IDENTIFYING WHAT ACTIONS READ AND CHANGE IN WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

World models support planning by predicting action consequences. Yet accurate prediction can leave latent state variables mixed across learned variables, obscuring which state variables the behavior policy reads (the read set) and which its actions change (the effect set). We jointly model how the behavior policy generates actions and how actions enter the transition, allowing next-step variables to depend on one another. We show that, under explicit conditions, video and logged actions reveal both roles of each variable that within-step dependence sets apart: its learned counterpart carries the same two labels, even while others stay mixed. The guarantee holds for models that reproduce the logged frame–action process with the fewest conditional dependencies among next-step variables. Under the same conditions, when next-step variables are conditionally independent given the current state and action, the guarantee covers every dynamic state variable. We learn one world model and estimate both sets on its frozen representation. Controlled latent experiments show closer variable matching and better read-set recovery than the same model trained for prediction alone, at comparable predictive accuracy. In a controlled RGB study on a fixed representation, read and effect estimates track policy and actuator switches, respectively. The selected variables also predict action responses more accurately than the full representation.

1 INTRODUCTION

World models support planning by predicting how a scene evolves under candidate actions (Hafner et al., 2019), for example how the cup in Figure 1a moves under different pushes. However, visual prediction quality does not consistently reflect downstream control performance (Tian et al., 2023; Qiu et al., 2026). Qiu et al. (2026) trace part of this gap to video-model features that stay more entangled with appearance than foundation-encoder features, while actions depend on a few interaction-relevant state variables. In a logged push, they can play two roles. Variables the behavior policy uses to select the push (e.g., cup position and contact force) form its *read set*. Variables whose next-step distribution depends on the push, with the whole current state held fixed (e.g., position, velocity, and contact force), form the *effect set*. Separating the two roles would let us audit which variables the logged policy relies on before imitating or evaluating it, check a model's predicted responses on the variables the push changes, and select variables for action and response prediction.

Prediction alone does not deliver this separation. A world model can keep the cup's position, velocity, and contact force mixed in its learned variables and still predict the cup's motion. Suppose one learned variable stores position plus velocity and another stores velocity (Figure 1b). Together they recover position, which the policy reads, so the second variable helps predict the logged push and looks like a policy input, although the policy never uses velocity. Likewise, if a learned variable mixes a state variable the push changes with one it does not, its response does not show which one responded. Predictive fit therefore cannot determine whether a learned variable's apparent role matches that of the underlying state variable.

Existing identifiability results (Appendix I) recover the state up to blocks or rotations, or its individual components. Control-oriented approaches organize states into action- and reward-defined blocks (Wang et al., 2022; Liu et al., 2023), and latent states can be identified up to a shared orthogonal transformation under linear Gaussian dynamics, without actions (Klindt et al., 2026) or with them (Zhang et al., 2026). Such blocks and rotations can still combine variables whose roles

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

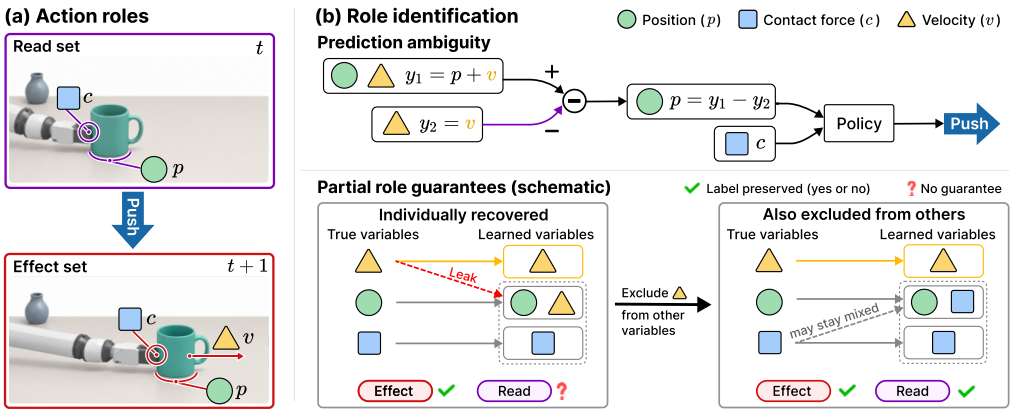

Figure 1: **The action–state interface.** **(a)** The policy reads position and contact force; the push changes all three next-step distributions. **(b)** Mixed variables still predict actions, making velocity look like a policy input. Under Theorem 1's role conditions, the lower schematics show sufficient conditions for preserving each label, including negative ones: velocity's Read check correctly excludes it from policy inputs.

differ, and reward relevance reflects an optimal policy's needs, not the behavior policy's inputs. Temporal identification recovers individual components through sparse changes or temporal dependencies, including instantaneous, or within-step, dependence (Klindt et al., 2021; Yao et al., 2022; Li et al., 2025), but the recovered components are not linked to what the policy reads or what the action changes. Sparse action effects yield partial disentanglement when variables do not affect one another within a step (Lachapelle et al., 2026), and observed intervention targets identify latent variables and their causal graph even when they do (Lippe et al., 2023). A logged action, however, does not record which variables it targets, yet each logged step carries information about both sets. Determining the roles of interest may require separating only part of the state. We therefore ask:

When do video and logged actions reveal what a policy reads and what its actions change?

To answer this, we show that, under appropriate conditions, video and logged actions can identify a state variable's two roles even while others stay mixed. Within-step dependence determines which variables this guarantee covers. For each such variable, its learned counterpart carries both labels, so one can tell whether a variable the policy reads is also one its actions change (Figure 1b). We jointly model how the behavior policy generates actions and how actions enter the transition, allowing next-step variables to depend on one another. The guarantee holds for models that reproduce the logged frame–action process with the fewest conditional dependencies among next-step variables. Under the same conditions, when next-step state variables are conditionally independent given the current state and action, the guarantee covers every dynamic state variable.[1]

We learn one world model and estimate both sets on its frozen representation, one estimator per relation (Section 4). In controlled latent processes (Section 5.1), our model surpasses the same model trained for prediction alone in variable matching and read-set recovery, at comparable predictive accuracy. In a controlled RGB study on a fixed representation with scene-specific priors (Section 5.2), read and effect estimates track policy and actuator switches, respectively, and the selected variables give action-prediction accuracy comparable to the full representation with lower action-response error. On BridgeData V2 robot logs (Walke et al., 2023), visual features from consecutive frames add small gains over robot state in decoding observed state changes.

2 Problem Formulation

From logged images and actions, we seek a world model's learned variables $(\hat{s}, \hat{d}_t)$ that preserve policy-input and next-step response roles and their overlap (Figure 1a). We call the components of $(\hat{s}, \hat{d}_t)$ coordinates.

[1] Fixed-covariance Gaussian transition noise fails the variation condition of Section 3.2.

Interaction model. Let s denote content fixed within a trajectory, such as table appearance, and d_t the dynamic state. Their known dimensions are n_s, n_d, with $[n_d] = \{1, \dots, n_d\}$. The static domain $\mathcal{S} \subseteq \mathbb{R}^{n_s}$ is open and connected, and the dynamic domain $\mathcal{D} = \prod_k \mathcal{D}_k$ is a product of open intervals. The observed image is generated by $i_t = g_I(s, d_t)$. The behavior policy generates $a_t = g_A(m^A \odot [s, d_t], \varepsilon_t^A)$. The selector m^A marks the state entries used by the policy; the private policy noise is independent of the current state. The action has a Lebesgue density that is positive on a common open support $\mathcal{A}$ for all states, allowing action comparisons at a fixed preceding state; positivity excludes deterministic policies.

The action enters the next-step distribution as $d_{t+1} \mid (s, d_t, a_t) \sim K_{s,a_t}(\cdot \mid d_t)$. Write $u = d_t$ for the current dynamic state and $x = d_{t+1}$ for the next. K is the transition kernel, and $K_{s,a}(x \mid u)$ denotes its density with respect to Lebesgue measure on $\mathcal{D}$. Next-step variables may depend on one another, allowing responses to propagate within the step; we call this within-step, or instantaneous, dependence.

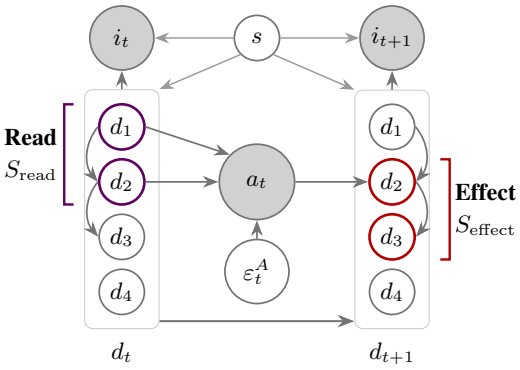

Figure 2: **Read and effect sets of the same action.** Read $\{1, 2\}$ and Effect $\{2, 3\}$ share d_2; d_3 responds through d_2. Shading marks observations; boxes group dynamic states. Arrows show the generating process, not the union CI graph. Static policy inputs are absent; transition noises are omitted. This fixed-covariance Gaussian illustration fails Assumption 2.

Definition 1 (Read and effect sets)**.** The read set contains dynamic inputs used by the behavior policy. The effect set contains state variables whose next-step marginal distributions vary with action at a fixed complete preceding state and static content:

$$\begin{aligned} S_{\text{read}} &= \{k \in [n_d] : n_s + k \in \operatorname{supp}(m^A)\}, \\ S_{\text{effect}} &= \{k \in [n_d] : \nabla_a p_k(x_k \mid s, u, a) \not\equiv 0\}. \end{aligned} \tag{1}$$

Here p_k is variable k's next-step marginal, with the other next-step variables integrated out (Appendix A.1), and $\not\equiv 0$ means that the action derivative is nonzero somewhere; Assumption 3 states the required regularity. Effect also includes indirect responses, such as that of d_3 through d_2 in Figure 2, and changes without a mean shift.

Process conditions. Given s, the process is first-order and time-homogeneous, initialized with a positive invariant density for the policy-averaged kernel. The state density is positive on $\mathcal{S} \times \mathcal{D}$. Under the joint law P_{S,D_t,A_t}, for almost every context (s, u, a), $x \mapsto K_{s,a}(x \mid u)$ is strictly positive and C^3 throughout the common domain $\mathcal{D}$.

Example (pushing a cup). Here d_t collects the cup's position p, velocity v, and contact force c, all dimensionless, and s the table's appearance, which plays neither role. The policy chooses the push from p and c with some randomness, so position and contact force form the read set; the push changes all three next-step marginals, so all three form the effect set. A model that stores $y_1 = p + v$ and $y_2 = v$ predicts as well as one that stores p and v, yet y_2 helps predict the push because $y_1 - y_2$ recovers p: velocity looks like a policy input although the policy never uses it (Figure 1b). Section 3 states which labels survive such mixing.

3 Identifying policy inputs and action responses

Under explicit conditions, some role labels survive even while other state variables stay mixed. To fix ideas, we use an idealized cup model in two versions, built in Appendix F to meet the conditions of Theorem 1. In one version, position, velocity, and contact force receive separate random jolts; in the other, position and velocity share one jolt whose split varies with where the cup is.

3.1 WHAT PRESERVES EACH ROLE LABEL

Why is read harder than effect? Whether the push changes a variable is a property of that variable's own next-step marginal, with the complete current state held fixed. Whether the policy reads a variable is not a property of that variable alone: the read test asks what it adds about the push beyond the other current variables. If another learned variable stores part of a policy input, as position-plus-velocity does (Figure 1b), the test can err in either direction. Response labels therefore tolerate more mixing elsewhere than read labels.

Formally, under the effect conditions of Section 3.2, an action-independent recoding of a state variable that is smooth with a smooth inverse responds to the push exactly when that variable does, whatever the other learned variables contain.

3.2 ASSUMPTIONS AND ESTIMATION CRITERION

The following assumptions distinguish what establishes variable correspondences from what preserves their role labels; the minimum-edge criterion separately specifies how candidates, which reproduce the logged frame–action process, are selected.

Assumption 1 (Observations and candidates)**.** The observation map g_I is a C^3 diffeomorphism onto its image $\mathcal{I}$, an embedded C^3 submanifold of the observation space, so a frame determines the modeled state.

A *candidate* has an open latent domain, the same block dimensions and logged frame–action process law, and a C^3 observation diffeomorphism onto the same $\mathcal{I}$. Its inverse encoder uses frames without actions, is shared across time, and keeps static outputs constant along trajectories. The candidate class contains the true model, and each candidate's dynamic fiber at a fixed static code is nonempty and connected (Appendix A).

Assumption 1's invertible observation model (von Kügelgen et al., 2021) rules out occlusion and velocities that a single frame cannot show; the cup model idealizes a frame as showing position, velocity, and contact force. Together with the interaction model, it identifies static content up to invertible recoding and dynamic variables up to an invertible, possibly mixed recoding that may depend on static content (Lemma 1, Appendix A.2).

Union conditional-independence graph. At a fixed time t, write the interaction variables in capitals: S, D_t, A_t, and D_{t+1}. The *union graph* G on $[n_d]$ records which next-step variables stay dependent once everything else in the step is known: its edge set $\mathcal{E}(G)$ omits $\{i, j\}$ exactly when $D_{t+1,i} \perp D_{t+1,j} \mid (D_{t+1,-\{i,j\}}, D_t, S, A_t)$, where $D_{t+1,-\{i,j\}}$ collects the other next-step variables. A non-edge holds at almost every context (s, u, a); an edge is present once the dependence appears on a positive-measure set of contexts, hence the name (Appendix B). A candidate's graph $\hat{G}$ is defined in the same way from its learned variables and its own process law.

Assumption 2 (Distributional variation)**.** Write $\ell(x \mid s, u, a) = \log K_{s,a}(x \mid u)$, with subscripts denoting derivatives in the next state x. Let $v_{s,x}(u, a) \in \mathbb{R}^q$, $q = 2n_d + |\mathcal{E}(G)|$, collect $\ell_{x_i x_i}$ and ℓ_{x_i} for all $i \in [n_d]$ and $\ell_{x_i x_j}$ over the edges $\{i, j\} \in \mathcal{E}(G)$, evaluated at (x, s, u, a) and ordered as in Eq. (B.2). There is a candidate-independent measurable set $\mathcal{S}_0 \subseteq \mathcal{S}$ with $P_S(\mathcal{S}_0) = 1$, where P_S is the static-state law, such that for every $s \in \mathcal{S}_0$, $x \in \mathcal{D}$, $c \in \mathbb{R}^q$, and $\kappa \in \mathbb{R}$,

$$c^\mathsf{T} v_{s,x}(u, a) = \kappa \text{ for Lebesgue-almost every } (u, a) \in \mathcal{D} \times \mathcal{A} \implies c = 0. \tag{2}$$

The condition asks that the next-step randomness change across preceding states and actions in enough independent ways and, for Gaussian noise, not only in its center. State-dependent variance alone need not suffice: a velocity jolt entering position with a fixed share, as fixed-step semi-implicit Euler integration gives, fails it whatever the variances do. Friction or contact can change how a jolt is shared between variables, which motivates state-dependent coupling; both cup versions provably meet the condition (Appendix F.3).

