence models and their graphical generalizations (staged trees, chain event graphs, context-specific trees) represent conditional independencies that hold only in certain regimes [Boutilier et al., 1996, Collazo et al., 2020, Riccomagno and Smith, 2007, Duarte and Solus, 2025]. Across all these frameworks, context or regime modulates *node-level* mechanisms, i.e., the full conditional $P(X_i|\mathrm{Pa}_i)$. They do not decompose mechanisms into edge-specific components that may vary independently with context. This matters in practice: for instance, genetically encoded tools can now ablate individual synapses in neural circuits [Bareghamyan et al., 2025], an intervention that targets a single edge mechanism rather than an entire node. POSCMs extend the context-dependence picture to this finer granularity: contexts are latent regime variables that jointly supervise both the node-level value mechanisms and the edge-formation law, so that a shift in context can alter not only how a variable responds to its parents, but which parents it has and how each parent's influence is transmitted.

Latent contexts and partial observability. Because contexts are latent, POSCMs connect causal discovery with latent variables and measurement error [Carroll et al., 2006, Schennach, 2016], and causal representation learning, which seeks latent factors that render mechanisms stable across environments [Schölkopf et al., 2021]. Additionally, since POSCMs model interacting entities whose latent contexts stochastically generate the interaction structure itself, they also connect to relational causal models [Maier et al., 2013] which formalize causal reasoning over repeated interacting units. However, the aforementioned models do not treat edge formation as an interventionable causal mechanism.

Functional decomposition and edge-level interventions. To intervene on individual edges, one needs a decomposition of node mechanisms into per-parent contributions. A long line of work exploits functional assumptions (e.g., additive noise) to

identify causal directionality and decompose mechanisms into interpretable components [Bühlmann et al., 2013]. Existing edge-level intervention formalisms take a different approach: the hierarchy of Shpitser and Tchetgen [2014] defines edge interventions as value-routing, i.e., sending different values along different pathways within a fixed mechanism, while soft interventions [Correa and Bareinboim, 2020, Jaber et al., 2020] modify mechanisms but typically at the node level. Our message-augmented representation (Sec. 3) plays an analogous role to additive-noise decompositions: it separates a context-to-structure kernel from dyad-local value mechanisms, providing the functional primitives on which true edge-mechanism interventions (Sec. 4) are defined.

B KAS EDGE-FUNCTIONAL DECOMPOSITION BEYOND CONTINUOUS COMPACT MECHANISMS

KAS-style decompositions are classically stated for *continuous* mechanisms on *compact* domains (e.g., $[0,1]^n$) [Kolmogorov, 1957, Arnold, 1957, Sprecher, 1996]. In contrast, SCMs (and hence POSCMs) typically allow arbitrary measurable mechanisms on non-compact domains (e.g., $\mathbb{R}^n$), and may be discontinuous. For probabilistic modeling, it is natural to measure approximation error in $L^p(\mu)$ under the parent distribution μ. The following standard truncation-approximation argument shows that KAS edge-functional decomposition remains dense in this sense.

Lemma B.1 (An $L^p(\mu)$ density statement for KAS edge-functional decompositions)**.** *Let $X \sim \mu$ be an $\mathbb{R}^n$-valued random vector and let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function with $\mathbb{E}[|f(X)|^p] < \infty$ for some $1 \le p < \infty$. For any $\varepsilon > 0$, there exists a measurable function $g : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\mathbb{E}[\,|f(X) - g(X)|^p\,] < \varepsilon,$$

and such that g is representable in KAS edge-functional decomposition form (on a sufficiently large compact cube), in the following sense: there exists $R < \infty$ for which $g|_{[-R,R]^n}$ admits a KAS representation after affine rescaling of $[-R,R]^n$ to $[0,1]^n$.

Proof. Write $\|\cdot\|_p$ for the $L^p(\mu)$ norm:

$$\|h\|_p := \Big(\mathbb{E}\big[|h(X)|^p\big]\Big)^{1/p}.$$

Fix $\varepsilon > 0$ and write $\varepsilon_p := \varepsilon^{1/p}$. It suffices to construct g with $\|f-g\|_p < \varepsilon_p$, since then $\mathbb{E}[|f(X)-g(X)|^p] = \|f-g\|_p^p < \varepsilon$. We split this $L^p(\mu)$ error budget evenly and set $\delta := \varepsilon_p/3$.

Step 1 (domain truncation). Let $C_R := [-R,R]^n$ and define $f_R(x) := f(x)\mathbf{1}\{x \in C_R\}$. Then

$$\|f - f_R\|_p^p = \mathbb{E}\big[\,|f(X)|^p\,\mathbf{1}\{\|X\|_\infty > R\}\big].$$

As $R \to \infty$, the indicator $\mathbf{1}\{\|X\|_\infty > R\}$ decreases pointwise to 0, and the integrand is dominated by $|f(X)|^p \in L^1$. By the dominated convergence theorem Lebesgue [1910], $\mathbb{E}[\,|f(X)|^p\mathbf{1}\{\|X\|_\infty > R\}] \to 0$. Hence we may choose R such that

$$\|f - f_R\|_p < \delta.$$

Step 2 (range truncation / clipping). For $B > 0$, define the clipping operator

$$\mathrm{clip}_B(y) := \mathrm{sign}(y)\min\{|y|, B\}.$$

This map is 1-Lipschitz on $\mathbb{R}$: for all $a, b \in \mathbb{R}$,

$$\big|\mathrm{clip}_B(a) - \mathrm{clip}_B(b)\big| \le |a-b|.$$

In particular, if $|b| \le B$ (so $\mathrm{clip}_B(b) = b$), then

$$|a - \mathrm{clip}_B(a)| \le |a-b| + \big|\mathrm{clip}_B(a) - \mathrm{clip}_B(b)\big| \le 2|a-b|.$$

Let $\tilde{f}(x) := \mathrm{clip}_B(f_R(x))$. Then $\tilde{f}$ is measurable, supported on C_R, and $|\tilde{f}| \le B$. Moreover, for each x,

$$|f_R(x) - \tilde{f}(x)| = \big(|f_R(x)| - B\big)_+ = \big(|f_R(x)| - B\big)\mathbf{1}\{|f_R(x)| > B\}, \qquad \text{so} \qquad |f_R(x) - \tilde{f}(x)|^p \le |f_R(x)|^p\,\mathbf{1}\{|f_R(x)| > B\}.$$