Estimation criterion (Fewest union-graph edges). At the population level, select from the class in Assumption 1 the candidates whose union graph $\hat{G}$ has the fewest edges.

Why fewest dependencies? A missing edge in a candidate's graph is a conditional independence between two learned variables, which must hold for almost every preceding state and action. The encoder sees only the frame it encodes, so how it combines true variables into learned ones does not change with the preceding dynamic state or the action. Under Assumption 2 the next-step randomness does change with them, in enough independent ways, so the learned pair can stay independent only if no pair of true variables joined by an edge is split across them. Since each true variable keeps a learned variable of its own, every candidate's graph has at least as many edges as the true graph. The true model attains this bound, so each selected candidate's graph matches the true graph edge for edge, and the candidate can mix variables only where the true graph already has an edge (Lemma 2, Appendix B). With separate jolts on position and velocity, the candidate of Section 2 that stores their sum has an edge the true graph lacks and is discarded; with one shared jolt, the same candidate adds no edge and is retained (Appendix F.5).

The variation condition follows the sufficient-change approach of Hyvärinen et al. (2019) and Zhang et al. (2024), and we adapt the latter's graph argument to logged actions. In a temporal process the choice of graph matters. The union graph counts only the within-step dependencies among next-step variables, and under Assumption 2 mixing cannot lower that count. A state-pair graph also counts lagged dependencies, which mixing can remove, so minimizing it can select only mixed encoders even when Assumption 2 holds (Appendix G). Partial guarantees still rest on the full condition.

Assumption 3 (Role conditions)**.** *(i) Read-faithfulness.* Each selected dynamic input carries action information that the rest of the state does not already supply:

$$I(A_t; D_{t,k} \mid S, D_{t,-k}) > 0 \qquad (k \in S_{\text{read}}), \tag{3}$$

where I is conditional mutual information and $D_{t,-k}$ collects the other current dynamic variables.

(ii) Effect regularity. For every $k \in [n_d]$ and every $(s, u, x_k) \in \mathcal{S} \times \mathcal{D} \times \mathcal{D}_k$, the map $a \mapsto p_k(x_k \mid s, u, a)$ is continuously differentiable on $\mathcal{A}$, where p_k is the true next-step marginal defined in Section 2.

For variables the policy does not use, this information is zero, since the policy depends only on its inputs and private noise. Without read-faithfulness, the read criterion can miss an input that the policy uses but that carries no conditional action information (cf. causal minimality, Peters et al., 2014; here the conditioning is on the rest of the state).

3.3 Partial identification of policy inputs and action responses

The missing edges do not exclude all mixing. A neighbor j is *intimate* to i if it also connects to every other neighbor of i; write $\Psi_G(i)$ for this set. In reconstructing a true variable from learned ones, apart from its own local match, only matches of intimate neighbors remain unexcluded (Appendix B.5); this is a restriction on reconstruction, not a construction of every remaining mixture.

Two properties yield different guarantees. If i has no intimate neighbor, it is matched to a learned variable by an invertible scalar transformation, so its counterpart is clean; the response guarantee uses this property, $U_\Psi = \{i : \Psi_G(i) = \varnothing\}$. If it also belongs to nobody else's intimate set, no other learned variable depends on it, so it does not leak into the variables the read test conditions on; the read guarantee uses both properties, $U_{\text{sep}} = \{i \in U_\Psi : \forall r \neq i,\ i \notin \Psi_G(r)\}$.

The candidate's read criterion applies Eq. (3) to the learned variables at time t: $\hat{S}_{\text{read}} = \{j : I(A_t; \hat{D}_{t,j} \mid \hat{S}, \hat{D}_{t,-j}) > 0\}$. Its response criterion uses the candidate's next-step marginals $\hat{p}_j(\hat{x}_j \mid \hat{s}, \hat{u}, a)$ on the guaranteed learned variables, with $\hat{u}$ the complete preceding dynamic code and $\hat{x}_j$ the candidate's jth learned next-state variable; their conditional versions inherit action differentiability from the true marginals (Appendix C.3).

Theorem 1 (Partial role identification)**.** *Under the model in Section 2 and Assumptions 1 and 2, each candidate selected by the minimum-edge criterion admits an injective map $\iota : U_\Psi \to [n_d]$, fixed across static values. For this same map, the first equality below holds under the effect part of Assumption 3, and the second under its read part:*

$$\begin{aligned} \{j \in \iota(U_\Psi) : \nabla_a \hat{p}_j(\hat{x}_j \mid \hat{s}, \hat{u}, a) \not\equiv 0\} &= \iota(S_{\text{effect}} \cap U_\Psi), \\ \hat{S}_{\text{read}} \cap \iota(U_{\text{sep}}) &= \iota(S_{\text{read}} \cap U_{\text{sep}}). \end{aligned} \tag{4}$$

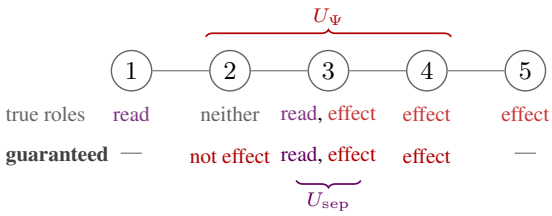

Figure 3: **Five-chain guarantees.** The lower row lists guarantees under Theorem 1's conditions; unlisted labels are not covered. Edges show union conditional dependence.

Labels that survive mixing elsewhere. When position and velocity share one random jolt, the two are linked and may stay mixed in a selected model, while the contact force is linked to neither. Under the theorem's conditions, every selected model, including one that mixes position and velocity, therefore carries both of the contact force's labels correctly (Appendix F.5).

When next-step variables are conditionally independent given the current state and action, as with separate jolts, no variable has a neighbor and, under the theorem's conditions, every label is guaranteed, whether yes or no. In the five-chain of Figure 3, each interior variable's two neighbors are unlinked, so the three interior variables keep clean counterparts and guaranteed response labels (negative for variable 2). Variables 2 and 4 may leak into the local matches of 1 and 5, so only the center variable's read label is guaranteed as well. A stationary five-chain process satisfies the theorem's premises, with its identity encoder attaining the edge minimum (Appendix E).

Applying the same neighbor rule to a selected candidate's own union graph returns exactly $\iota(U_\Psi)$ and $\iota(U_{\text{sep}})$ (Remark 1, Appendix C), so the guaranteed variables can in principle be read off the selected model itself. Our estimators do not compute this graph. The ranges mark where the guarantee applies, not where identification fails: other labels, and labels in processes that fail Assumption 2, may still be identifiable.

Section 4 estimates both sets on one frozen representation, and Section 5.1 asks how well they are recovered when the truth is known.

4 Estimating the Action–State Interface

From time-paired observations and logged actions, we learn a world model's variables, freeze them, and estimate the read and effect sets separately. Three choices follow from Section 3. A read label is tested through the conditional action law given the rest of the state, whereas an effect label is a property of one variable's next-step marginal given the complete preceding state (Section 3.1), so each set has its own estimator. An effect need not move the mean (Definition 1), so the effect estimator probes the next-step distribution beyond its mean. Theorem 1 states for which state variables the two tests of Section 3.3, applied to a learned dynamic variable, return the true labels once a candidate is selected by the minimum-edge criterion. Training constrains gates, not graph edges, and no step selects the candidate with the fewest edges, so we run both tests on the frozen representation as an *interface audit* rather than reading labels from the trained heads and gates. We describe the controlled-study implementation; pixel front ends and diagnostics appear with their experiments (Appendix J).

4.1 Learning a shared state representation

An encoder produces static and dynamic blocks $\hat{z}_\tau = (\hat{s}_\tau, \hat{d}_\tau)$. A nonlinear read head predicts the action from the dynamic block weighted by a read gate. To predict the next dynamic state, we add a transition base, taking the complete current representation, to an effect-gated linear action branch. This branch receives the action minus a stop-gradient copy of the read prediction. A decoder reconstructs observations and maps predicted states back to observation space.

We minimize the objective $\mathcal{J} = \mathcal{J}_{\text{fit}} + \mathcal{J}_{\text{act}} + \mathcal{J}_{\text{stat}} + \mathcal{J}_{\text{reg}}$: fit terms (reconstruction, transition negative log-likelihood, next-observation prediction), action terms (action-mean regression, contrastive alignment of the gated dynamic block and the action), static consistency, and regularizers (variance, covariance, an ℓ_1 penalty on both gates). Action terms train the read gate, transition and

next-observation terms the effect gate (modules, gradient paths, and coefficients in Appendices J.1 and J.2).

4.2 Estimating the read and effect sets

Read set. The read audit measures the action information a variable adds given the remaining current coordinates, testing the conditional action law beyond the mean fitted by the training head. For each of the n_s+n_d coordinates we estimate the conditional mutual information $\hat{r}_j = \hat{I}(a_\tau; \hat{z}_{\tau,j} \mid \hat{z}_{\tau,-j})$ by the Frenzel–Pompe conditional variant of KSG (Kraskov et al., 2004; Frenzel & Pompe, 2007) and set $\hat{S}^{\text{est}}_{\text{read}} = \{j : \hat{r}_j > \delta_r\}$ (settings in Appendix J.3); static coordinates are scored as well, beyond the theorem's range.

Effect set. Action and next-step variables can be associated through their shared dependence on the preceding state. To adjust for this, we residualize their features against features of the complete current representation. First and second moments and Fourier features probe distributional changes beyond the mean. As in Definition 1, other next-step variables remain marginalized, retaining indirect responses in the target.

Two-fold ridge regressions predict features of the action, $\Phi(a_{\tau-1})$, and of each next-step coordinate, $\Psi(\hat{z}_{\tau,j})$, from features of the complete current representation $\hat{z}_{\tau-1}$, and each residual is evaluated out of fold (Chernozhukov et al., 2018). The score $\hat{e}_j$ is the largest absolute cross-correlation between the standardized action and next-step residuals over feature pairs, and $\hat{S}^{\text{est}}_{\text{effect}} = \{j : \hat{e}_j > \delta_e\}$, which may share coordinates with $\hat{S}^{\text{est}}_{\text{read}}$. It is a residual-correlation statistic, not a general conditional-independence test; Appendix J.3 gives the formulas and their scope and Appendix K.1 the calibration of δ_e.

5 Experiments

We first ask, in a latent process with known truth, whether the estimates recover the read and effect sets and whether predictive error alone would have shown the same (§5.1); we then ask whether the two sets follow policy and actuator changes in visual observations and what the selected coordinates retain (§5.2).

5.1 Recovering read and effect sets in a controlled process

We simulate five dynamic state variables and one static variable; the policy reads d_0, d_2, and actions affect next-state d_1, d_3. Eight learned methods are evaluated with frozen encoders and common post-hoc readouts under lag-only and instantaneous dynamics (Appendix K.1). Probe MSE measures predictive fit; under full-space matching, MCC measures mean absolute correlation with true dynamic variables, and Read/Effect F_1 measures set membership. Both processes have fixed innovation covariances and fail Assumption 2; they test the estimators against known roles, not the theorem's guarantee.

Recovery. Our model achieves the highest dynamic MCC and Read F_1 among the eight learned methods in both settings, with probe MSE comparable to that of the temporal backbone, which uses the same architecture without interface objectives (Table 1). Read and Effect F_1 also exceed those of IDOL+A and Mechanism Sparsity in both settings. The Read advantage over the backbone extends to three more instantaneous topologies (Appendix Table 10). In the matched ablation, mean Read F_1 is similar with and without the read objective (Appendix Table 12).

Coordinate mixing. Orthogonally mixing the true dynamic variables leaves refitted probe MSE essentially unchanged (relative changes below 3.3×10^{-9}) while reducing both role F_1 scores by .390 on average (Figure 4a).

Action responses. Our model's responses are more concentrated on the true effect variables than the backbone's (Figure 4b). The encoded next-state means respond strongly on d_1, d_3 and more weakly on d_0, d_2 than the backbone's, with a residual response on d_4.

Table 1: **Predictive fit and recovery.** Five-seed means with sample SDs in subscripts. Blocks: baselines; our model and same-architecture controls; oracle evaluation ceiling. Bold/underline: best/second-best unrounded means among the eight learned methods per column. Absolute MSE is comparable only within a DGP. IDOL+A is our action-conditioned IDOL variant.

	Lag-only dynamics				Instantaneous dynamics			
Method	MSE ↓	MCC ↑	Read F_1 ↑	Effect F_1 ↑	MSE ↓	MCC ↑	Read F_1 ↑	Effect F_1 ↑
TDRL	$.643_{\pm.102}$	$.651_{\pm.092}$	$.622_{\pm.116}$	$.548_{\pm.073}$	$.647_{\pm.088}$	$.660_{\pm.094}$	$\underline{.667}_{\pm.000}$	$.574_{\pm.130}$
IDOL	$.643_{\pm.095}$	$.696_{\pm.043}$	$\underline{.687}_{\pm.124}$	$.500_{\pm.000}$	$.647_{\pm.086}$	$\underline{.684}_{\pm.047}$	$.556_{\pm.097}$	$.533_{\pm.075}$
IDOL+A	$.642_{\pm.095}$	$.692_{\pm.047}$	$.641_{\pm.113}$	$.500_{\pm.000}$	$.647_{\pm.086}$	$.682_{\pm.046}$	$.529_{\pm.124}$	$.533_{\pm.075}$
Mechanism Sparsity	$.635_{\pm.090}$	$.658_{\pm.098}$	$.608_{\pm.127}$	$.529_{\pm.039}$	$.642_{\pm.082}$	$.655_{\pm.068}$	$.561_{\pm.114}$	$.529_{\pm.039}$
SlowVAE	$1.004_{\pm.215}$	$.141_{\pm.037}$	$.000_{\pm.000}$	$.500_{\pm.000}$	$.985_{\pm.194}$	$.151_{\pm.024}$	$.000_{\pm.000}$	$.500_{\pm.000}$
Temporal backbone	$\mathbf{.633}_{\pm.090}$	$\underline{.708}_{\pm.045}$	$.655_{\pm.094}$	$.548_{\pm.073}$	$\mathbf{.638}_{\pm.080}$	$.678_{\pm.053}$	$.600_{\pm.091}$	$.560_{\pm.134}$
Static alignment	$.678_{\pm.123}$	$.525_{\pm.041}$	$.629_{\pm.052}$	$\underline{.603}_{\pm.114}$	$.676_{\pm.104}$	$.541_{\pm.058}$	$.583_{\pm.142}$	$\mathbf{.641}_{\pm.113}$
Ours	$\underline{.635}_{\pm.089}$	$\mathbf{.754}_{\pm.084}$	$\mathbf{.853}_{\pm.145}$	$\mathbf{.636}_{\pm.100}$	$\underline{.639}_{\pm.078}$	$\mathbf{.743}_{\pm.029}$	$\mathbf{.960}_{\pm.089}$	$\underline{.595}_{\pm.071}$
Oracle coordinates	$.632_{\pm.090}$	$1.000_{\pm.000}$	$1.000_{\pm.000}$	$1.000_{\pm.000}$	$.636_{\pm.078}$	$1.000_{\pm.000}$	$1.000_{\pm.000}$	$.960_{\pm.089}$

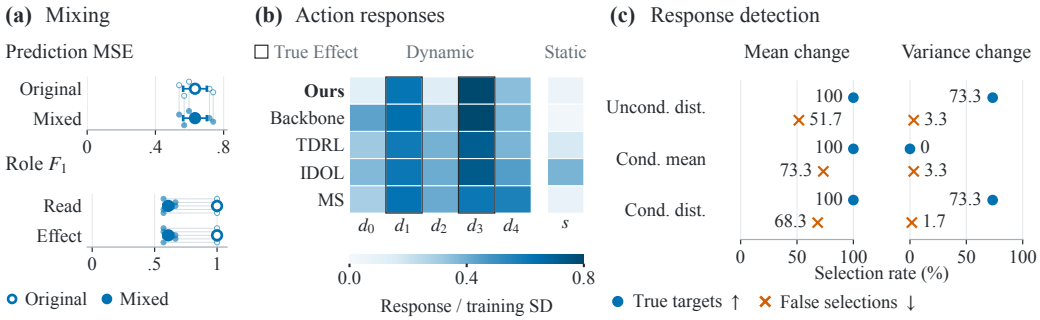

Figure 4: **Coordinate mixing and action responses.** (a) Prediction and set recovery under orthogonal mixing of the true dynamic variables. (b) RMS responses of encoded next-state means to action changes with the true current state fixed and noise paired. (c) Selection rates on our model's coordinates. Five-seed means; whiskers in (a) are 95% paired-bootstrap intervals; MS denotes Mechanism Sparsity; protocols and no-effect selection rates in Appendix K.3.

Distributional changes. When actions change only the variance of the true variables, the conditional-distribution statistic detects responses missed by the conditional-mean statistic (Figure 4c). Under mean changes, all three statistics detect the targets but also select many non-targets.

5.2 Visual role separation and downstream utility

We study separate and joint policy and actuator changes in a controlled RGB scene, using a common fixed representation with scene, color, and single-axis-motion priors (learned representations trained on this scene did not match its six state variables; Appendix L.1); the read and effect sets are re-estimated in each of the four settings. We also evaluate learned representations on BridgeData V2 through paired-frame decoding of observed state changes, using current proprioception and visual features at t and $t+1$ (Appendix M).

Role separation. Read selection follows the policy switch from object A to C, while Effect selection follows the actuator switch from B to A (Figure 5a). In the actuator-only setting, the read and effect sets overlap on object A. Our conditional-association estimator reaches Read/Effect F_1 of $.967/.950$.

Set utility. Using the same prediction-head families, estimated sets retain action-prediction accuracy comparable to all coordinates and yield lower error on short-term responses to action perturbations, particularly with fewer head-fitting windows (Figure 5b). Size-matched random subsets perform worse on both tasks, while Group Lasso has higher response error.

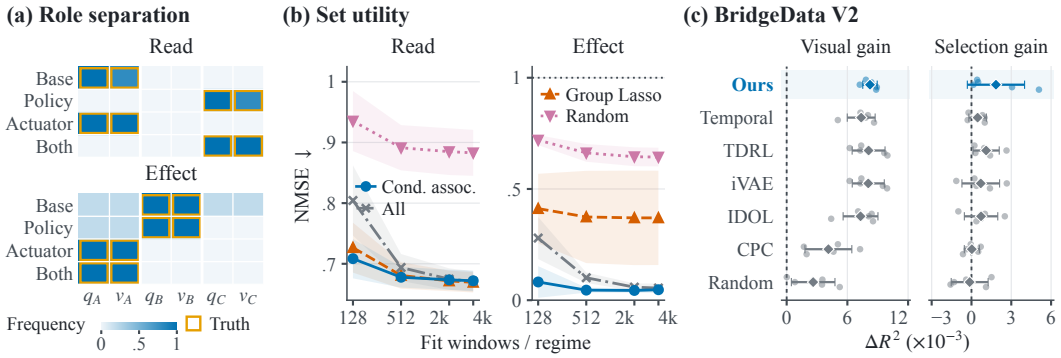

Figure 5: **Action-role separation and coordinate utility.** (a) Selection frequencies in four mechanism settings (q, v: position and velocity of objects A–C). (b) Action-prediction (Read) and action-response (Effect) NMSE versus head-fitting windows (Cond. assoc.: our estimator); the dotted line marks zero-response error. (c) BridgeData V2: visual gain, incremental R^2 over state-only decoding; selection gain, selected minus size-matched random-coordinate R^2; baselines are local adaptations (Appendix M). Five seeds; dots runs, curves and diamonds means, bars and shading one SD; (b) averages settings within seeds.

Robot-log decoding. Visual features provide small gains in state-change decoding over the state-only baseline across the compared representations (Figure 5c). Our model's selected coordinates also yield higher decoding accuracy than equal-size random subsets of the same representation.

5.3 What this means for evaluating world models

Three findings follow for evaluating a world model's variables. First, predictive accuracy does not measure role recovery: methods with comparable probe MSE differ in Read F_1 (Table 1), and mixing the true variables leaves probe MSE unchanged while lowering both role scores (Figure 4a). Second, with the representation and its coordinate correspondence held fixed, the read set follows the policy's inputs and the effect set the actuator's targets (Figure 5a), so each change in membership can be traced to one state variable. Third, the selected variables serve action and response prediction: in the fixed RGB representation, a few selected coordinates give action-prediction accuracy comparable to the full representation with lower action-response error (Figure 5b).

6 Conclusion

Under the stated conditions, video and logged actions identify both roles of each state variable that within-step dependence sets apart (whether the behavior policy reads it and whether its actions change it), even while the remaining variables stay mixed. Such a variable's roles can thus be judged without first separating the rest of the state. Both labels attach to a single learned variable, so one can tell whether the policy reads the same state variable its actions change. Predictive accuracy does not settle these labels: in the controlled latent studies, mixing the true variables left prediction error unchanged but lowered read- and effect-set recovery. The identification results are population statements, and the estimators take these conditions as premises rather than verifying them; Appendix N states the scope of results, estimators, and studies.

Two boundaries of these results point to future work. First, the guarantees assume synchronized frames and actions; identifying unknown action delays would be one step toward raw robot recordings, where such delays occur. Second, our guarantees do not cover pretrained world-model encoders; with a label-free threshold calibration, our audit could describe which of their variables track the behavior policy's inputs and the one-step effects of its actions. Knowing which variables an action changes within one step could then focus action-conditioned evaluation of world models on those variables. Whether the resulting labels help planning or control remains untested; the audit supplies the variables and labels that a closed-loop study would compare.

AI USE STATEMENT

We used generative AI tools to assist with drafting and polishing the manuscript, including improving clarity, grammar, and organization. We also used these tools for literature retrieval and discovery, including finding potentially relevant work. The authors take responsibility for the final manuscript, including its wording, citations, technical claims, and reported results.

ETHICS STATEMENT

This work primarily studies the theoretical foundations of identifying policy inputs and action responses in world models. We do not identify ethical concerns specific to this study beyond those generally associated with research on world models.

REPRODUCIBILITY STATEMENT

Sections 2 and 3.2 state the model and assumptions, with proofs and constructions in Appendices A–H. Appendix J describes the implementation and training settings. Experimental setups and evaluation protocols are provided in Appendices K–M.

REFERENCES

Jake Bruce, Michael Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, Yusuf Aytar, Sarah Bechtle, Feryal Behbahani, Stephanie Chan, Nicolas Heess, Lucy Gonzalez, Simon Osindero, Sherjil Ozair, Scott Reed, Jingwei Zhang, Konrad Zolna, Jeff Clune, Nando de Freitas, Satinder Singh, and Tim Rocktäschel. Genie: Generative interactive environments. In *International Conference on Machine Learning (ICML)*, 2024. Best Paper; arXiv:2402.15391.

Baiming Chen, Mengdi Xu, Liang Li, and Ding Zhao. Delay-aware model-based reinforcement learning for continuous control. *Neurocomputing*, 450:119–128, 2021. doi: 10.1016/j.neucom.2021.04.015. URL https://doi.org/10.1016/j.neucom.2021.04.015. arXiv:2005.05440.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Imant Daunhawer, Alice Bizeul, Emanuele Palumbo, Alexander Marx, and Julia E. Vogt. Identifiability results for multimodal contrastive learning. In *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2303.09166.

Stefan Frenzel and Bernd Pompe. Partial mutual information for coupling analysis of multivariate time series. *Physical Review Letters*, 99(20):204101, 2007.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2555–2565. PMLR, 2019. URL https://proceedings.mlr.press/v97/hafner19a.html. arXiv:1811.04551.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640:647–653, 2025. doi: 10.1038/s41586-025-08744-2. URL https://www.nature.com/articles/s41586-025-08744-2. arXiv:2301.04104.

Aapo Hyvärinen and Hiroshi Morioka. Unsupervised feature extraction by time-contrastive learning and nonlinear ICA. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2016. arXiv:1605.06336.

Aapo Hyvärinen, Hiroaki Sasaki, and Richard E. Turner. Nonlinear ICA using auxiliary variables and generalized contrastive learning. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019. arXiv:1805.08651.

Petr Ivashkov, Randall Balestriero, and Bernhard Schölkopf. Sensorimotor world models: Perception for action via inverse dynamics. *arXiv preprint arXiv:2606.20104*, 2026. doi: 10.48550/arXiv.2606.20104. URL https://arxiv.org/abs/2606.20104.

Ilyes Khemakhem, Diederik P. Kingma, Ricardo Pio Monti, and Aapo Hyvärinen. Variational autoencoders and nonlinear ICA: A unifying framework. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020. arXiv:1907.04809.

David Klindt, Yann LeCun, and Randall Balestriero. When does LeJEPA learn a world model?, 2026. URL https://arxiv.org/abs/2605.26379.

David A. Klindt, Lukas Schott, Yash Sharma, Ivan Ustyuzhaninov, Wieland Brendel, Matthias Bethge, and Dylan M. Paiton. Towards nonlinear disentanglement in natural data with temporal sparse coding. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2007.10930. arXiv:2007.10930.

Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.

Sébastien Lachapelle. On the identifiability of latent action policies. In *NeurIPS 2025 Workshop on Unifying Representations in Neural Models (UniReps)*, 2025. arXiv:2510.01337.

Sébastien Lachapelle, Pau Rodríguez López, Yash Sharma, Katie Everett, Rémi Le Priol, Alexandre Lacoste, and Simon Lacoste-Julien. Nonparametric partial disentanglement via mechanism sparsity: Sparse actions, interventions and sparse temporal dependencies. *Journal of Machine Learning Research*, 27(71), 2026. URL https://www.jmlr.org/papers/volume27/24-0771/24-0771.pdf. Extends the CLeaR 2022 paper; arXiv:2401.04890.

Steffen L. Lauritzen. *Graphical Models*, volume 17 of *Oxford Statistical Science Series*. Oxford University Press, 1996.

Zijian Li, Yifan Shen, Kaitao Zheng, Ruichu Cai, Xiangchen Song, Mingming Gong, Guangyi Chen, and Kun Zhang. On the identification of temporal causal representation with instantaneous dependence. In *International Conference on Learning Representations (ICLR)*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/5726513facc85f802be4a25e77fb9765-Paper-Conference.pdf.

Zijian Li, Yifan Shen, Kaitao Zheng, Ruichu Cai, Xiangchen Song, Mingming Gong, Guangyi Chen, and Kun Zhang. On the identification of temporally causal representation with instantaneous dependence. *arXiv preprint arXiv:2405.15325*, 2026. URL https://arxiv.org/abs/2405.15325v4. Version 4, January 18, 2026.

Juan K. Lin. Factorizing multivariate function classes. In *Advances in Neural Information Processing Systems 10 (NIPS)*, pp. 563–569, 1997.

Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Efstratios Gavves. CITRIS: Causal identifiability from temporal intervened sequences. In *International Conference on Machine Learning (ICML)*, 2022. PMLR 162; arXiv:2202.03169.

Phillip Lippe, Sara Magliacane, Sindy Löwe, Yuki M. Asano, Taco Cohen, and Efstratios Gavves. Causal representation learning for instantaneous and temporal effects in interactive systems. In *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2206.06169.

Yuren Liu, Biwei Huang, Zhengmao Zhu, Honglong Tian, Mingming Gong, Yang Yu, and Kun Zhang. Learning world models with identifiable factorization. In *Advances in Neural Information Processing Systems*, volume 36, pp. 31831–31864. Curran Associates, Inc., 2023. doi: 10.52202/075280-1381. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/65496a4902252d301cdf219339bfbf9e-Abstract-Conference.html.

Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2 edition, 2009.

Alexander Nikulin, Ilya Zisman, Denis Tarasov, Nikita Lyubaykin, Andrei Polubarov, Igor Kiselev, and Vladislav Kurenkov. Latent action learning requires supervision in the presence of distractors. In *International Conference on Machine Learning (ICML)*, 2025. arXiv:2502.00379.

Jonas Peters, Joris M. Mooij, Dominik Janzing, and Bernhard Schölkopf. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15(58):2009–2053, 2014. URL https://jmlr.org/papers/v15/peters14a.html.

Lu Qiu, Yizhuo Li, Yi Chen, Yuying Ge, Yixiao Ge, and Xihui Liu. Making foresight actionable: Repurposing representation alignment in world action models. *arXiv preprint arXiv:2606.12217*, 2026. URL https://arxiv.org/abs/2606.12217. Version 1.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning (ICML)*, 2021. arXiv:2103.00020.

Dominik Schmidt and Minqi Jiang. Learning to act without actions. In *International Conference on Learning Representations (ICLR)*, 2024. Spotlight; arXiv:2312.10812.

Stephen Tian, Chelsea Finn, and Jiajun Wu. A control-centric benchmark for video prediction. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://jiajunwu.com/papers/vp2_iclr.pdf. arXiv:2304.13723.

Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, and Francesco Locatello. Self-supervised learning with data augmentations provably isolates content from style. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. arXiv:2106.04619.

Homer Walke, Kevin Black, Abraham Lee, Moo Jin Kim, Max Du, Chongyi Zheng, Tony Zhao, Philippe Hansen-Estruch, Quan Vuong, Andre He, Vivek Myers, Kuan Fang, Chelsea Finn, and Sergey Levine. BridgeData V2: A dataset for robot learning at scale. In *Conference on Robot Learning (CoRL)*, 2023. URL https://arxiv.org/abs/2308.12952. arXiv:2308.12952.

Lirui Wang, Kevin Zhao, Chaoqi Liu, and Xinlei Chen. Learning real-world action-video dynamics with heterogeneous masked autoregression. *arXiv preprint arXiv:2502.04296*, 2025.