Since $|f_R(X)|^p \le |f(X)|^p$ and $|f(X)|^p$ is integrable, the dominated convergence theorem [Lebesgue, 1910] gives $\mathbb{E}[|f_R(X)|^p \mathbf{1}\{|f_R(X)| > B\}] \to 0$ as $B \to \infty$. Hence we may choose B such that

$$\|f_R - \tilde{f}\|_p < \delta.$$

Step 3 (Lusin + Tietze: continuous approximation on C_R). Set $\eta := (\delta/2B)^p$. By Lusin's theorem [Lusin, 1912], there exists a compact $K \subseteq C_R$ such that $\tilde{f}|_K$ is continuous and $\mu(C_R \setminus K) < \eta$. By the Tietze extension theorem (since C_R is normal) [Tietze, 1915], there exists a continuous function $h : C_R \to \mathbb{R}$ such that $h = \tilde{f}$ on K and $\sup_{x \in C_R} |h(x)| \le B$. Since $h = \tilde{f}$ on K and both are bounded by B on C_R, for $x \in C_R \setminus K$ we have $|\tilde{f}(x) - h(x)| \le 2B$. Therefore,

$$\|\tilde{f} - h\|_p^p = \mathbb{E}\Big[|\tilde{f}(X) - h(X)|^p \, \mathbf{1}\{X \in C_R\}\Big] \le (2B)^p \, \mu(C_R \setminus K) < (2B)^p \, \eta = \delta^p,$$

hence

$$\|\tilde{f} - h\|_p < \delta.$$

Step 4 (KAS representation on the compact cube). Define the affine homeomorphism $T_R : C_R \to [0,1]^n$ by

$$T_R(x) := \frac{x + R\mathbf{1}}{2R} \quad \text{(componentwise)}, \qquad T_R^{-1}(u) := 2Ru - R\mathbf{1},$$

where $\mathbf{1}$ is the all-ones vector in $\mathbb{R}^n$. Let $\bar{h}(u) := h(T_R^{-1}(u))$, which is continuous on $[0,1]^n$.

By the Kolmogorov-Arnold-Sprecher representation theorem for continuous functions on $[0,1]^n$ [Kolmogorov, 1957, Arnold, 1957, Sprecher, 1996], there exist real constants $\eta, \lambda_1, \ldots, \lambda_n$, a continuous function $\Phi : \mathbb{R} \to \mathbb{R}$, and a continuous increasing function $\varphi : [0,1] \to [0,1]$ such that for all $u \in [0,1]^n$,

$$\bar{h}(u) = \sum_{q=0}^{2n} \Phi\left(\sum_{p=1}^{n} \lambda_p \, \varphi(u_p + \eta q) + q \right).$$

Define $k(u)$ to be the right-hand side, and set

$$g_R(x) := k(T_R(x)) \quad \text{for } x \in C_R.$$

Then *exactly* $g_R(x) = h(x)$ for all $x \in C_R$. Finally, define a measurable $g : \mathbb{R}^n \to \mathbb{R}$ by

$$g(x) := \begin{cases} g_R(x), & x \in C_R, \\ 0, & x \notin C_R. \end{cases}$$

Clarification (global domain). Classical KAS is stated on compact domains; we therefore interpret "g is KAS-representable" as "$g|_{C_R}$ is KAS-representable after rescaling". The arbitrary definition of g outside C_R is immaterial for the $L^p(\mu)$ approximation since $\mu(C_R^c)$ can be made arbitrarily small by Step 1.

Step 5 (combine errors with Minkowski). On $\mathbb{R}^n$ we have

$$f - g = (f - f_R) + (f_R - \tilde{f}) + (\tilde{f} - h) + (h - g).$$

Moreover, $h - g = 0$ μ-a.e. because $h = g$ on C_R and both vanish outside C_R under our extensions. Hence, by Minkowski's inequality (valid for $p \ge 1$) [Minkowski, 1910],

$$\|f - g\|_p \le \|f - f_R\|_p + \|f_R - \tilde{f}\|_p + \|\tilde{f} - h\|_p < \delta + \delta + \delta = \varepsilon^{1/p}.$$

Therefore $\mathbb{E}[|f(X) - g(X)|^p] = \|f - g\|_p^p < \varepsilon$, as desired.

Lemma B.1 is not an identifiability result; it simply supports our use of KAS edge-functional decomposition as an existence theorem for well-posed surgical edge interventions under a distributional notion of approximation. □

Identification target (latent A)	β-node	V-node	β-edge	V-edge
Kernels $(\alpha, \{\phi_i\}, \Gamma)$ (Thm. 5.4)	sufficient	sufficient†	not needed	not needed
Value dyadic primitives $\{H^V_{i\leftarrow j}\}$ and slot-response kernels of F_i (Thm. C.3)	not needed	used†	not needed	used
Context dyadic primitives $\{H^\beta_{i\leftarrow j}\}$ and slot-response kernels of Φ_i (Thm. C.3)	used	used†	used	not needed
Realized adjacency $A^\star$ (Prop. C.4(i))	not needed	used	not needed	not needed

Table 5: Summary of which primitive intervention types are used by different identification targets in this paper when adjacency is not directly observed. "Sufficient/used" refers to sufficient intervention families under the stated assumptions; "not needed" means the primitive does not appear in that result. † V-node interventions enter through the probing-based structure readout protocol (Assumption D.1). Necessity in the strict sense is relative to $(\mathcal{I}, \mathcal{O})$; we give explicit impossibility results for missing β-level access (Prop. 5.2) and for predicting edge-intervention effects under latent edges (Lemma 5.3).

C INTERVENTION HIERARCHY EXTENSIONS

C.1 NECESSARY VS. SUFFICIENT INTERVENTION TYPES ACROSS THE HIERARCHY

Which intervention tiers are (and are not) needed for identifying $(\alpha, \{\phi_i\}, \Gamma)$ when A is unobserved. Theorem 5.4 shows that to identify the *population* kernels $(\alpha, \{\phi_i\}, \Gamma)$ in the sense of conditional laws, it suffices to use *node* interventions at the context and value levels, $\mathcal{I}_{\beta\text{-node}} \cup \mathcal{I}_{V\text{-node}}$; edge-message interventions are not required for this kernel-level target. At the same time, "necessity" is relative to the chosen $(\mathcal{I}, \mathcal{O})$ (Def. 5.1); our results establish two concrete barriers and one additional readout requirement:

(i) If contexts are unobserved and *no* β-level interventions are available, then $(\alpha, \{\phi_i\}, \Gamma)$ are not identifiable beyond context reparameterization (Proposition 5.2); V-node/V-edge interventions in Phase II cannot break a symmetry that lives in Phase I.

(ii) If adjacency is unobserved, some *structure readout* resource is required to condition on events such as $\{\mathrm{Pa}_A(i) = S\}$; in this paper it is provided by repeated Phase II probing using V-node interventions (Assumption D.1), with a sufficient condition stated in Proposition C.4(i).

(iii) Edge-message interventions become essential only when the identification target is *finer* than kernels; for instance, to predict the effects of edge interventions themselves under latent edges (Lemma 5.3), or to identify dyadic message primitives in a message-augmented representation (Theorem C.3).