Tongzhou Wang, Simon Du, Antonio Torralba, Phillip Isola, Amy Zhang, and Yuandong Tian. Denoised MDPs: Learning world models better than the world itself. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 22591–22612. PMLR, 2022. URL https://proceedings.mlr.press/v162/wang22c.html.

Zizhao Wang, Chang Shi, Jiaheng Hu, Kevin Rohling, Roberto Martín-Martín, Amy Zhang, and Peter Stone. Factored latent action world models. In *International Conference on Machine Learning (ICML)*, 2026. PMLR 306; arXiv:2602.16229.

John Wikman, Alexandre Proutiere, and David Broman. Adaptive reinforcement learning for unobservable random delays. In *Proceedings of the 43rd International Conference on Machine Learning*, volume 306 of *Proceedings of Machine Learning Research*. PMLR, 2026. URL https://arxiv.org/abs/2506.14411. arXiv:2506.14411v2.

Shaoan Xie, Lingjing Kong, Yujia Zheng, Yu Yao, Zeyu Tang, Eric P. Xing, Guangyi Chen, and Kun Zhang. SmartCLIP: Modular vision-language alignment with identification guarantees. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Dingling Yao, Danru Xu, Sébastien Lachapelle, Sara Magliacane, Perouz Taslakian, Georg Martius, Julius von Kügelgen, and Francesco Locatello. Multi-view causal representation learning with partial observability. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2311.04056.

Weiran Yao, Guangyi Chen, and Kun Zhang. Temporally disentangled representation learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. arXiv:2210.13647.

Seonghyeon Ye, Joel Jang, Byeongguk Jeon, Sejune Joo, Jianwei Yang, Baolin Peng, Ajay Mandlekar, Reuben Tan, Yu-Wei Chao, Yuchen Lin, Lars Liden, Kimin Lee, Jianfeng Gao, Luke Zettlemoyer, Dieter Fox, and Minjoon Seo. Latent action pretraining from videos. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2410.11758.

Chuheng Zhang, Tim Pearce, Pushi Zhang, Kaixin Wang, Xiaoyu Chen, Wei Shen, Li Zhao, and Jiang Bian. What do latent action models actually learn? In *Advances in Neural Information Processing Systems*, volume 38, pp. 146676–146697, 2025. doi: 10.52202/085713-4906. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/d7b3cef7c31b94a4a533db83d01a8882-Abstract-Conference.html. arXiv:2506.15691.

Kun Zhang, Shaoan Xie, Ignavier Ng, and Yujia Zheng. Causal representation learning from multiple distributions: A general setting. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 60057–60075, 2024. URL https://proceedings.mlr.press/v235/zhang24br.html.

Xiangteng Zhang, Yang Guan, Bo Zhang, Hongyang Li, Ya-Qin Zhang, and Shengbo Eben Li. On the identifiability of controlled world models. *arXiv preprint arXiv:2607.22430*, 2026. doi: 10.48550/arXiv.2607.22430. URL https://arxiv.org/abs/2607.22430. Version 2.

Chuning Zhu, Raymond Yu, Siyuan Feng, Benjamin Burchfiel, Paarth Shah, and Abhishek Gupta. Unified world models: Coupling video and action diffusion for pretraining on large robotic datasets. In *Proceedings of Robotics: Science and Systems*, 2025. doi: 10.15607/RSS.2025.XXI.015. URL https://www.roboticsproceedings.org/rss21/p015.html. arXiv:2504.02792.

Roland S. Zimmermann, Yash Sharma, Steffen Schneider, Matthias Bethge, and Wieland Brendel. Contrastive learning inverts the data generating process. In *International Conference on Machine Learning (ICML)*, 2021. arXiv:2102.08850.

Appendix Table of Contents

A Model identities and static correspondence **14**

B Union graph and local mixing restrictions **16**

C Partial correspondence and role preservation **19**

D Shared certificates for the constructions **21**

E A five-chain process satisfying the conditions **22**

F The cup model in two versions **24**

G Why state-pair edge minimality is insufficient **27**

H Scope of full correspondence and a four-cycle example **29**

I Related Work **32**

J Method and audit implementation details **33**

K Additional Controlled Numerical Results **37**

L Physics-Simulation Details **43**

M BridgeData V2: paired-frame decoding **44**

N Scope of the results, the estimators, and the studies **44**

A Model identities and static correspondence

Machine-checked proofs. Lean 4 with Mathlib checks three groups of results under the manuscript's model, assumptions, and minimum-edge criterion, with its conditional versions and almost-everywhere quantifiers: the correspondence and role-preservation chain, its full-correspondence case, and the read-sufficiency and conditional-KL results in Appendix J.5; Proposition 1 and the moral-graph argument in Appendix H.1; and the three constructions' stationarity, variation, role sets, and edge minima. The shared lemmas in Appendix D are checked on these instances; the cup model of Appendix F reuses them and is checked by hand only. All checked statements use only Lean's standard axioms (propositional extensionality, choice, and quotient soundness), with no placeholders. The correspondence to the manuscript was checked by hand; compilation certifies the formal statements.

Where to look. Appendices A–C prove Theorem 1 in three steps: static correspondence (Lemma 1), graph identification with its local mixing restriction (Lemma 2), and a common matching on the guarantee ranges that preserves both roles (Lemma 3, Appendices C.2 and C.3). The table below locates other checks.

To check	Go to
Why the read guarantee covers U_{sep} and the effect guarantee all of U_Ψ	Appendices C.1–C.3
What each cup version guarantees, and when the mixed encoder is retained	Appendix F (Table 2)
Which five-chain labels are guaranteed, including a negative one	Appendix E
How full correspondence can hold despite within-step dependence	Appendix H
Why minimizing state-pair edges can select only mixed encoders	Appendix G (Table 3)
How the constructions verify stationarity and the variation condition	Appendix D
How the method is implemented and instantiated in each study	Appendices J.1–J.4; Tables 5–6
What an ideal read criterion selects under full correspondence	Appendix J.5
How the comparison in Table 1 is set up, and its complete results	Appendices K.1 and K.2
How the three panels of Figure 4 are defined	Appendix K.3
Which further topologies are tested, and how the read objective is ablated	Appendix K.4
Which Effect errors a larger audit sample removes, and which errors remain on aligned learned coordinates	Appendix K.5
How the fixed RGB study and paired-frame decoding of Figure 5 are run	Appendices L.1 and M
What the theory, the estimators, and the studies each cover	Appendix N

Notation. The following symbols recur across appendices; auxiliary matrices such as B, C, and J are defined where used.

Symbol	Meaning	Defined
$\mu = P_S$, ρ_s	static law; invariant dynamic density given s	Section 2
$K_{s,a}(x \mid u)$, $\bar{K}_s(x \mid u)$	transition kernel version; its policy average	Section 2; Eq. (A.1)
u, x; y	current and next true dynamic states; a candidate's next-state code $y = h_s(x)$	Section 2; Lemma 2
ℓ, ℓ_{x_i}, $\ell_{x_i x_j}$	$\log K_{s,a}(x \mid u)$ and its next-state derivatives	Assumption 2
$v_{s,x}(u,a)$, q	variation vector at fixed (s,x); its length $2n_d + \lvert\mathcal{E}(G)\rvert$	Eq. (B.2)
G, $\hat{G}$; $N_G(i)$, $\Psi_G(i)$	true and candidate union graphs; neighbor and intimate sets	Section 3.2; Eq. (B.3)
U_Ψ, U_{sep}	guarantee ranges	Section 3.3
$\mathcal{S}_0$, $\mathcal{S}^\star$	static set of Assumption 2; candidate-dependent μ-full subset	Assumption 2; Lemma 2
$\Phi = E \circ g_I = (\beta, h)$	candidate encoder composed with the true observation map	Appendix A.1
b, h_s, $H_s = h_s^{-1}$	static map; dynamic map at static value s and its inverse	Lemma 1
$\pi_{s,y}$, ι	local graph isomorphism at (s,y); common injection on U_Ψ	Lemmas 2 and 3
$T_{i,s}$, $F_{i,s}$	scalar map of factor i; map of the complementary coordinates	Lemma 3
S_{read}, S_{effect}, $\hat{S}_{\text{read}}$	role sets of the true and learned variables	Definition 1; Section 3.3
S_0, $S^\star$	S_{read} at a fixed time; its image $\pi(S_{\text{read}})$	Appendix C.2; Corollary 1

This appendix uses the process in Section 2 and the observation and candidate conditions in Assumption 1: the image $\mathcal{I}$ is an embedded C^3 submanifold of the observation space, and a candidate's dynamic fiber at each static code is nonempty and connected, though not necessarily a product. We first establish the static correspondence needed to compare the conditional transition laws of two representations.

A.1 Conditional laws and notation

Write $\mu = P_S$, and let ρ_s denote the invariant dynamic density from Section 2. Throughout, densities are jointly measurable versions; no kernel continuity or differentiability in (s,u,a) beyond the process conditions of Section 2 is required for identification. The next-step marginal of factor k is $p_k(x_k \mid s,u,a) = \int_{\mathcal{D}_{-k}} K_{s,a}(x_k, x_{-k} \mid u)\, dx_{-k}$, and $\nabla_a p_k \not\equiv 0$ in Definition 1 means that the action derivative is nonzero somewhere on $\mathcal{D}_k \times \mathcal{S} \times \mathcal{D} \times \mathcal{A}$. The policy-averaged transition density and the conditional two-frame density are

$$\bar{K}_s(x \mid u) = \int_{\mathcal{A}} K_{s,a}(x \mid u) p(a \mid s,u)\, da, \tag{A.1}$$

$$p(u,x \mid s) = \rho_s(u) \bar{K}_s(x \mid u). \tag{A.2}$$

The densities in the integrand and ρ_s are positive, so the two-frame density is positive throughout $\mathcal{D} \times \mathcal{D}$ for almost every s. Positive joint state density also gives $\mu(V) > 0$ for every nonempty relatively open $V \subseteq \mathcal{S}$. In particular, a μ-full subset of $\mathcal{S}$ is dense there. This support property follows from the model.

For a candidate, write its open latent domain as $\hat{\Omega}$, its observation map as $\hat{g}_I$, and its encoder as $E = \hat{g}_I^{-1}$. Define

$$\hat{\mathcal{S}} = \text{proj}_s\, \hat{\Omega}, \qquad \hat{\Omega}_{\hat{s}} = \{y : (\hat{s}, y) \in \hat{\Omega}\}, \qquad \Phi = E \circ g_I = (\beta, h).$$

The two observation maps are diffeomorphisms onto the same embedded image $\mathcal{I}$. Consequently $\Phi : \mathcal{S} \times \mathcal{D} \to \hat{\Omega}$ is a C^3 diffeomorphism onto the entire candidate domain. The embedded image is Borel, and the observation maps and their inverses are measurable in its relative Borel structure. Thus equality of the logged process laws transports measurable properties of encoded frames between the two models.

A.2 STATIC CORRESPONDENCE

Lemma 1 (Static correspondence)**.** *Under the model of Section 2 and Assumption 1, every candidate admits*

$$\Phi(s,d) = \big(b(s), h_s(d)\big) \qquad ((s,d) \in \mathcal{S} \times \mathcal{D}), \tag{A.3}$$

where $b : \mathcal{S} \to \hat{\mathcal{S}}$ is a C^3 diffeomorphism and, for each s, $h_s : \mathcal{D} \to \hat{\Omega}_{b(s)}$ is a C^3 diffeomorphism onto the whole fiber. The maps $(s,d) \mapsto h_s(d)$ and $(s,y) \mapsto H_s(y) := h_s^{-1}(y)$ are jointly C^3 on their respective domains. When $n_s = 0$, the static map is omitted.

Proof. Static outputs do not vary with the dynamic state. The static output of the time-shared candidate encoder is constant along candidate trajectories, so equality of the logged process laws and measurability of the encoder give $\beta(S, D_t) = \beta(S, D_{t+1})$ almost surely under the true process law. By Eq. (A.2) and Fubini's theorem, for μ-almost every s the continuous function $\beta(s,u) - \beta(s,x)$ vanishes for Lebesgue-almost every $(u,x) \in \mathcal{D} \times \mathcal{D}$, hence everywhere. Fix $d^\circ \in \mathcal{D}$ and set $b(s) = \beta(s, d^\circ)$; then b is C^3, $\beta(s,d) = b(s)$ on a μ-full and hence dense set of static values, and continuity in s extends this equality to all of $\mathcal{S} \times \mathcal{D}$.

The two diagonal Jacobian blocks are nonsingular. The form $\Phi(s,d) = (b(s), h_s(d))$ just established and the equal block dimensions make $D\Phi(s,d)$ block lower triangular with diagonal blocks $Db(s)$ and $Dh_s(d)$, so $\det Db(s) \det Dh_s(d) = \det D\Phi(s,d) \neq 0$. For fixed s, injectivity of Φ makes h_s injective, and the inverse function theorem makes it an open map with a C^3 inverse onto its image, a nonempty open subset of $\mathbb{R}^{n_d}$.

A candidate fiber has only one static preimage. Fix $\hat{s} \in \hat{\mathcal{S}}$. Surjectivity of Φ gives

$$\hat{\Omega}_{\hat{s}} = \bigsqcup_{s:\, b(s) = \hat{s}} h_s(\mathcal{D}), \tag{A.4}$$

a disjoint union by injectivity of Φ, each member a nonempty relatively open subset of the fiber. Two or more members would disconnect $\hat{\Omega}_{\hat{s}}$, contrary to Assumption 1. Hence b is bijective onto $\hat{\mathcal{S}}$, its local C^3 inverses combine into a global one, and $h_s(\mathcal{D}) = \hat{\Omega}_{b(s)}$ for every s.

Joint smoothness. Joint C^3 smoothness of h follows from $\Phi \in C^3$, and composing $\Phi^{-1}(\hat{s}, y) = \big(b^{-1}(\hat{s}), H_{b^{-1}(\hat{s})}(y)\big)$ with $(s,y) \mapsto (b(s), y)$ gives joint C^3 smoothness of $H_s(y)$ on the open set $\{(s,y) : (b(s), y) \in \hat{\Omega}\}$. For $n_s = 0$, $\Phi = h$ is already the required dynamic diffeomorphism. □

The dynamic encoding may depend on s, and candidate fibers may vary with the static code and need not be products. Fiber connectedness excludes multiple static branches in Eq. (A.4); connectedness of the true static domain is used in Lemma 3 for a common matching of the guaranteed variables.