An intermediate regime: β-edge interventions without β-node interventions. Proposition 5.2 rules out identification when β is unobserved and no β-*level* interventions are available. A natural intermediate experimental regime, relevant in settings where one can perturb a specific communication channel (synapse, interaction) but cannot directly "set" a unit's latent type, is to allow β-edge interventions while disallowing β-node interventions. Because an edge intervention replaces an internal message map with an *externally specified* function $\widetilde{H}$, it need not respect the reparameterization symmetry used in Proposition 5.2. In particular, constant clamps $\widetilde{H}(\cdot) \equiv m$ erase dependence on the latent context and therefore do not, by themselves, anchor a continuous context coordinate system. In contrast, non-constant replacements can in principle break the symmetry: under a reparameterization γ of the latent context space, the post-intervention message becomes $\widetilde{H}(\beta_s)$ in one model and $\widetilde{H}(\gamma(\beta_s))$ in a reparameterized model, which generically induces different interventional laws unless $\widetilde{H}$ is invariant to γ. This suggests that sufficiently rich β-edge interventions may partially substitute for β-node interventions for kernel identification at non-root nodes, whereas root contexts remain intrinsically harder to probe in Phase I because they have no incoming β-channels. We leave a formal minimality characterization of intervention families in such restricted regimes to future work.

Assumption C.1 (Edge-intervention richness and channel distinguishability)**.** *The family $\mathcal{I}$ contains a sufficiently rich class of edge-message replacements (Def. 4.2) to probe a single dyadic channel while holding other inputs fixed. In particular, for each dyad $j \to i$ the intervention family can implement constant "message clamps" $\widetilde{H}(\cdot) \equiv m$ for m in a set with non-empty interior in the message space.*

Moreover, fixing all other parent values by V-node interventions (Assumption D.6) and conditioning on the event $\{A_{ji} = 1\}$, the map

$$m \;\longmapsto\; \mathcal{L}\big(V_i \mid do_V(j \to i; \widetilde{H} \equiv m),\ do(V_{Pa_A(i)} = v),\ \beta_i = b,\ Pa_A(i) = S,\ A_{ji} = 1\big)$$

is injective on that reachable set (non-degenerate message channel).

Replay variant. *For the replay-based Route (C) in Theorem C.3, we additionally assume a pointwise (noise-fixed) version of channel distinguishability: for any fixed (v, b, S) as above (and fixed values of the other message slots induced by (v, S)), the map*

$$m \longmapsto F_i(\dots, m, \dots, u)$$

is injective on the reachable set for $\mathcal{L}(U_i^V \mid \beta_i = b, Pa_A(i) = S)$-almost every u.

Assumption C.2 (Paired-world replay (idealized))**.** *In a paired multi-world experiment, multiple interventions may be applied to the* same *underlying exogenous draw (i.e., the same latent noise realization), so that the only difference between two "worlds" is the intervention itself. This assumption is* not *required for Theorem 5.4; it is only used as an* idealized *route for identifying fine-grained dyadic message primitives in Theorem C.3(C).*

Theorem 5.4 is the *kernel-level* identification result: it establishes identifiability of the causal conditionals α, $\mathcal{L}(\beta_i \mid \beta_{\mathrm{Pa}_A(i)}, \mathrm{Pa}_A(i))$, and $\mathcal{L}(V_i \mid V_{\mathrm{Pa}_A(i)}, \beta_i, \mathrm{Pa}_A(i))$ from ordinary (independently sampled) interventional distributions. Under Assumption D.1, the adjacency is *never* directly observed but is recovered via a probing-based readout procedure (Remark D.2), so that events such as $\{\mathrm{Pa}_A(i) = S\}$ may be conditioned on after this recovery.

From kernels to dyadic message mechanisms. Theorem 5.4 identifies node-level *kernels* (conditional laws) but does not identify a particular dyadic decomposition of those kernels (Remark 3.4). The next theorem shows that, under stronger experimental access, the dyadic message primitives used to define edge interventions become identifiable (up to gauge). This refinement is not required for kernel identification, but it clarifies when edge-mechanism questions are learnable from data.

Theorem C.3 (Dyadic message identifiability (three routes))**.** *Assume the POSCM admits a message-augmented representation (Def. 3.1). Assume structure readout (Assumption D.1),* context readout *(Assumption D.3), value readout (Assumption D.4), positivity (Assumption D.7), and edge-intervention richness (Assumption C.1). Then the* value *message primitives $\{H^V_{i\leftarrow j}\}_{j<i}$ are identifiable on the intervention-reachable support, up to representation gauge (Remark 3.4). Moreover, under Routes (A)–(B) below, for each target node i and dyad $j \to i$, the induced j-slot* response kernel

$$m \longmapsto \mathcal{L}\big(V_i \mid do_V(j \to i; \widetilde{H} \equiv m),\ do(V_{Pa_A(i)} = v),\ \beta_i = b,\ Pa_A(i) = S,\ A_{ji} = 1\big)$$

is identifiable on the reachable support. (In general this identifies only these pushforward kernels, not the deterministic map F_i separately from the latent noise U_i^V; cf. Proposition C.4(iii).) The message-primitive identification holds under any one of the following routes:

- *(A)* ***Joint node controls (no replay).*** *If value-intervention coverage holds (Assumption D.6), then for each target node i and dyad $j \to i$, combining (i) joint V-node interventions that fix all parent values $V_{Pa_A(i)}$ and (ii) a family of* single-edge *V-edge message clamps on $j \to i$ with varying constants m isolates the jth dyadic channel across independent experimental units, identifying $H^V_{i\leftarrow j}$ and the associated j-slot response kernel on the reachable support.*
- *(B)* ***Message clamping (no replay).*** *Under constant clamps $\widetilde{H}(\cdot) \equiv m$, comparing the baseline distribution at $do(V_j = v)$ (holding other parent values fixed) to the family of clamped distributions recovers the unique message value $m = H^V_{i\leftarrow j}(v)$ that matches the baseline (uniqueness by Assumption C.1), thereby identifying $H^V_{i\leftarrow j}(\cdot)$ on the reachable support.*
- *(C)* ***Paired-world replay (idealized).*** *Under paired-world replay (Assumption C.2) together with the pointwise distinguishability condition in Assumption C.1, replay blocks allow within-unit calibration across clamp values while holding all other inputs and the full exogenous draw fixed, identifying $H^V_{i\leftarrow j}$ on the reachable support* without *joint parent-value controls (provided V_j varies across replay blocks over that support).*

Analogous statements hold for the context *message primitives $\{H^\beta_{i\leftarrow j}\}$ and the induced slot-response kernels of $\{\Phi_i\}$, replacing V-operations with β-operations (and using context readout to access the relevant conditioning events).*

Population kernels versus realized instances. Theorem 5.4 concerns identifiability of population-level kernels from interventional distributions across independently sampled units. In applications one may also wish to recover the realized adjacency and mechanisms of a *particular* stationary instance. The following proposition clarifies this distinction and provides a sufficient-condition companion to Assumption D.1.