B UNION GRAPH AND LOCAL MIXING RESTRICTIONS

We establish the graph restriction used in Theorem 1. All conditions are stated in Sections 2–3. Throughout, G and $\hat{G}$ are the true and candidate union graphs of Section 3.2. In the appendix notation $U = D_t$ and $X = D_{t+1}$, their non-edges read

$$\{i,j\} \notin \mathcal{E}(G) \quad \Longleftrightarrow \quad X_i \perp X_j \mid (X_{-\{i,j\}}, U, S, A), \tag{B.1}$$

with the analogous definition for $\hat{G}$ in learned variables. A non-edge is a property of the joint law and holds at almost every context (s,u,a); an edge is present as soon as the dependence appears on a positive-measure set of contexts, which is why the graph is called a union. Assumption 2 uses the variation vector

$$v_{s,x}(u,a) = (\ell_{x_i x_i})_{i \in [n_d]} \oplus (\ell_{x_i})_{i \in [n_d]} \oplus (\ell_{x_i x_j})_{\{i,j\} \in \mathcal{E}(G)} \in \mathbb{R}^q, \qquad q = 2n_d + |\mathcal{E}(G)|, \tag{B.2}$$

with $\oplus$ denoting vector concatenation and all derivatives evaluated at (x,s,u,a); in Eq. (2), c and κ may depend on (s,x) but are fixed as (u,a) varies. For the neighbor set $N_G(i)$, the intimate set is

$$\Psi_G(i) = \{j \in N_G(i) : N_G(i) \setminus \{j\} \subseteq N_G(j)\}. \tag{B.3}$$

Lemma 2 (Graph identification and local mixing)**.** *Under the model of Section 2, Assumption 1, and Assumption 2, every candidate satisfies $|\mathcal{E}(\hat{G})| \geq |\mathcal{E}(G)|$. For a candidate selected by the minimum-edge criterion, the edge counts are equal. There is a measurable μ-full set $\mathcal{S}^\star \subseteq \mathcal{S}_0$, which, unlike $\mathcal{S}_0$, may depend on that candidate, such that for every $s \in \mathcal{S}^\star$ and every $y \in h_s(\mathcal{D})$, some graph isomorphism $\pi_{s,y} : V(G) \to V(\hat{G})$ satisfies*

$$\partial_{y_{\pi_{s,y}(i)}} H_{s,i}(y) \neq 0 \qquad (i \in [n_d]), \tag{B.4}$$

$$\partial_{y_{\pi_{s,y}(j)}} H_{s,i}(y) = 0 \qquad \big(j \notin \{i\} \cup \Psi_G(i)\big). \tag{B.5}$$

Here h_s, H_s are supplied by Lemma 1.

The proof occupies the rest of this appendix: a conditional change of variables (Appendix B.1), non-edge identities on a common context set (Appendix B.2), elimination of the non-edge coefficients by Assumption 2 (Appendix B.3), the all-candidate lower bound with equality under selection (Appendix B.4), and the remaining local mixing (Appendix B.5).

B.1 Conditional change of variables

The static correspondence gives, at both consecutive times, $(\hat{S}, \hat{D}_t) = (b(S), h_S(D_t))$. Since the full state map is a bimeasurable bijection,

$$\sigma(\hat{S}, \hat{D}_t, A_t) = \sigma(S, D_t, A_t).$$

Using the fixed true conditional version from Section 2, define the candidate version by its pushforward:

$$\hat{K}_{b(s),a}(y \mid h_s(u)) = K_{s,a}(H_s(y) \mid u)|\det DH_s(y)|, \qquad y \in h_s(\mathcal{D}). \tag{B.6}$$

Equality of the conditioning sigma-fields makes this a version of the candidate conditional law. The formula holds by construction for every next-state value at each context where the true kernel is specified. Writing $\ell = \log K$ and $\hat{\ell} = \log \hat{K}$ gives

$$\hat{\ell}(y \mid b(s), h_s(u), a) = \ell(H_s(y) \mid s, u, a) + \Lambda_s(y), \qquad \Lambda_s(y) = \log|\det DH_s(y)|. \tag{B.7}$$

For fixed context, this expression is C^2 in y: $H_s \in C^3$ gives $\Lambda_s \in C^2$, which suffices for the two next-state derivatives below.

B.2 Non-edges on a common full-measure context set

We use only the following implication on candidate domains: for a positive C^2 density f on an open set O, conditional independence of variables Y_k, Y_l given the others implies $\partial_{y_k y_l} \log f = 0$ throughout O (Lin, 1997). The factorization argument applies rectangle by rectangle within each section of O, and continuity of the mixed derivative extends the zero value from almost every section to all of O, so no product structure of O is needed.

For a true non-edge, the union graph definition gives, for almost every context (s, u, a), conditional independence under $K_{s,a}(\cdot \mid u)$. The preceding implication yields

$$\ell_{x_i x_j}(x \mid s, u, a) = 0 \quad \text{for all } x \in \mathcal{D} \quad (\{i, j\} \notin \mathcal{E}(G)). \tag{B.8}$$

For a candidate non-edge, the same argument applied to Eq. (B.6) gives

$$\partial_{y_k y_l}\big[\ell(H_s(y) \mid s, u, a) + \Lambda_s(y)\big] = 0 \quad \text{for all } y \in h_s(\mathcal{D}) \quad (\{k, l\} \notin \mathcal{E}(\hat{G})). \tag{B.9}$$

The context map $(s, u, a) \mapsto (b(s), h_s(u), a)$ is a diffeomorphism, and the true context density is positive by the state and action density conditions. Thus the full-measure statements can all be read on the true context space with Lebesgue measure. Intersect their sets of validity over the finitely many true and candidate non-edges and with the valid-kernel set. Fubini's theorem, followed by intersection with $\mathcal{S}_0$ from Assumption 2, gives a μ-full set $\mathcal{S}^\star$ such that, for each $s \in \mathcal{S}^\star$, Eqs. (B.8) and (B.9) hold for almost every (u, a) and all next-state values.

We never differentiate the kernel in s, u, or a: the model requires regularity only in the next-state variables, and Assumption 2 applies to an identity valid for almost every history–action pair.

B.3 DISTRIBUTIONAL VARIATION REMOVES THE NON-EDGE COEFFICIENTS

Fix $s \in \mathcal{S}^\star$, $y \in h_s(\mathcal{D})$, and put $x = H_s(y)$. Write

$$B_{ik} = \partial_{y_k} H_{s,i}(y), \qquad C_{ikl} = \partial_{y_k y_l} H_{s,i}(y), \qquad J_{kl} = \partial_{y_k y_l} \Lambda_s(y).$$

For a candidate non-edge $\{k, l\}$, the chain rule in Eq. (B.9), with true non-edge terms removed by Eq. (B.8), gives

$$\begin{aligned} 0 = &\sum_i \ell_{x_i x_i} B_{ik} B_{il} + \sum_i \ell_{x_i} C_{ikl} \\ &+ \sum_{\{i,j\} \in \mathcal{E}(G)} \ell_{x_i x_j} (B_{ik} B_{jl} + B_{jk} B_{il}) + J_{kl}. \end{aligned} \tag{B.10}$$

Every derivative of ℓ here is evaluated at (x, s, u, a). For the fixed (s, y), the coefficients B, C, J do not depend on (u, a): the encoder is frame-only, action-free, and shared across time. In the ordering of Eq. (B.2), Eq. (B.10) reads $c^\mathsf{T} v_{s,x}(u, a) = -J_{kl}$ for almost every (u, a), with coefficients c built from B and C. Assumption 2, in the form of Eq. (2) applied at the fixed (s, x), therefore makes all these coefficients zero:

$$\begin{aligned} B_{ik} B_{il} = 0, \qquad C_{ikl} = 0 \qquad (i \in [n_d]), \\ B_{ik} B_{jl} + B_{jk} B_{il} = 0 \qquad (\{i, j\} \in \mathcal{E}(G)). \end{aligned} \tag{B.11}$$

For a true edge $\{i, j\}$ and candidate non-edge $\{k, l\}$, set $P = B_{ik} B_{jl}$ and $Q = B_{jk} B_{il}$. Equation (B.11) gives $P + Q = 0$ and $PQ = (B_{ik} B_{il})(B_{jk} B_{jl}) = 0$. Hence

$$B_{ik} B_{jl} = B_{jk} B_{il} = 0. \tag{B.12}$$

The two products vanish separately; cancellation between them cannot mask a true edge.

B.4 ALL-CANDIDATE LOWER BOUND AND MINIMUM-EDGE SELECTION

At every fixed (s, y) above, $B = DH_s(y)$ is nonsingular. Its determinant has a nonzero permutation term, so there is a permutation $\pi_{s,y}$ with $B_{i,\pi_{s,y}(i)} \neq 0$ for every i. If a true edge $\{i, j\}$ mapped to a candidate non-edge, Eq. (B.12) would force $B_{i,\pi_{s,y}(i)} B_{j,\pi_{s,y}(j)} = 0$, a contradiction. Thus this permutation embeds the true edge set into the candidate edge set, and

$$|\mathcal{E}(G)| \leq |\mathcal{E}(\hat{G})| \qquad \text{for every candidate.} \tag{B.13}$$

The candidate class contains the true model by Assumption 1, so the minimum-edge criterion gives the reverse inequality for a selected candidate. Edge counts are integers in a finite set and the class is nonempty, so a minimum is attained. Equality of edge counts makes each edge embedding $\pi_{s,y}$ a graph isomorphism, preserving non-edges as well as edges.

B.5 THE REMAINING LOCAL MIXING

For a minimum candidate, fix (s, y) and abbreviate $\pi = \pi_{s,y}$. Let $j \notin \{i\} \cup \Psi_G(i)$. If i, j are nonadjacent, then $\pi(i), \pi(j)$ are a candidate non-edge. The first equality in Eq. (B.11) yields $B_{i,\pi(i)} B_{i,\pi(j)} = 0$; since the first factor is nonzero, $B_{i,\pi(j)} = 0$.

Otherwise, $j \in N_G(i)$ but $j \notin \Psi_G(i)$. The definition of $\Psi_G(i)$ supplies $r \in N_G(i) \setminus \{j\}$ with $r \notin N_G(j)$. Hence $\{\pi(r), \pi(j)\}$ is a candidate non-edge, whereas $\{r, i\}$ is a true edge. Equation (B.12) gives $B_{r,\pi(r)} B_{i,\pi(j)} = 0$, and again $B_{i,\pi(j)} = 0$. This proves Eq. (B.5) and completes Lemma 2.

Two consequences will be used in the role proof. If $i \in U_\Psi(G)$, row i of $DH_s(y)$ has exactly one nonzero entry, at column $\pi_{s,y}(i)$. If $i \in U_{\mathrm{sep}}(G)$, that column also has exactly one nonzero entry: for $r \neq i$, the definition of U_{sep} gives $i \notin \Psi_G(r)$, so $B_{r,\pi_{s,y}(i)} = 0$. For general variables the permutation can still depend on (s, y), and the permitted entries indexed by $\Psi_G(i)$ need not vanish. The global correspondences for the guaranteed variables are established separately in Appendix C.

C PARTIAL CORRESPONDENCE AND ROLE PRESERVATION

We first establish the variable correspondences using the model in Section 2, Assumptions 1 and 2, and selection by the minimum-edge criterion. The read and effect arguments then add their respective parts of Assumption 3 to prove Theorem 1 under the same matching. Lemmas 1 and 2 supply $\Phi(s,d) = (b(s), h_s(d))$ and the local restriction (B.5). Write $H_s = h_s^{-1}$ and abbreviate $U_\Psi(G), U_{\mathrm{sep}}(G)$ by U_Ψ, U_{sep}. Only these subsets receive global variable correspondences.

C.1 FROM LOCAL RESTRICTIONS TO SCALAR CORRESPONDENCES

Lemma 3 (Partial correspondence with a common matching)**.** *There is an injection $\iota : U_\Psi \to [n_d]$, independent of the static value, such that, for every $s \in \mathcal{S}$ and $i \in U_\Psi$, $j = \iota(i)$ satisfies*

$$h_{s,j}(d) = T_{i,s}(d_i), \tag{C.1}$$

where $T_{i,s}$ is a C^3 diffeomorphism from $\mathcal{D}_i$ onto an open interval. For $i \in U_{\mathrm{sep}}$, there is also a C^3 diffeomorphism $F_{i,s}$ from $\mathcal{D}_{-i}$ onto its open image such that

$$h_{s,-j}(d) = F_{i,s}(d_{-i}). \tag{C.2}$$

Both families are jointly C^3 in their static and dynamic arguments.

Proof. On the true domain $\mathcal{S}\times\mathcal{D}$, put $B(s,x) = DH_s(h_s(x))$ and $J(s,x) = Dh_s(x)$; by Lemma 1, B is continuous and nonsingular with $BJ = I$ at every (s,x). The identities of Eq. (B.11) involving B hold on $\mathcal{S}^\star \times \mathcal{D}$, which is dense by positive state density, and are continuous in (s,x), so they hold on all of $\mathcal{S} \times \mathcal{D}$. Since the determinant and graph arguments in Appendices B.4–B.5 use only these identities, nonsingularity, and equality of the edge counts, they give Eqs. (B.4)–(B.5) at every (s,y) with $y = h_s(x)$.

For $i \in U_\Psi$, row i of $B(s,x)$ consequently has exactly one nonzero entry. Its position is locally constant: that entry remains nonzero in a neighborhood, where no other entry in the row can be nonzero. The true domain $\mathcal{S} \times \mathcal{D}$ is connected, so the position is one fixed index $\iota(i)$ for all (s,x). Two such rows cannot share an index without making B singular. Thus ι is injective and agrees with every local matching on U_Ψ.

For any s and $j = \iota(i)$, row i of $BJ = I$ gives

$$J_{jm} = 0 \quad (m \neq i), \qquad J_{ji} = B_{ij}^{-1} \neq 0.$$

Each slice of the true product domain with x_i fixed is connected. Integration of these zero derivatives along that slice yields $h_{s,j}(x) = T_{i,s}(x_i)$. Its derivative is everywhere nonzero, hence has constant sign on the interval $\mathcal{D}_i$. Therefore $T_{i,s}$ is strictly monotone and is a diffeomorphism onto its open interval image.

If $i \in U_{\mathrm{sep}}$, each local matching sends i to j. Equation (B.5) applied to every row $r \neq i$ therefore gives $B_{rj} = 0$, since $i \notin \Psi_G(r)$. After ordering rows as $(i, -i)$ and columns as $(j, -j)$, B is block diagonal. Its inverse consequently satisfies $J_{li} = 0$ for $l \neq j$, with a nonsingular complementary block. Integration along the connected interval $\mathcal{D}_i$ gives $h_{s,-j}(x) = F_{i,s}(x_{-i})$. To prove global injectivity of $F_{i,s}$, suppose it takes the same value at v and v'. Pair both with any common x_i. Their full h_s images agree, so injectivity of h_s implies $v = v'$. The inverse function theorem now gives a smooth inverse on the open image of $F_{i,s}$. In particular, $h_s(\mathcal{D}) = T_{i,s}(\mathcal{D}_i) \times F_{i,s}(\mathcal{D}_{-i})$ in this ordering.

Evaluating the jointly C^3 map h at fixed complementary arguments shows that $T_{i,s}$ and $F_{i,s}$ are jointly C^3 in their respective static and dynamic arguments. When $n_s = 0$, the same proof uses a singleton static domain. □

Remark 1 (Reading the ranges on the candidate graph). Each $\pi_{s,y}$ of Lemma 2 is a graph isomorphism, so it maps intimate sets to intimate sets and the subsets U_Ψ and U_{sep} of G onto those of $\hat{G}$, and it agrees with ι on U_Ψ because row i of DH_s has a single nonzero entry for each $i \in U_\Psi$. Hence

$$\iota(U_\Psi(G)) = U_\Psi(\hat{G}), \qquad \iota(U_{\mathrm{sep}}(G)) = U_{\mathrm{sep}}(\hat{G}), \tag{C.3}$$

and the guaranteed ranges can be read off any selected candidate's own union graph, without any role condition.