Proposition C.4 (Identifying a realized draw $(A^\star, f^\star)$ from identified kernels)**.** *Theorem 5.4 identifies* population-level *kernels. A particular realized draw $A^\star$ and mechanism collection $f^\star$ (Remark 2.3) is* not *identifiable from population-level interventional distributions alone, since kernels do not reveal which random outcome occurred. However, if the protocol permits repeated probing of the* same *underlying instance and the instance is stationary across probes, then $A^\star$ and (in some regimes) $f^\star$ become identifiable.*

(i) If (a) the instance is stationary across repeated probes with fresh Phase II value noise and measurement noise resampled i.i.d. across probes (Assumption D.1), (b) joint V-node interventions are available (Assumption D.6), (c) value readout holds (Assumption D.4), and (d) each true edge is controlled edge-faithful *in the following sense: whenever $A^\star_{ji} = 1$ there exist values $v \neq v'$ and an assignment v_{-j} to $V_{-j} := (V_k)_{k<i,\, k\neq j}$ such that*

$$\mathcal{L}\big(V_i \mid do(V_j = v,\ V_{-j} = v_{-j})\big) \neq \mathcal{L}\big(V_i \mid do(V_j = v',\ V_{-j} = v_{-j})\big),$$

then $A^\star$ is identifiable in the large-sample limit. Moreover, for any desired error level $\delta > 0$, a finite number of probes per tested intervention setting suffices to recover $A^\star$ with probability at least $1 - \delta$ under standard finite-sample two-sample distinguishability/testing conditions.

(ii) If Γ is deterministic *given $(\beta_i^\star, Pa_{A^\star}(i))$ (i.e., $f_i = \Gamma(\beta_i, Pa_A(i))$ with no mechanism-randomness), then once $(A^\star, \beta^\star)$ are identified we have $f_i^\star = \Gamma(\beta_i^\star, Pa_{A^\star}(i))$.*

(iii) If Γ is stochastic*, then repeated probes identify the* unit-specific response kernels $v \mapsto \mathcal{L}(V_i \mid do(V_{Pa_{A^\star}(i)} = v), f_i^\star)$ *on the probed domain. Moreover, if the mechanism is distinguishable from its probed response kernel in the sense that the map*

$$f_i \longmapsto \big\{\mathcal{L}(V_i \mid do(V_{Pa_{A^\star}(i)} = v), f_i)\big\}_{v \in \mathcal{V}_{\mathrm{probe}}}$$

is injective on the candidate class (modulo observational equivalence on $\mathcal{V}_{\mathrm{probe}}$), then $f_i^\star$ is identified up to that equivalence. Identifying the structural function *$f_i^\star(\cdot,\cdot)$ itself is strictly stronger and, in general, requires additional noise-model structure (e.g., noise-free mechanisms, known additive noise, monotone structural quantiles, or a parametric mechanism class).*

A full statement and proof sketch are given in Appendix D.5.

The readout assumptions in Theorem 5.4 directly address the barriers identified in Proposition 5.2 (context readout) and Lemma 5.3 (structure readout).

D IDENTIFIABILITY PROOFS

Readout and coverage assumptions for Theorem 5.4.

Assumption D.1 (Experimental structure readout (latent adjacency))**.** *The adjacency A is latent (not directly observed). Instead, the protocol provides a structure readout mechanism* via multiple intervention experiments*: each experimental unit may be probed repeatedly in Phase II while its realized $(A^\star, \beta^\star, f^\star)$ remains stationary across probes. Across probes, Phase II exogenous randomness (value noise) and measurement noise are resampled i.i.d., so repeated probes yield i.i.d. draws conditional on $(A^\star, \beta^\star, f^\star)$. The experimenter may apply joint V-node interventions (Assumption D.6) across probes. These repeated probes suffice to identify the realized adjacency $A^\star$ in the large-sample limit (a sufficient condition is given in Proposition C.4(i)). Whenever results below condition on events involving A (e.g., $\{Pa_A(i) = S\}$), this is understood in terms of the identified adjacency returned by the probing procedure.*

Remark D.2 (Structure readout via repeated probing)**.** *Assumption D.1 rules out direct observation of A. Instead, it posits that adjacency can be recovered from* multiple intervention experiments *on a stationary instance. A sufficient condition is that value-intervention coverage holds and each realized edge is edge-faithful in a* controlled *sense (Proposition C.4(i)), in which case repeated Phase II probes identify the realized adjacency $A^\star$ (Proposition C.4(i)). Proposition C.4(i) is a sufficient-condition companion to Assumption D.1 and is proved independently of Theorem 5.4. Kernel identification in Theorem 5.4 still relies on sampling* independent *instances under interventions; repeated probing of a single instance identifies only $A^\star$ and, when Γ is stochastic, unit-specific response kernels (Proposition C.4(iii)).*

Assumption D.3 (Context readout)**.** *Contexts are observed (possibly noisily) through $\tilde{\beta} = \mathcal{O}^\beta(\beta, U_\beta^{\mathcal{O}})$ in a way that makes β identifiable up to any explicitly stated equivalence (e.g., label permutation).*

Assumption D.4 (Value readout)**.** *Endogenous values are observed (possibly noisily) through $\tilde{V} = \mathcal{O}^V(V, U_V^{\mathcal{O}})$ in a way that makes V identifiable up to any explicitly stated equivalence.*

Assumption D.5 (Context-intervention coverage)**.** *For each node j, the family of β-node interventions can set β_j to values in a set with non-empty interior (continuous case) or can visit all labels (discrete case). Moreover, the protocol supports* joint *β-node interventions: for any subset $S \subseteq \{0, \ldots, N-1\}$ and any tuple b_S in the intervention-reachable set, the intervention $do(\beta_S = b_S)$ is admissible.*

Assumption D.6 (Value-intervention coverage)**.** *For each node j, the family of V-node interventions can set V_j to values in a set with non-empty interior (continuous case) or over full support (discrete case). Moreover, the protocol supports* joint *V-node interventions: for any subset $S \subseteq \{0, \dots, N-1\}$ and any tuple v_S in the intervention-reachable set, the intervention $do(V_S = v_S)$ is admissible. In particular, for each node i and each realized parent set $S \subseteq \{j : j < i\}$, the family of V-node interventions can set $(V_j)_{j \in S}$ jointly to any tuple v_S in a set with non-empty interior (continuous case) or over full support (discrete case). Interventions at the β level and the V level may be combined in the same experiment (Phase I + Phase II).*

Assumption D.7 (Positivity)**.** *On the intervention-reachable support, for all dyads (j, i) with $j < i$ and all reachable values b,*

$$0 < \mathbb{P}\big(A_{ji} = 1 \mid do(\beta_j = b)\big) < 1.$$

Scope and assumption boundary. The results in this section are *conditional identifiability* statements: they show that, assuming the experimental capabilities in Assumptions D.1–C.2, the corresponding kernels/messages/mechanisms are identifiable from the stated interventional distributions. In particular, the readout/coverage/injectivity conditions are not derived from the bare POSCM generative semantics; they encode additional experimental structure that breaks the symmetries highlighted in Proposition 5.2 and Lemma 5.3.

D.1 PROOF OF PROPOSITION 5.2

Proof. We show that when contexts are unobserved and no β-level interventions are available, the model admits a context-reparameterization symmetry that cannot be broken by V-level data.

Fix any POSCM $\mathcal{M}$ and any diffeomorphism $\gamma : \mathcal{B} \to \mathcal{B}$. We construct a second POSCM $\mathcal{M}'$ with different $(\mathbb{P}(A \mid \beta), \{\phi_i\}, \Gamma)$ that is $(\mathcal{I}, \mathcal{O})$-equivalent to $\mathcal{M}$ for $\mathcal{I} = \mathcal{I}_{\text{obs}} \cup \mathcal{I}_{V\text{-node}} \cup \mathcal{I}_{V\text{-edge}}$.

Construction of $\mathcal{M}'$. Let $\beta'_i := \gamma(\beta_i)$ for all i and define $\mathcal{M}'$ via pushforward/composition:

$$\mathbb{P}_{\beta'_1} := \gamma_{\#}\mathbb{P}_{\beta_1}, \tag{7}$$

$$\mathbb{P}'(A_{ji} = 1 \mid \beta'_j) := \mathbb{P}(A_{ji} = 1 \mid \gamma^{-1}(\beta'_j)), \tag{8}$$

$$\phi'_i(\{\beta'_k\}_{k \in \mathrm{Pa}(i)}, U_i^\beta) := \gamma\Big(\phi_i(\{\gamma^{-1}(\beta'_k)\}_{k \in \mathrm{Pa}(i)}, U_i^\beta)\Big), \tag{9}$$

$$\mathbb{P}'(f_i \mid \beta'_i, \mathrm{Pa}_A(i)) := \mathbb{P}(f_i \mid \gamma^{-1}(\beta'_i), \mathrm{Pa}_A(i)). \tag{10}$$

All other components (value-noise laws, etc.) are unchanged, and the measurement model $\mathcal{O}$ is held fixed (Def. 2.2).

Equality of observable interventional laws. Consider any intervention $\iota \in \mathcal{I}$, which acts only at the V level (Phase II). By construction (7)–(10), $\mathcal{M}$ and $\mathcal{M}'$ induce the same interventional law for (A, V) after marginalizing out β (since β' is a reparameterization of β and all kernels/mechanism-assignment factors are composed accordingly). Under the factorized measurement model of Def. 2.2, the observable channels $\tilde{A} = \mathcal{O}^A(A, U_A^{\mathcal{O}})$ and $\tilde{V} = \mathcal{O}^V(V, U_V^{\mathcal{O}})$ do not depend on β. When β is unobserved, $\mathcal{O}^\beta$ outputs nothing, so the induced interventional distributions over observables $(\tilde{A}, \tilde{V})$ coincide under $\mathcal{M}$ and $\mathcal{M}'$ for all $\iota \in \mathcal{I}$. Therefore $\mathbb{P}(A \mid \beta)$, $\{\phi_i\}$, and Γ are not $\mathcal{I}$-identifiable beyond context reparameterization. □

D.2 PROOF OF LEMMA 5.3

Proof. We construct an explicit counterexample showing that two POSCMs can agree on all node-level interventional distributions yet disagree on V-edge intervention counterfactuals.

Setup. Consider a two-node ordered POSCM with $\tau(1) < \tau(2)$ and a single potential dyad $1 \to 2$. Let contexts be degenerate (omit β). Let $V_1 \in \{0, 1\}$ be endogenous variable. Let $A_{12} \sim \mathrm{Bern}(p)$ be unobserved.

Define the value mechanism at node 2 as follows:

- If $A_{12} = 0$: node 2 has no parents and outputs $V_2 \sim \mathrm{Bern}(1/2)$.
- If $A_{12} = 1$: node 2 has parent set $\{1\}$ and outputs $V_2 \sim \mathrm{Bern}(q_{V_1})$ for parameters $q_0, q_1 \in (0, 1)$.

Node-level interventional distributions. For each node intervention $\mathrm{do}(V_1 = v)$, $v \in \{0, 1\}$, we have:

$$\mathbb{P}(V_2 = 1 \mid \mathrm{do}(V_1 = v)) = (1 - p) \cdot \tfrac{1}{2} + p \cdot q_v.$$

Model $\mathcal{M}$. Fix $p = \frac{1}{2}$ and $(q_0, q_1) = (0.2, 0.8)$. This yields:

$$\mathbb{P}_{\mathcal{M}}(V_2 = 1 \mid \mathrm{do}(V_1 = 0)) = 0.5 \cdot 0.5 + 0.5 \cdot 0.2 = 0.35,$$
$$\mathbb{P}_{\mathcal{M}}(V_2 = 1 \mid \mathrm{do}(V_1 = 1)) = 0.5 \cdot 0.5 + 0.5 \cdot 0.8 = 0.65.$$

Model $\mathcal{M}'$. Choose a different edge probability $p' = 0.8$ and define (q'_0, q'_1) by solving:

$$(1 - p') \cdot \tfrac{1}{2} + p' \cdot q'_v = (1 - p) \cdot \tfrac{1}{2} + p \cdot q_v, \quad v \in \{0, 1\}.$$

Substituting:

$$0.2 \cdot 0.5 + 0.8 \cdot q'_0 = 0.35 \quad \Rightarrow \quad q'_0 = (0.35 - 0.1)/0.8 = 0.3125,$$
$$0.2 \cdot 0.5 + 0.8 \cdot q'_1 = 0.65 \quad \Rightarrow \quad q'_1 = (0.65 - 0.1)/0.8 = 0.6875.$$

Both $q'_0, q'_1 \in (0, 1)$, so $\mathcal{M}'$ is a valid POSCM.

Node-level equivalence. By construction:

$$\mathbb{P}_{\mathcal{M}}(V_2 = 1 \mid \mathrm{do}(V_1 = v)) = \mathbb{P}_{\mathcal{M}'}(V_2 = 1 \mid \mathrm{do}(V_1 = v)) \quad \text{for } v \in \{0, 1\}.$$

Message-augmented realization (Def. 3.1). To match the edge-intervention definition in Def. 4.2, realize the above mechanism in message-augmented form with message dimension $d_2 = 2$. Let $U_2^V \sim \mathrm{Unif}(0, 1)$ be exogenous variable and define the (baseline) value-message primitive

$$H^V_{2\leftarrow 1}(v) := (1, q_v), \qquad M^V_{2\leftarrow 1} := A_{12} H^V_{2\leftarrow 1}(V_1) = (A_{12}, A_{12} q_{V_1}),$$

where the first coordinate records whether the edge is present. Define the aggregator

$$F_2\big((m_1, m_2), u\big) := \mathbf{1}\Big\{u < (1 - m_1)\tfrac{1}{2} + m_2\Big\}.$$

Then $V_2 = F_2(M^V_{2\leftarrow 1}, U_2^V)$ yields $V_2 \sim \mathrm{Bern}(1/2)$ when $A_{12} = 0$ and $V_2 \sim \mathrm{Bern}(q_{V_1})$ when $A_{12} = 1$, as above.