C.2 READ-SET SOUNDNESS AND INVARIANCE

For the read equality, assume the read part of Assumption 3. At a fixed time, write $(S, D, A) = (s, d_t, a_t)$, $\hat{S} = b(S)$, and $\hat{D} = h_S(D)$. With $S_0 = S_{\mathrm{read}}$, the policy equation is

$$A = \phi(S, D_{S_0}, E), \qquad E \perp (S, D).$$

For any bounded measurable f, define $K_f(S, D_{S_0}) = \int f(\phi(S, D_{S_0}, e))P_E(de)$. Private-noise independence gives

$$\mathbb{E}[f(A) \mid S, D] = K_f(S, D_{S_0}). \tag{C.4}$$

If $k \notin S_0$, this expression is measurable with respect to (S, D_{-k}). Taking a conditional expectation onto that information shows $A \perp D_k \mid (S, D_{-k})$. This proves read-set soundness; read-faithfulness supplies the converse for selected factors.

For $i \in U_{\mathrm{sep}}$ and $j = \iota(i)$, the maps $(s, d_i) \mapsto (b(s), T_{i,s}(d_i))$ and $(s, d_{-i}) \mapsto (b(s), F_{i,s}(d_{-i}))$ are measurable bijections with measurable inverses. In particular,

$$\sigma(\hat{S}, \hat{D}_{-j}) = \sigma(S, D_{-i}).$$

Thus the read test conditions on the same information before adding the tested coordinate. Given this information, S is fixed and $T_{i,S}$ is a fixed measurable bijection. The joint conditional law of (A, D_i) and the product of its conditional marginals are both pushed forward by $\mathrm{id} \times T_{i,S}$. Invariance of relative entropy under a common measurable bijection, followed by expectation, gives

$$I(A; \hat{D}_j \mid \hat{S}, \hat{D}_{-j}) = I(A; D_i \mid S, D_{-i}), \qquad i \in U_{\mathrm{sep}},\ j = \iota(i), \tag{C.5}$$

including the value $+\infty$. If $i \notin S_{\mathrm{read}}$, private-noise independence and Eq. (C.4) make this quantity zero. If $i \in S_{\mathrm{read}}$, read-faithfulness makes it positive. Hence

$$\hat{S}_{\mathrm{read}} \cap \iota(U_{\mathrm{sep}}) = \iota(S_{\mathrm{read}} \cap U_{\mathrm{sep}}).$$

This proves the read identity in Eq. (4) in both directions. The scalar correspondence alone does not ensure that the conditioning information is preserved: other learned variables can still depend on D_i when $i \in U_\Psi \setminus U_{\mathrm{sep}}$.

C.3 EFFECT INVARIANCE

For the effect equality, assume the effect part of Assumption 3. The joint state map $\Phi(s, u) = (b(s), h_s(u))$ is a measurable bijection with measurable inverse, so

$$\sigma(\hat{S}, \hat{D}, A) = \sigma(S, D, A).$$

The complete preceding state therefore supplies the same conditioning information in both representations. Other next-step coordinates are integrated out when forming the tested marginal. Fix the true conditional kernel version specified in Section 2 and its marginals, and use their pushforwards as the corresponding candidate versions. Let $i \in U_\Psi$ and $j = \iota(i)$. Time sharing gives $\hat{X}_j = T_{i,S}(X_i)$ for $X = d_{t+1}$, using the same scalar map as for the current state. Scalar change of variables yields the first identity below. Since $T_{i,s}$ is action-independent, the marginal action regularity in Assumption 3 gives the second by differentiating this density identity:

$$\begin{aligned} \hat{p}_j(T_{i,s}(x_i) \mid b(s), h_s(u), a) &= \frac{p_i(x_i \mid s, u, a)}{|T'_{i,s}(x_i)|}, \\ \nabla_a \hat{p}_j(T_{i,s}(x_i) \mid b(s), h_s(u), a) &= \frac{\nabla_a p_i(x_i \mid s, u, a)}{|T'_{i,s}(x_i)|}. \end{aligned} \tag{C.6}$$

These are pointwise identities for the paired versions. When a varies, both $T_{i,s}(x_i)$ and the finite nonzero denominator remain fixed. Consequently the candidate marginal inherits C^1 action regularity and an action-independent support on this guaranteed variable; the response criterion uses only such variables. The action derivative is taken directly in the marginal-density identity, so no differentiation under the joint-density integral is needed for this step. This argument uses the marginal

law's scalar pushforward, not separation of the remaining candidate variables or a product candidate fiber.

On a connected action domain a zero action gradient everywhere makes the marginal constant in the action, whereas on a disconnected domain different components may carry different constant densities while all derivatives vanish; the effect label therefore stays derivative-based unless connectedness is added.

The map $(x_i, s, u, a) \mapsto (T_{i,s}(x_i), b(s), h_s(u), a)$ is a bijection between the true and corresponding candidate argument domains. Thus a nonzero action derivative exists on one side exactly when it exists on the other, proving the effect identity in Eq. (4) in both directions. Together with Appendix C.2, this gives the read and effect identities of Theorem 1 for the same injection ι of Lemma 3, each under its own part of Assumption 3.

For a concrete model, the marginal regularity in Assumption 3 can be verified from explicit marginal formulas or by differentiating under the marginal integral, justified by the kernel's action regularity, finite marginal integrals, and an integrable bound on its action derivatives over the marginalized coordinates, locally uniform in action; Appendix E provides both calculations for the five-chain.

C.4 Full correspondence as a special case

If $\Psi_G(i) = \varnothing$ for every vertex, then $U_\Psi = U_{\mathrm{sep}} = [n_d]$ and the injection is a permutation π:

$$\hat{S} = b(S), \qquad \hat{D}_{\pi(i)} = T_{i,S}(D_i), \qquad i \in [n_d]. \tag{C.7}$$

The scalar maps can still depend on S. Under their respective parts of Assumption 3, both set equalities cover every coordinate; the effect part also gives every candidate marginal action regularity. The ideal read-subset selection built on this full correspondence is stated in Appendix J.5.

D Shared certificates for the constructions

The constructions in Appendices E–H verify the process conditions of Section 2 and Assumption 2 through two lemmas: a stationarity certificate for Gaussian kernel families with bounded parameters (Lemma 4) and a pair of certificates for Assumption 2 (Lemma 5). That assumption requires the vector of next-state derivatives of the log kernel to have full essential affine span across contexts (u, a); the constructions realize this through Gaussian kernels whose means and precision matrices depend on the history. Certificate (i) (finitely many contexts) is used for the five-chain (Appendix E.2) and the cup model (Appendix F.3); certificate (ii) (derivative span on almost every slice) for the state-pair example (Appendix G.2) and the four-cycle (Appendix H.5). Each construction applies Lemma 4 with its own spectral and mean bounds.

D.1 Stationarity of bounded Gaussian kernel families

Lemma 4 (Bounded Gaussian kernel families)**.** *Let $\mathcal{D} = \mathbb{R}^n$, let $\mathcal{A}$ be open, and let $K_a(x \mid u) = \mathcal{N}\big(x; m(u,a), Q(u)^{-1}\big)$, where $(u,a) \mapsto (m(u,a), Q(u))$ is measurable, $\lambda I \preceq Q(u) \preceq \Lambda I$ for constants $0 < \lambda \le \Lambda < \infty$, and $\|m(u,a)\| \le M < \infty$ in the Euclidean norm for all (u,a). Let the policy density $p(a \mid u)$ be jointly measurable and positive on $\mathcal{A}$. Then, for all (x,u,a),*

$$K_a(x \mid u) \ge g(x) := \Big(\frac{\lambda}{2\pi}\Big)^{n/2} \exp\Big[-\frac{\Lambda}{2}\big(\|x\| + M\big)^2\Big]. \tag{D.1}$$

The policy-averaged kernel $\bar{K}(x \mid u) = \int K_a(x \mid u) p(a \mid u)\, da$ satisfies $\bar{K} \ge \epsilon\nu$ with $\epsilon = \int g \in (0,1]$ and $\nu = g/\epsilon$, is a contraction in total variation with coefficient $1 - \epsilon$, and has a unique invariant law, whose density ρ is C^∞ with $\rho \ge g > 0$. Initialized from ρ, the process is strictly stationary, and its state–action density is positive on $\mathbb{R}^n \times \mathcal{A}$.

Proof. Since $\det Q(u) \ge \lambda^n$ and $(x-m)^\top Q(u)(x-m) \le \Lambda\|x-m\|^2 \le \Lambda(\|x\| + M)^2$, the Gaussian density is bounded below by g; averaging over the policy density preserves the bound, and $\epsilon \le 1$ because $g \le \bar{K}$ and $\bar{K}$ integrates to one. For probability laws P, P', the minorization gives $\|P\bar{K} - P'\bar{K}\|_{\mathrm{TV}} \le (1-\epsilon)\|P - P'\|_{\mathrm{TV}}$, so completeness in total variation and the

contraction mapping theorem give a unique invariant law (Meyn & Tweedie, 2009), with density $\rho(x) = \int \bar{K}(x \mid u)\rho(du) \geq g(x)$. Every x-derivative of the Gaussian kernel is a polynomial in $x - m$ and the entries of Q times the kernel, hence bounded on compact x sets uniformly in (u, a), and differentiation under the probability integral gives $\rho \in C^\infty$. Initialization from ρ makes the state chain, and with it the logged state–action process, strictly stationary, with joint density $\rho(u)p(a \mid u) > 0$. □

D.2 Two certificates for Assumption 2

Assumption 2 concerns, at a fixed (s, x), the vector $v_{s,x}(u, a)$ of Eq. (B.2) as a function of the context (u, a) on $\mathcal{D} \times \mathcal{A}$, with $q = 2n_d + |\mathcal{E}(G)|$.

Lemma 5 (Certificates for distributional variation)**.** *Fix $s \in \mathcal{S}_0$ and $x \in \mathcal{D}$. Each of the following implies Eq.* (2) *at (s, x).*

- *(i)* (Finitely many contexts.) *The map $(u, a) \mapsto v_{s,x}(u, a)$ is continuous on $\mathcal{D} \times \mathcal{A}$, and there are contexts $(u_0, a_0), \ldots, (u_q, a_q)$ whose q difference vectors $v_{s,x}(u_b, a_b) - v_{s,x}(u_0, a_0)$, $b = 1, \ldots, q$, are linearly independent.*
- *(ii)* (Derivative span on almost every slice.) *For almost every $a \in \mathcal{A}$ the kernel version $K_{s,a}(x \mid u)$ is jointly C^3 in (x, u), and there is a history index m such that, for almost every (a, u_{-m}), the derivatives $V'(t)$ of the slice $V(t) = v_{s,x}(u, a)|_{u_m = t}$ span $\mathbb{R}^q$ as t ranges over $\mathcal{D}_m$.*

Proof. (i) The determinant of the $q \times q$ matrix of difference vectors is continuous in the $q+1$ contexts and nonzero at the displayed witnesses, hence nonzero on a product of open neighborhoods of them. A Lebesgue-full subset of $\mathcal{D} \times \mathcal{A}$ meets each neighborhood, so if $c^\mathsf{T} v_{s,x}(u, a) = \kappa$ held for almost every (u, a), it would hold at $q + 1$ contexts with linearly independent differences; subtracting the base context gives c^T times each difference equal to zero, hence $c = 0$.

(ii) Suppose $c^\mathsf{T} v_{s,x}(u, a) = \kappa$ for almost every (u, a). By Fubini's theorem, for almost every (a, u_{-m}) the equality holds for almost every u_m. On such a slice that also has full derivative span, joint C^3 regularity makes V continuously differentiable in u_m, so the equality extends to every $u_m \in \mathcal{D}_m$, and differentiation gives $c^\mathsf{T} V'(t) = 0$ for all t. Full derivative span forces $c = 0$. □

Both certificates verify the same spanning condition. Along a slice in which one history variable $t = u_m$ varies on its interval, for any fixed interior t_0,

$$\operatorname{span}\{V'(t) : t \in \mathcal{D}_m\} = \operatorname{span}\{V(t) - V(t_0) : t \in \mathcal{D}_m\}, \tag{D.2}$$

since $V(t) - V(t_0) = \int_{t_0}^{t} V'(r)\, dr$ lies in the closed derivative span and each derivative is a limit of difference quotients in the closed difference span. This is the local form of the sufficient-change conditions of Zhang et al. (2024, Section 3.1). Certificate (i) needs an explicit kernel whose derivative vector is continuous in the context, which the five-chain and the cup model have; certificate (ii) needs no continuity in the action. The identification proofs use only Eq. (2).

E A five-chain process satisfying the conditions

This construction is a complete stationary process with the guarantees shown in Figure 3. Its union graph is the path $1 - 2 - 3 - 4 - 5$, with intimate sets (Eq. (B.3)) $\Psi_G(1) = \{2\}$, $\Psi_G(5) = \{4\}$, and $\Psi_G(2) = \Psi_G(3) = \Psi_G(4) = \varnothing$, so $U_\Psi = \{2, 3, 4\}$ and $U_{\text{sep}} = \{3\}$. Variable 3 in U_{sep} has guaranteed positive read and effect labels; in U_Ψ, variable 4 has a guaranteed positive effect label and variable 2 a guaranteed negative one; the endpoints lie outside both ranges.

We verify the process conditions (Appendices E.1–E.3) and show that the identity candidate attains the edge minimum, so the guarantee applies to every minimum candidate (Appendix E.4). The construction uses $n_s = 0$, $\mathcal{D} = \mathbb{R}^5$, $\mathcal{A} = (-1, 1)$, and $g_I = \mathrm{id}$.

E.1 Transition kernel and contemporaneous equations

For $0 < \delta \le 1/100$, define

$$\begin{aligned} z &= \tanh u_1, & F_j(u) &= z^j/j \quad (j = 1, \dots, 14), \\ r_i(u) &= \delta F_i(u), & Q_{ii}(u) &= 4 + \delta F_{5+i}(u), \\ Q_{i,i+1}(u) &= Q_{i+1,i}(u) = -1 + \delta F_{10+i}(u). \end{aligned} \tag{E.1}$$

The diagonal indices run from 1 to 5 and the edge indices from 1 to 4; all other off-diagonal entries vanish. Since $|F_j| \le 1/j$, Gershgorin's bounds give, uniformly over u,

$$1.9I \preceq Q(u) \preceq 6.1I, \qquad Q_{i,i+1}(u) < 0.$$

Set $\mu_0 = Q^{-1}r$ and define the backward Schur recursion

$$\lambda_5 = Q_{55}, \quad \lambda_i = Q_{ii} - Q_{i,i+1}^2/\lambda_{i+1}, \quad \beta_k = -Q_{k-1,k}/\lambda_k.$$

Here $i = 4, \dots, 1$ and $k = 2, \dots, 5$. The bounds $Q_{ii} \in [3.99, 4.01]$ and $|Q_{i,i+1}| \le 1.01$ imply by backward induction that $3 \le \lambda_i \le 4.01$ and $0 < \beta_k < 1/2$. All these functions are smooth. Define

$$K_a(x \mid u) = \mathcal{N}\big(x; \mu_0(u) + aw(u), Q(u)^{-1}\big), \qquad w = (0, 0, 1, \beta_4, \beta_5\beta_4)^\top. \tag{E.2}$$

For independent fresh standard Gaussian noises ε_k, the corresponding structural equations are

$$\begin{aligned} X_1 &= \mu_{0,1} + \lambda_1^{-1/2}\varepsilon_1, \\ X_k &= \mu_{0,k} + \beta_k(X_{k-1} - \mu_{0,k-1}) + \mathbf{1}_{\{k=3\}}a + \lambda_k^{-1/2}\varepsilon_k, \quad k = 2, \dots, 5. \end{aligned}$$

To check the kernel, let L have unit diagonal and $L_{k,k-1} = -\beta_k$, with other entries zero. The recursion gives $Q = L^\top \operatorname{diag}(\lambda) L$, and the chosen w satisfies $Lw = e_3$. Thus the equations have exactly the mean and covariance in Eq. (E.2). The action enters equation 3 directly and propagates to equations 4 and 5. The coefficients β_k vary with history; they are determined by Q.

The kernel is positive on the common support $\mathbb{R}^5$ and jointly C^∞ in (x, u, a). A Gaussian conditional-independence graph is given by the precision matrix's nonzero off-diagonal entries (Lauritzen, 1996, Chapter 5). Consequently the actual graph at every context, and hence the union graph, is $1 - 2 - 3 - 4 - 5$. It has four edges and $q = 2n_d + |\mathcal{E}(G)| = 14$. Its agreement with the moral graph of the displayed chain equations follows from this explicit calculation (Appendix H.1).

E.2 Distributional variation at every next-state value

Let $\ell = \log K$ and take the 15 contexts

$$u_b = (\operatorname{arctanh}(b/15), 0, 0, 0, 0), \qquad a_b = 0, \qquad b = 0, \dots, 14.$$

For every fixed $x \in \mathbb{R}^5$, form the derivative vector in Assumption 2, ordered as the five diagonal second derivatives, five first derivatives, and four edge second derivatives. At zero action,

$$\ell_{x_i} = r_i - \sum_k Q_{ik}x_k, \qquad \ell_{x_ix_k} = -Q_{ik}.$$

Write Δ for differences from $(u_0, 0)$. The row operation

$$\Delta\ell_{x_i} - x_i\Delta\ell_{x_ix_i} - \sum_{k \in N_G(i)} x_k\Delta\ell_{x_ix_k} = \Delta r_i$$

removes all x dependence from each first-derivative row. Its coefficients are the same for all 14 difference columns, and its determinant is one: it adds only curvature rows to first-derivative rows. After reordering rows and changing signs, the difference matrix is

$$M_{jb} = \delta\frac{(b/15)^j}{j}, \qquad j, b = 1, \dots, 14.$$

The Vandermonde determinant gives

$$|\det M| = \frac{\delta^{14}}{14!}\left(\prod_{b=1}^{14}\frac{b}{15}\right)\prod_{1 \le b < c \le 14}\frac{c-b}{15} = \frac{\delta^{14}\prod_{k=1}^{13}k!}{15^{105}} > 0. \tag{E.3}$$

Thus the same 15 contexts give full rank at every x.

For this explicit kernel the derivative vector is continuous in (u, a) at each fixed x, so Lemma 5(i) with these 15 contexts proves Assumption 2 for every x and every $0 < \delta \le 1/100$.

E.3 POLICY, STATIONARITY, AND ROLE LABELS

For $\theta_1, \theta_3 > 0$, use the policy density

$$p(a \mid u) = \frac{\exp(a\eta(u))}{\int_{-1}^{1} \exp(v\eta(u))\, dv}, \qquad \eta(u) = \theta_1 \tanh u_1 + \theta_3 \tanh u_3. \tag{E.4}$$

It is smooth and positive on the common open action support. The inverse conditional CDF applied to an independent uniform random variable realizes it using private policy noise and the selector $S_{\mathrm{read}} = \{1, 3\}$. Use fresh policy noise at each time and transition disturbances independent of the past and of this policy noise.

The parameter bounds imply $\|Q^{-1}r\| < 1/10$ and $\|w\| < 19/10$, so $\|\mu_0 + aw\| < 2$ for $|a| < 1$, and Lemma 4 applies with $\lambda = 1.9$, $\Lambda = 6.1$, and $M = 2$; its stationary process with a positive smooth state–action density is what the model and read-faithfulness require.

Distinct values of the parameter η in Eq. (E.4) give distinct action distributions. Holding the remaining state fixed, each of u_1 and u_3 changes η strictly monotonically. The positive state density gives a nondegenerate conditional distribution for each variable given the others, for almost every conditioning value. Hence $I(A; D_i \mid D_{-i}) > 0$ for $i = 1, 3$. For the other variables, the policy kernel depends only on the remaining state and private noise, giving zero conditional mutual information. Thus the selector satisfies read-faithfulness.

Each next-state marginal is Gaussian with variance $\sigma_i^2(u) = [Q(u)^{-1}]_{ii}$ and mean $\mu_{0,i}(u) + aw_i(u)$, so

$$\partial_a p_i(x_i \mid u, a) = w_i(u) \frac{x_i - \mu_{0,i}(u) - aw_i(u)}{\sigma_i^2(u)} p_i(x_i \mid u, a). \tag{E.5}$$

Since $w_1 = w_2 = 0$ and $w_3, w_4, w_5 > 0$, the derivative-defined effect set is $S_{\mathrm{effect}} = \{3, 4, 5\}$. Every marginal is smooth in action on an action-independent support, verifying the effect part of Assumption 3. The derivative in Eq. (E.5) can also be computed under the marginal integral: $\partial_a K_a = w^\top Q(x - \mu)K_a$ has an integrable Gaussian bound when marginalizing over x_{-i}, locally in the fixed arguments. This bound justifies the interchange for this kernel.

E.4 MINIMUM CANDIDATES AND THE NONEMPTY GUARANTEE

Lemma 2, whose process conditions were verified above, gives $|\mathcal{E}(\hat{G})| \geq 4$ for every candidate. The identity encoder belongs to the class of Assumption 1 and its union graph has four edges, so it attains the minimum. Every candidate selected by the minimum-edge criterion therefore falls within Theorem 1 and carries the labels listed at the start of this appendix.

The construction satisfies every condition above for every $0 < \delta \leq 1/100$ and $\theta_1, \theta_3 > 0$, in particular on the parameter box $\delta \in [1/200, 1/100]$, $\theta_1, \theta_3 \in [1/2, 3/2]$.

F THE CUP MODEL IN TWO VERSIONS

This construction realizes the cup example of Sections 2 and 3 as two processes that satisfy the model of Section 2 and Assumptions 1–3, with the identity encoder attaining the edge minimum, so that Theorem 1 applies to every candidate selected by the minimum-edge criterion. The two versions differ only in whether position and velocity share a jolt. We verify the union graph, the variation condition, stationarity, and the role conditions for both, compute the guarantee ranges, and check what the criterion does with the mixed encoder of Section 2; Table 2 summarizes the outcome.

The construction uses $n_s = 0$, $\mathcal{D} = \mathbb{R}^3$, $\mathcal{A} = (-1, 1)$, and $g_I = \mathrm{id}$: coordinates 1, 2, and 3 are position, velocity, and contact force, measured as deviations from nominal values, and a frame shows all three (the idealization stated in Section 3.2). Static content such as the table's appearance enters neither relation and is omitted, as in Appendix E; a static block that neither the policy nor the kernel uses could be appended without changing any step below.

F.1 Transition kernel and structural equations

For $0 < \delta \le 1/100$, define

$$\begin{aligned} z = \tanh u_1, \qquad & F_j(u) = z^j/j \quad (j = 1, \dots, 7), \\ r_i(u) = \delta F_i(u), \qquad & Q_{ii}(u) = 4 + \delta F_{3+i}(u) \quad (i = 1, 2, 3), \\ Q_{13} = Q_{23} = 0, \quad & Q_{12}(u) = Q_{21}(u) = \begin{cases} -1 + \delta F_7(u) & \text{(shared jolt)}, \\ 0 & \text{(separate jolts)}. \end{cases} \end{aligned} \tag{F.1}$$

Since $|F_j| \le 1/j$, every diagonal entry lies in $[3.99, 4.01]$ and, in the shared-jolt version, $Q_{12} \in [-1.002, -0.998]$; Gershgorin's bounds give $2.9I \preceq Q(u) \preceq 5.1I$ uniformly over u in both versions. Set $\mu_0 = Q^{-1}r$, fix $w = (1/2, 1, 1/2)^\mathsf{T}$, and define

$$K_a(x \mid u) = \mathcal{N}\big(x; \mu_0(u) + aw, Q(u)^{-1}\big). \tag{F.2}$$

All entries are smooth in u; the kernel is positive on the common support $\mathbb{R}^3$ and jointly C^∞ in (x, u, a). The next-step randomness depends on where the cup is, through z, as friction may vary across a table; the push a shifts all three means.

For independent standard Gaussian noises ε_i and $\mu = \mu_0(u) + aw$, the shared-jolt version has the structural equations

$$\begin{aligned} X_2 &= \mu_2 + \lambda_2^{-1/2}\varepsilon_2, \\ X_1 &= \mu_1 + \beta(u)\,(X_2 - \mu_2) + Q_{11}^{-1/2}\varepsilon_1, \qquad \beta = -\frac{Q_{12}}{Q_{11}}, \quad \lambda_2 = Q_{22} - \frac{Q_{12}^2}{Q_{11}}. \\ X_3 &= \mu_3 + Q_{33}^{-1/2}\varepsilon_3, \end{aligned} \tag{F.3}$$

The velocity jolt ε_2 enters position with the state-dependent share $\beta(u) \in (0.249, 0.251)$; position also has a jolt of its own, and the contact force has one that no other coordinate shares. To check the kernel, let L have unit diagonal, $L_{12} = -\beta$, and other entries zero; then $Q = L^\mathsf{T}\,\mathrm{diag}(Q_{11}, \lambda_2, Q_{33})L$, so the equations have exactly the mean and covariance in Eq. (F.2). In the separate-jolt version, $\beta \equiv 0$ and $\lambda_2 = Q_{22}$: each coordinate is $X_i = \mu_i + Q_{ii}^{-1/2}\varepsilon_i$ with its own jolt.

F.2 Union graphs and guarantee ranges

A Gaussian conditional-independence graph is given by the nonzero off-diagonal entries of the precision matrix (Lauritzen, 1996, Chapter 5). In the shared-jolt version $Q_{12} < 0$ at every context and $Q_{13} = Q_{23} = 0$, so the union graph is the single edge 1–2, with $q = 2n_d + |\mathcal{E}(G)| = 7$. In the separate-jolt version Q is diagonal at every context, so the union graph is empty and $q = 6$. Reading the intimate sets off these graphs gives the ranges in Table 2; its last row is verified in Appendix F.5.

Table 2: Union graph, guarantee ranges, and whether the criterion retains the mixed encoder, in the two cup versions of Eq. (F.1); coordinates 1, 2, and 3 are position, velocity, and contact force. Inside the ranges, every selected candidate carries the true labels, negative ones included; labels outside them are not covered by Theorem 1 but may still be identifiable.

	Separate jolts	Shared jolt
Union graph G	empty	the single edge 1–2
Intimate sets	$\Psi_G(i) = \varnothing$ for $i = 1, 2, 3$	$\Psi_G(1) = \{2\}$, $\Psi_G(2) = \{1\}$, $\Psi_G(3) = \varnothing$
$U_\Psi = U_{\mathrm{sep}}$	$\{1, 2, 3\}$	$\{3\}$
Guaranteed labels	all six, including velocity's negative read label	both labels of the contact force
Mixed encoder $(x_1 + x_2, x_2, x_3)$	discarded: one edge, none in G	retained: one edge, one in G

F.3 Distributional variation at every next-state value

Let $\ell = \log K$ and take the $q+1$ contexts

$$u_b = \big(\operatorname{arctanh}(b/(q+1)), 0, 0\big), \qquad a_b = 0, \qquad b = 0, \dots, q.$$

For every fixed $x \in \mathbb{R}^3$, form the vector of Eq. (B.2): three diagonal second derivatives, three first derivatives, and, in the shared-jolt version, the edge second derivative. At zero action,

$$\ell_{x_i} = r_i - \sum_k Q_{ik} x_k, \qquad \ell_{x_i x_k} = -Q_{ik}.$$

Write Δ for differences from $(u_0, 0)$. The row operation

$$\Delta\ell_{x_i} - x_i \Delta\ell_{x_i x_i} - \sum_{k \in N_G(i)} x_k \Delta\ell_{x_i x_k} = \Delta r_i$$

removes all x dependence from the first-derivative rows; its coefficients are the same in every difference column, and its determinant is one. After reordering rows and changing signs, the difference matrix is

$$M_{jb} = \delta \frac{(b/(q+1))^j}{j}, \qquad j, b = 1, \dots, q,$$

whose Vandermonde determinant is nonzero:

$$|\det M| = \frac{\delta^q}{q!} \prod_{b=1}^{q} \frac{b}{q+1} \prod_{1 \le b < c \le q} \frac{c-b}{q+1} > 0. \tag{F.4}$$

The derivative vector is continuous in (u, a) at each fixed x, so Lemma 5(i) with these contexts proves Assumption 2 for every x and every $0 < \delta \le 1/100$, in both versions.

In equations of the form of Eq. (F.3), the condition requires any nonzero share β to vary with the state. If velocity's jolt instead entered position with a fixed coefficient, as semi-implicit Euler integration with a fixed step gives, $X_1 = \mu_1 + \Delta t\,(X_2 - \mu_2) + \sigma(u)\,\varepsilon_1$ with constant $\Delta t > 0$ and any positive $\sigma(u)$, then $Q_{11} = \sigma(u)^{-2}$ and $Q_{12} = -\Delta t\,\sigma(u)^{-2}$, so $\ell_{x_1 x_2} = -\Delta t\,\ell_{x_1 x_1}$ at every context, whatever $\sigma(u)$ and the velocity variance do. The vector c with entry Δt on the $\ell_{x_1 x_1}$ component, entry 1 on the $\ell_{x_1 x_2}$ component, and zeros elsewhere, together with $\kappa = 0$, then violates Eq. (2). The encoder $(x_1 - \Delta t\, x_2, x_2, x_3)$ has an empty union graph and would be selected in place of the true model.