V-edge intervention counterfactuals differ. Consider the V-edge message intervention of Def. 4.2

$$\iota := \mathrm{do}_V(1 \to 2;\, \widetilde{H}^V_{2\leftarrow 1}), \qquad \widetilde{H}^V_{2\leftarrow 1}(v) := (1, v).$$

Equivalently, when $A_{12} = 1$ this replaces q_v by $\tilde{q}_v = v$ and hence forces $V_2 = V_1$. Because messages are gated by A_{12}, ι is a no-op when $A_{12} = 0$. Under ι, for $v \in \{0, 1\}$:

$$\mathbb{P}\big(V_2 = 1 \mid \iota,\, \mathrm{do}(V_1 = v)\big) = (1 - p) \cdot \tfrac{1}{2} + p \cdot v.$$

In $\mathcal{M}$ (with $p = 0.5$):

$$\mathbb{P}_{\mathcal{M}}(V_2 = 1 \mid \iota,\, \mathrm{do}(V_1 = 0)) = 0.5 \cdot 0.5 + 0.5 \cdot 0 = 0.25,$$
$$\mathbb{P}_{\mathcal{M}}(V_2 = 1 \mid \iota,\, \mathrm{do}(V_1 = 1)) = 0.5 \cdot 0.5 + 0.5 \cdot 1 = 0.75.$$

In $\mathcal{M}'$ (with $p' = 0.8$):

$$\mathbb{P}_{\mathcal{M}'}(V_2 = 1 \mid \iota,\, \mathrm{do}(V_1 = 0)) = 0.2 \cdot 0.5 + 0.8 \cdot 0 = 0.10,$$
$$\mathbb{P}_{\mathcal{M}'}(V_2 = 1 \mid \iota,\, \mathrm{do}(V_1 = 1)) = 0.2 \cdot 0.5 + 0.8 \cdot 1 = 0.90.$$

Since $p \neq p'$, the post-edge-intervention distributions differ between $\mathcal{M}$ and $\mathcal{M}'$.

Conclusion. Node-level interventional data do not determine V-edge intervention counterfactuals in the nonparametric latent-edge setting. The underlying ambiguity is *structure-mechanism confounding*: a model with frequent edges and weak edge-gated mechanisms can be observationally equivalent (at the node level) to a model with rare edges and strong edge-gated mechanisms, yet these models make different predictions about what would happen under edge-level interventions. □

D.3 PROOF OF THEOREM 5.4 (KERNEL IDENTIFIABILITY; NO PAIRED-WORLD REPLAY)

Proof. Let $\mathcal{M}$ and $\mathcal{M}'$ be two POSCMs that are $(\mathcal{I}_{\text{kern}}, \mathcal{O})$-equivalent, where $\mathcal{I}_{\text{kern}} := \mathcal{I}_{\text{obs}} \cup \mathcal{I}_{\beta\text{-node}} \cup \mathcal{I}_{V\text{-node}}$. By Assumptions D.3 and D.4, equality of interventional distributions over the observable readouts implies equality of the corresponding interventional distributions over (β, V) (up to any explicitly stated measurement equivalences). Moreover, by Assumption D.1 the adjacency A is identifiable under each intervention via the probing-based structure readout (Remark D.2), so the interventional law of A is identified as well. We therefore reason below as if (A, β, V) were available, with A obtained via the probing readout.

Proof template (decoder/inversion ⇒ identifiability). For each target kernel family in Theorem 5.4, the goal is to show that it is a (deteministic) function of the restricted interventional law family $\{\mathcal{L}(A, \beta, V \mid \iota)\}_{\iota \in \mathcal{I}_{\text{kern}}}$. Equivalently, we exhibit an explicit *decoder* (an inversion map) from these laws to the target kernel. Therefore, if $\mathcal{M}$ and $\mathcal{M}'$ are $(\mathcal{I}_{\text{kern}}, \mathcal{O})$-equivalent, they induce the same restricted interventional laws, hence the same decoder output, and thus the same target kernel (up to the measurement equivalences in Assumptions D.3–D.4).

We prove identifiability of each kernel family from interventional distributions collected from *independent experimental units* (no paired-world replay).

(i) Structure kernel α. Fix a source node j and a reachable value b. Under the intervention $\text{do}(\beta_j = b)$ (admissible by Assumption D.5), the Phase I edge-formation step generates the outgoing edge vector $A_{j,>j}$ using the structure kernel evaluated at $\beta_j = b$. Therefore the interventional law

$$\mathcal{L}(A_{j,>j} \mid \text{do}(\beta_j = b))$$

is exactly the supervising measure $\mu_j(b)$ appearing in Theorem 5.4(i). By structure readout (Assumption D.1), this law is identified from the observable interventional distribution for each reachable b. Varying b over the intervention-reachable set identifies the supervising measures $\{\mu_j(\cdot)\}_j$ and, in particular, each dyadwise marginal $b \mapsto \mathbb{P}(A_{ji} = 1 \mid \text{do}(\beta_j = b))$ for $i > j$. Hence α is $(\mathcal{I}_{\text{kern}}, \mathcal{O})$-identifiable on the reachable support.

(ii) Context propagation kernels induced by $\{\phi_i\}$. Fix a node i and a parent set S with $\mathbb{P}(\text{Pa}_A(i) = S \mid \iota) > 0$ for some $\iota \in \mathcal{I}_{\text{kern}}$. Consider a joint β-node intervention $\text{do}(\beta_S = b_S)$, admissible by Assumption D.5. Under ordered generation, the Phase I context mechanism generates

$$\beta_i \;=\; \phi_i\big(\beta_S,\, U_i^{\beta}\big) \qquad \text{on the event } \{\text{Pa}_A(i) = S\}.$$

Thus, conditioning on $\text{Pa}_A(i) = S$ (which is identifiable by Assumption D.1) and on the intervention value $\beta_S = b_S$, the conditional interventional law

$$\mathcal{L}\big(\beta_i \mid \text{do}(\beta_S = b_S),\; \text{Pa}_A(i) = S\big)$$

coincides with the kernel $K^{\beta}_{i,S}(\cdot \mid b_S)$ in Theorem 5.4(ii). By context readout (Assumption D.3) and structure readout (Assumption D.1), this conditional law is identified from interventional data for each reachable b_S; varying b_S over the reachable set identifies $K^{\beta}_{i,S}$ on its reachable support. Equivalently, $\mathcal{L}(\beta_i \mid \beta_{\text{Pa}_A(i)}, \text{Pa}_A(i))$ is identified (up to any context reparameterization/gauge equivalences discussed in the main text when interventions are not anchored).

No replay is required: conditioning on identified (A, β_S) across independent units plays the same isolating role that replay played in the original paired-world argument.

(iii) Endogenous-variable mechanism kernels induced by Γ. Fix a node i and a parent set S in the support of $\text{Pa}_A(i)$ under $\mathcal{I}_{\text{kern}}$. Consider a V-node intervention $\text{do}(V_S = v_S)$, admissible by Assumption D.6. Since V-node interventions occur in Phase II, they do not alter Phase I variables (A, β), so we may condition on $\beta_i = b_i$ and $\text{Pa}_A(i) = S$ (identifiable by Assumptions D.1 and D.3).

Under the Phase II value-generation step, conditional on $\beta_i = b_i$ and $\text{Pa}_A(i) = S$, the node-i mechanism is sampled via Γ and then applied to the intervened parent values:

$$f_i \sim \mathbb{P}(\,\cdot \mid \beta_i = b_i, \text{Pa}_A(i) = S), \qquad V_i = f_i(v_S, U_i^V).$$

Therefore the conditional interventional law

$$\mathcal{L}\big(V_i \mid \text{do}(V_S = v_S),\; \beta_i = b_i,\; \text{Pa}_A(i) = S\big)$$

coincides with $K^V_{i,S}(\cdot \mid v_S, b_i)$ in Theorem 5.4(iii). By value readout (Assumption D.4) together with structure/context readout, this conditional law is identified for each reachable (v_S, b_i); varying (v_S, b_i) over the intervention-reachable set identifies $K^V_{i,S}$ on its reachable support. Equivalently, $\mathcal{L}(V_i \mid V_{\mathrm{Pa}_A(i)}, \beta_i, \mathrm{Pa}_A(i))$ is identified. Note that if Γ is stochastic, $K^V_{i,S}$ is a *population* conditional law that averages over the mechanism draw f_i; repeated probing of a single stationary instance instead identifies the corresponding *unit-specific* response kernel (Proposition C.4(iii)).

Again, no replay is required: the causal isolation comes from (a) ordered generation, and (b) the ability to fix parent values and condition on $\mathrm{Pa}_A(i)$ across independent experimental units. □

D.4 PROOF OF THEOREM C.3 (DYADIC MESSAGE IDENTIFIABILITY)

Proof. We sketch the argument for the *value* message primitives; the context case is analogous.

Fix a target node i and an upstream node $j < i$. In a message-augmented representation (Def. 3.1), the value-update admits a decomposition of the form

$$M^V_{i\leftarrow k} \;=\; A_{ki}\, H^V_{i\leftarrow k}(V_k), \qquad V_i \;=\; F_i\big(\{M^V_{i\leftarrow k}\}_{k<i},\, U^V_i\big),$$

up to representation gauge (Remark 3.4).

Throughout, we condition on a realized parent set $S = \mathrm{Pa}_A(i)$ and on a fixed context value $\beta_i = b$; these conditioning events are operational under Assumptions D.1 and D.3.

Route A (joint node controls; no replay). Assume value-intervention coverage (Assumption D.6) and edge-intervention richness (Assumption C.1). Fix a realized parent set $S = \mathrm{Pa}_A(i)$ and fix parent values v_S via a joint V-node intervention $\mathrm{do}(V_S = v_S)$. By structure readout we may condition on $\{\mathrm{Pa}_A(i) = S\}$ and (if needed) on $\{A_{ji} = 1\}$.

Under the baseline (no edge intervention), the jth message slot equals $m^\star := H^V_{i\leftarrow j}(v_j)$, and all other slots are fixed at $\{H^V_{i\leftarrow k}(v_k)\}_{k\in S\setminus\{j\}}$ or 0 for $k \notin S$. Therefore, for fixed (v_S, b, S) the conditional law of V_i is the pushforward of U^V_i through the (unknown) aggregator at message value $m^\star$:

$$\mathcal{L}\big(V_i \mid \mathrm{do}(V_S = v_S),\, \beta_i = b,\, \mathrm{Pa}_A(i) = S,\, A_{ji} = 1\big) \;=\; \mathcal{L}\big(F_i(\dots, m^\star, \dots, U^V_i) \mid \beta_i = b,\, \mathrm{Pa}_A(i) = S\big).$$

Now apply a V-edge message intervention on $j \to i$ (Def. 4.2) that replaces $H^V_{i\leftarrow j}$ by a *known* constant clamp $\widetilde{H}(\cdot) \equiv m$. Under the same joint parent-value control and the same conditioning on $(\beta_i = b, \mathrm{Pa}_A(i) = S, A_{ji} = 1)$, the resulting conditional law is

$$\mathcal{L}\big(V_i \mid \mathrm{do}_V(j \to i; \widetilde{H} \equiv m),\, \mathrm{do}(V_S = v_S),\, \beta_i = b,\, \mathrm{Pa}_A(i) = S,\, A_{ji} = 1\big) \;=\; \mathcal{L}\big(F_i(\dots, m, \dots, U^V_i) \mid \beta_i = b,\, \mathrm{Pa}_A(i) = S\big).$$

In particular, taking $m = m^\star$ reproduces the baseline conditional law, since F_i depends on the jth input only through the resulting message value. Varying m over a rich set identifies the induced *response kernel* $m \mapsto \mathcal{L}(V_i \mid \mathrm{do}_V(j \to i; \widetilde{H} \equiv m), \dots)$ on the reachable support. By injectivity in Assumption C.1, the baseline distribution matches *exactly one* member of this family, so $m^\star$ (and hence $H^V_{i\leftarrow j}(v_j)$) is uniquely identified on the reachable support (up to gauge). Repeating the argument for varying v_j (via V-node interventions) identifies the function $H^V_{i\leftarrow j}(\cdot)$ on the reachable support, and iterating over dyads identifies all value message primitives.

What is (and is not) identified. The same experiments identify the induced response kernels of the aggregator in the relevant slot. In general, they do *not* identify the deterministic map F_i separately from the latent noise U^V_i without additional noise-model structure (cf. Proposition C.4(iii)).

Route B (message clamping; no replay). This is the matching step in Route A emphasized as a calibration procedure: for fixed (v_S, b, S), the family of clamped interventional laws parameterized by m provides a "response curve" for the jth slot, and the baseline distribution at $\mathrm{do}(V_j = v)$ selects the unique $m = H^V_{i\leftarrow j}(v)$ on the reachable support.

Route C (paired-world replay; idealized). Under paired-world replay (Assumption C.2), we may evaluate a baseline world and multiple V-edge clamps on $j \to i$ on the *same* exogenous draw. Within a replay block ω, all non-descendants of the intervention target (including A, all β, and all upstream V_k for $k < i$) are identical across worlds, and the realized value

noise $U_i^V(\omega)$ is shared. Fix a block with $A_{ji} = 1$ and $\mathrm{Pa}_A(i) = S$, and write $v_j := V_j(\omega)$ and $m^\star := H^V_{i\leftarrow j}(v_j)$. Then, for each clamp value m we observe

$$V_i^{(m)}(\omega) \;=\; F_i(\ldots, m, \ldots, U_i^V(\omega)).$$

By the *pointwise* distinguishability condition in Assumption C.1 (Replay variant), the map $m \mapsto V_i^{(m)}(\omega)$ is injective for almost every block, so matching the baseline value $V_i^{(\mathrm{base})}(\omega) = V_i^{(m^\star)}(\omega)$ to the clamped curve identifies $m^\star$ within that block. Repeating across replay blocks whose V_j values range over the intervention-reachable support (either by natural variation or by intervening on V_j) identifies $H^V_{i\leftarrow j}(\cdot)$ pointwise on that support, up to gauge, *without* requiring joint interventions to fix the other parent values.

This completes the proof sketch for value messages; the context-message case is analogous. □

D.5 PROOF OF PROPOSITION C.4 (INSTANCE-LEVEL IDENTIFICATION)

Proof. We justify each item of Proposition C.4.

(i) Identifying $A^\star$ by repeated probing (stationary instance). Fix a dyad $j < i$. By joint value interventions (Assumption D.6), we may set $V_j = v$ while clamping the other potential parents $V_{-j} := (V_k)_{k<i,\, k\neq j}$ to an arbitrary assignment v_{-j}, i.e., apply $\mathrm{do}(V_j = v,\ V_{-j} = v_{-j})$. By Assumption D.1, repeated probes under a fixed intervention resample Phase II value noise and measurement noise i.i.d., so we obtain i.i.d. samples from the induced conditional law of V_i (or its readout, by Assumption D.4).

If $A^\star_{ji} = 0$, then the Phase II structural equation for V_i does not take V_j as an argument; once V_{-j} is clamped, changing v cannot change the law of V_i. If instead $A^\star_{ji} = 1$, controlled edge-faithfulness provides values $v \neq v'$ and some clamp v_{-j} such that the induced laws under $\mathrm{do}(V_j = v,\ V_{-j} = v_{-j})$ and $\mathrm{do}(V_j = v',\ V_{-j} = v_{-j})$ differ.

With sufficiently many repeated probes per intervention setting, empirical convergence together with any consistent two-sample test distinguishes equality versus inequality of these induced laws, yielding a consistent decision rule for the edge indicator $A^\star_{ji}$. Applying this decision over all dyads recovers $A^\star$.

(ii) Identifying $f^\star$ when Γ is deterministic. If Γ is deterministic given $(\beta_i, \mathrm{Pa}_A(i))$, then once $\beta^\star$ and $A^\star$ are identified we have

$$f_i^\star = \Gamma(\beta_i^\star, \mathrm{Pa}_{A^\star}(i)).$$

No additional data beyond identifying $(A^\star, \beta^\star)$ are required.

(iii) Stochastic Γ: identifying the realized response kernel vs. the structural function. If Γ is stochastic, then $f_i^\star$ is a latent draw that is fixed across probes. For any probed parent-value assignment v (implemented via $\mathrm{do}(V_{\mathrm{Pa}_{A^\star}(i)} = v)$), repeated probes yield i.i.d. samples from

$$\mathcal{L}\big(V_i \mid \mathrm{do}(V_{\mathrm{Pa}_{A^\star}(i)} = v),\ f_i^\star\big),$$

hence the unit-specific response kernel $v \mapsto \mathcal{L}(V_i \mid \mathrm{do}(V_{\mathrm{Pa}_{A^\star}(i)} = v), f_i^\star)$ is identified on the probed domain in the large-sample limit.

We can make the "observational equivalence on $\mathcal{V}_{\mathrm{probe}}$" precise by defining

$$f_i \sim_{\mathrm{probe}} f_i' \quad\Longleftrightarrow\quad \forall v \in \mathcal{V}_{\mathrm{probe}},\ \mathcal{L}(V_i \mid \mathrm{do}(V_{\mathrm{Pa}_{A^\star}(i)} = v), f_i) = \mathcal{L}(V_i \mid \mathrm{do}(V_{\mathrm{Pa}_{A^\star}(i)} = v), f_i').$$

The condition is that the map $f_i \mapsto \{\mathcal{L}(V_i \mid \mathrm{do}(V_{\mathrm{Pa}_{A^\star}(i)} = v), f_i)\}_{v\in\mathcal{V}_{\mathrm{probe}}}$ is injective modulo $\sim_{\mathrm{probe}}$, and the conclusion is that the identified response kernel pins down $f_i^\star$ up to $\sim_{\mathrm{probe}}$.

Finally, identifying the deterministic map $f_i^\star(\cdot,\cdot)$ itself (separating it from value noise) is strictly stronger: without additional restrictions on how noise enters, many different deterministic functions can induce the same conditional law. Standard sufficient conditions include noise-free mechanisms, known additive noise, monotone structural quantiles, or a known parametric mechanism family; in these cases $f_i^\star$ becomes identifiable from the collection of interventional response laws. □

D.6 TABLES

POSCM component	Retina instantiation
Node ordering τ	Laminar order: PR (layer 0) $\to$ HZ (1) $\to$ BC (2) $\to$ AC (3) $\to$ RGC (4).
Context β_i	Discrete cell type (e.g., $\{\text{rod}, \text{cone}\}$ for PRs; $\{\text{ON-BC}, \text{OFF-BC}, \text{rod-BC}\}$ for BCs; $\{\text{AII}\}$ for ACs; $\{\text{ON}, \text{OFF}\}$ for RGCs).
Structure kernel $\alpha(\beta_j, \cdot)$	Type- and distance-dependent connectivity rule (e.g., dendritic field overlap).
Mechanism operator $\Gamma(\beta_i, \text{Pa}_A(i))$	Mapping from (cell type, parent set) to synaptic transfer parameters (e.g., sign-inverting vs. sign-preserving synapses; $(g_{\max}, V_{\text{thr}}, V_{\text{slope}}, \tau)$).
Endogenous variables V_i	Membrane potential time series (mV) over a 200 ms epoch (0.1 ms step).
Exogenous noise U	Ion-channel noise, synaptic release noise, photon shot noise.

Table 6: POSCM–retina dictionary used throughout Sec. 6.

Parameter	Value	Description
Patch size	30×30 μm	Reduced for tractability
Eccentricity	-1.2 mm	Mid-peripheral retina
Simulation time	200 ms	Per trial (stimulus epoch)
Time step	0.1 ms	NEURON integration step
Stimulus	120 pA flash	Full-field; rods and cones
Seeds	2	Independent network realizations

Table 7: Shared simulation parameters (unless otherwise noted).

Theoretical result	Experiment	Key finding	Quantitative evidence
Prop. 5.2 (no β access)	Exp. 1	Type-swapped twins are indistinguishable under observational and V-node data	KS p-values 0.96–1.00 across conditions
Lemma 5.3 (latent-edge confounding)	Exp. 2	Calibrated (p, g) pairs match under node-do but diverge under edge-do	Mean MMD: node-do 0.044 vs. edge-do 0.131
Thm. 5.4 (kernel identifiability)	Exp. 3	Context sweeps modulate composition; voltage sweeps recover a transfer curve	Sigmoidal BC$\to$RGC curve with transition near -45 mV

Table 8: Theory–experiment correspondence in the virtual retina experiments.