F.4 Policy, stationarity, and role labels

For $\theta_1, \theta_3 > 0$, use the policy density

$$p(a \mid u) = \frac{\exp(a\eta(u))}{\int_{-1}^{1} \exp(v\eta(u))\, dv}, \qquad \eta(u) = \theta_1 \tanh u_1 + \theta_3 \tanh u_3, \tag{F.5}$$

which reads position and contact force with some randomness. It is smooth and positive on the common open action support; the inverse conditional CDF applied to an independent uniform random variable realizes it with private policy noise and a selector that marks coordinates 1 and 3, so that $S_{\mathrm{read}} = \{1, 3\}$. Use fresh policy noise at each time and transition disturbances independent of the past and of this policy noise.

The bounds above give $\|\mu_0\| \le \|r\|/2.9 < 1/200$ and $\|w\| < 5/4$, so $\|\mu_0 + aw\| < 2$ for $|a| < 1$, and Lemma 4 applies with $\lambda = 2.9$, $\Lambda = 5.1$, and $M = 2$; its stationary process with a positive smooth state–action density is what the model and read-faithfulness require.

Distinct values of η give distinct action distributions. Holding the remaining state fixed, each of u_1 and u_3 changes η strictly monotonically, and the positive state density gives a nondegenerate conditional distribution of each variable given the others for almost every conditioning value. Hence $I(A; D_i \mid D_{-i}) > 0$ for $i = 1, 3$, while for velocity the policy kernel depends only on the remaining state and private noise, giving zero conditional mutual information. Thus the selector satisfies read-faithfulness, in both versions.

Each next-state marginal is Gaussian with variance $\sigma_i^2(u) = [Q(u)^{-1}]_{ii}$ and mean $\mu_{0,i}(u) + aw_i$, so

$$\partial_a p_i(x_i \mid u, a) = w_i \frac{x_i - \mu_{0,i}(u) - aw_i}{\sigma_i^2(u)}\, p_i(x_i \mid u, a), \tag{F.6}$$

and $w_i > 0$ gives $S_{\text{effect}} = \{1, 2, 3\}$ in both versions. Every marginal is smooth in the action on an action-independent support, verifying the effect part of Assumption 3; the derivative can also be computed under the marginal integral, since $\partial_a K_a = w^{\mathsf{T}} Q(x - \mu) K_a$ has an integrable Gaussian bound when marginalizing over x_{-i}, locally in the fixed arguments.

F.5 Minimum candidates and the mixed encoder

Lemma 2, whose process conditions were verified above, gives $|\mathcal{E}(\hat{G})| \geq |\mathcal{E}(G)|$ for every candidate. The identity encoder belongs to the class of Assumption 1 and its union graph has one edge (shared jolt) or none (separate jolts), so it attains the minimum, and every candidate selected by the minimum-edge criterion falls within Theorem 1. With a shared jolt, the contact force lies in U_{sep}: every selected candidate has a learned variable whose read and effect labels are those of the contact force, both positive, while position and velocity lie outside both ranges. With separate jolts, every factor lies in U_{sep} and all six labels are guaranteed, including velocity's negative read label.

The mixed encoder of Section 2, $E(x) = (x_1 + x_2, x_2, x_3)$, is a linear bijection of $\mathbb{R}^3$, so the model with observation map E^{-1} and encoder E satisfies Assumption 1 and reproduces the logged process law. Its kernel at context (Eu, a) is Gaussian with precision $E^{-\mathsf{T}} Q(u) E^{-1}$, whose off-diagonal entries are $Q_{12} - Q_{11}$ for the pair $\{1, 2\}$ and zero for the pairs $\{1, 3\}$ and $\{2, 3\}$. Since $Q_{12} - Q_{11} < 0$ in both versions, this candidate's union graph has exactly one edge. With separate jolts the true graph has none, so the criterion discards the candidate; with a shared jolt the counts agree, so the criterion retains it. In the retained candidate, learned variable 2 is velocity, and conditioning on $(\hat{Y}_1, \hat{Y}_3) = (x_1 + x_2, x_3)$ makes $\hat{Y}_2 = x_2$ determine position $x_1 = \hat{Y}_1 - \hat{Y}_2$, on which η depends strictly monotonically; with the positive state density this gives $I(A; \hat{Y}_2 \mid \hat{Y}_1, \hat{Y}_3) > 0$, so velocity acquires a positive read label in this candidate although the policy never uses it. Its variable 3 equals the contact force and keeps both labels, as the theorem guarantees.

The construction satisfies every condition above for every $0 < \delta \leq 1/100$ and $\theta_1, \theta_3 > 0$.

G Why state-pair edge minimality is insufficient

Let the state-pair graph be the conditional-independence graph on $(d_{\tau-1}, d_\tau)$ given $(s, a_{\tau-1})$, with every pair tested given the remaining coordinates. Its current-time induced subgraph is the union graph G.

Minimizing the total number of state-pair edges can reward removal of lag dependencies even when the resulting coordinates mix true factors. The construction below retains single-step distributional variation and uses the same smooth encoding at both times. Its true factors have five pair edges, while a mixed representation has at most four; this strict gap excludes every componentwise representation from the minima.

Proposition 1 (Mixed state-pair minima)**.** *There exists a stationary, positive, smooth two-factor process with independent transition innovations, an empty union graph, and full $q = 4$ history variation at every current point, for which every encoder minimizing the total number of state-pair edges over the observationally equivalent smooth, block-pure, time-shared model class is non-componentwise.*

Block purity means that the static and dynamic codes depend only on their respective true blocks; with the empty static block used below it restricts nothing, whereas Assumption 1 generally permits static-dependent dynamic recoding.

G.1 A positive stationary process

Take two dynamic factors on $\mathbb{R}^2$, choose $0 < \epsilon < 1$, and write

$$r = u_1 + u_2, \qquad f(r) = \tanh r, \qquad \alpha_1(r) = 2 + \epsilon \sin r, \qquad \alpha_2(r) = 2 + \epsilon \cos r.$$

Define the transition by

$$X_{\tau,i} = f(U_1 + U_2) + \alpha_i(U_1 + U_2)^{-1/2}E_{\tau,i}, \qquad E_{\tau,i} \overset{\text{iid}}{\sim} N(0,1), \tag{G.1}$$

with $U = X_{\tau-1}$. Conditional on history, the current factors are independent, so G is empty and every intimate set is empty. Take $n_s = 0$, set the frame map to the identity, and append independent Gaussian control noise as the action, with an empty policy selector and no action argument in the dynamics.

The precision matrix $\mathrm{diag}(\alpha_1, \alpha_2)$ has spectrum in $[2-\epsilon, 2+\epsilon]$ and the mean (f, f) has norm below $\sqrt{2}$, so Lemma 4 applies with $\lambda = 2 - \epsilon$, $\Lambda = 2 + \epsilon$, $M = \sqrt{2}$, and the Gaussian action density; writing ρ for its invariant law, the stationary pair density $\rho(u)K(x \mid u)$ is positive and smooth on $\mathbb{R}^4$.

G.2 Full $q = 4$ variation at every current point

Set $b_i = \alpha_i f$. For either $m = 1$ or $m = 2$,

$$Q_i := \ell_{x_i x_i u_m} = -\alpha_i'(r), \qquad L_i := \ell_{x_i u_m} = -x_i\alpha_i'(r) + b_i'(r). \tag{G.2}$$

For every fixed x, the vector (Q_1, Q_2, L_1, L_2) is a fixed invertible linear transformation of $V(r) = (\alpha_1', \alpha_2', b_1', b_2')$. To prove its full span, suppose constants A, B, C, D satisfy

$$A\alpha_1' + B\alpha_2' + C(\alpha_1 f)' + D(\alpha_2 f)' \equiv 0. \tag{G.3}$$

Along $r = \theta + 2\pi k$ as $k \to +\infty$, $f \to 1$ and $f' \to 0$. Eq. (G.3) gives $(A + C)\epsilon\cos\theta - (B + D)\epsilon\sin\theta = 0$ for every θ, hence $A + C = B + D = 0$. As $k \to -\infty$, $f \to -1$ and $f' \to 0$, giving $A - C = B - D = 0$. Thus all four coefficients vanish.

The values of V span $\mathbb{R}^4$; otherwise a nonzero linear functional would annihilate Eq. (G.3). Therefore four values of the history variable with linearly independent vectors exist. This holds for every fixed x and either history index m, since varying u_m at fixed u_{-m} lets r range over $\mathbb{R}$. The derivative span is therefore full at every action and complementary history with $q = 4$, and Lemma 5(ii) gives Assumption 2.

G.3 Five true pair edges, at most four after mixing

The true pair log density is $\log\rho(u) + \ell(x \mid u)$. There is no current edge $\{x_1, x_2\}$. All four lag–current edges $\{u_m, x_i\}$ are present because the mixed derivatives L_i in Eq. (G.2) are not identically zero. The historical edge $\{u_1, u_2\}$ is present: the coefficient of x_1^2 in its mixed log-density derivative is $-\alpha_1''(r)/2$, which is not identically zero and cannot be canceled by $\log\rho(u)$. For a positive smooth density on this product domain, a pair is conditionally independent given all remaining variables exactly when its mixed log-density derivative vanishes identically. The true pair graph therefore has exactly five edges.

Apply the same orthogonal coordinate change at every time:

$$h(x) = \left(\frac{x_1 + x_2}{\sqrt{2}}, \frac{x_1 - x_2}{\sqrt{2}}\right). \tag{G.4}$$

Writing $v = h(u)$ and $y = h(x)$, the transformed transition depends on history only through v_1, because $r = \sqrt{2}v_1$. The pair density is $\hat\rho(v)\hat K(y \mid v_1)$; hence both lag edges from v_2 to current coordinates are absent. At most four edges remain: two from v_1 to y, one within the current pair, and one within the historical pair. Table 3 compares the counts.

This map is a shared C^∞ diffeomorphism, is block-pure in the present setting with no static variables, and reproduces the same observations through h^{-1}. The transformed Gaussian conditional covariance is positive definite and admits a smooth Cholesky factorization, so it is representable with independent Gaussian innovations and a contemporaneous acyclic ordering. It therefore remains within the observationally equivalent smooth, block-pure, time-shared model class of Proposition 1.

The attainable pair-edge counts form a nonempty finite set of integers, so a minimum is attained and is at most four. Every componentwise bijection and permutation preserves conditional-independence

Table 3: State-pair edges in the true coordinates and after the mixing h of Eq. (G.4). Componentwise recodings keep all five true edges; the mixed coordinates lose both lag–current edges from v_2, so at most four remain and every componentwise encoder is excluded from the minima.

Coordinates	Historical pair	Lag–current	Current pair	Total
True, or any componentwise recoding	1	4	0	5
Mixed, $v = h(u)$ and $y = h(x)$	≤ 1	≤ 2	≤ 1	≤ 4

sigma-fields and hence all five true pair edges. No pair-edge minimizer can be componentwise. This proves Proposition 1.

The true union graph is empty and already attains the zero-edge minimum in the minimum-edge criterion. The state-pair objective instead favors removing lag edges through a within-state mixture shared across time, despite full single-step variation.

Independent actions give empty read and effect sets in this example, and the process satisfies the single-step condition of Assumption 2 but not the two-step window condition of Li et al. (2025).[2]

H Scope of full correspondence and a four-cycle example

The history-dependent four-cycle realizes the full-correspondence case $\Psi_G(i) = \varnothing$ for every node, so that $U_\Psi = U_{\mathrm{sep}} = [n_d]$, despite nonempty contemporaneous dependencies. Its behavior policy uses factor 1 only, while all four next-step marginals respond to action.

This full case has a structural restriction. Under an acyclic model with independent innovations, Appendix H.1 shows that a nonempty G with all intimate sets empty requires conditional independences beyond those implied by the generating structure. The fixed four-cycle of Appendix H.2 illustrates the required cancellation, but its constant covariance violates Assumption 2. The history-dependent four-cycle of Appendices H.3–H.6 satisfies the model, Assumptions 1 and 2, and the role conditions; the edge minimum is attained, and every minimum candidate has full correspondence and both complete role sets.

H.1 The union graph and the moral graph

Suppose the next-step variables follow a contemporaneous acyclic structural model with generating DAG D whose innovations are mutually independent and independent of the context $c = (u, s, a)$, $u = d_{\tau-1}$. The acyclic recursion factorizes $K_c(x) = \prod_i p(x_i \mid x_{\mathrm{pa}_D(i)}, c)$; two variables nonadjacent in the moral graph $M(D)$ share no factor, so the mixed derivative of $\log K_c$ in that pair vanishes and, on the positive product support of Section 2, they are conditionally independent. Hence $G \subseteq M(D)$, the inclusion of Zhang et al. (2024, Lemma 1), whose Proposition 2 gives equality under faithfulness variants. The reverse inclusion is not assumed here, since mixed derivatives of different factors can cancel; Theorem 1 imposes no faithfulness condition on the contemporaneous structure, and Appendices H.3–H.5 give a process satisfying Assumption 2 while violating such faithfulness.

If D has at least one edge and $G = M(D)$, then some node of G has a nonempty intimate set: following directed edges from any edge reaches a sink v with a parent; a sink has no children and hence no co-parents, so its neighbors in $M(D)$ are exactly $\mathrm{pa}_D(v)$, which form a clique, and $\Psi_G(v) = \mathrm{pa}_D(v) \neq \varnothing$. A second consequence needs no equality: if D is a directed chain, then G is a subgraph of a path and hence a forest, and any nonempty forest has a leaf whose only neighbor is intimate. Thus no chain with a nonempty G, including the chain of Figure 2, satisfies the all-node condition $\Psi_G(i) = \varnothing$ for every i that defines the full-correspondence case, and any example with a nonempty G satisfying that condition exhibits a conditional independence that its generating

[2]That conference version uses in its Theorem 1 the window $c_\tau = (z_{\tau-1}, z_\tau)$ with density $p(c_\tau \mid z_{\tau-2})$ and a $4n + |\mathcal{E}(\mathcal{M}_{c_\tau})|$ variability vector, whereas the arXiv v4 formulation (Li et al., 2026) uses a single-transition condition, as here. For this first-order Markov process, with $v = z_{\tau-2}$, $u = z_{\tau-1}$, and $x = z_\tau$, $\log p(u, x \mid v) = \log K(u \mid v) + \log K(x \mid u)$, so $\partial_{v_m}\partial_{x_i} \log p(u, x \mid v)$ and $\partial_{v_m}\partial^2_{x_i} \log p(u, x \mid v)$ vanish identically: the window vector has zero rows and cannot satisfy its full-rank premise.

structure does not imply. This restricts that all-node condition, not every route to full identifiability; the ranges U_Ψ and U_{sep} are determined by neighborhood structure. The four-cycle below realizes such an unimplied independence through an exact cancellation, and the five-chain of Appendix E shows that the partial guarantee needs none.

H.2 Cancellation in the fixed four-cycle

Remark 2 (The fixed four-cycle). The kernel family of Appendix H.3 reduces at $u = 0$ and $a = 0$ to the fixed precision matrix

$$Q = \begin{pmatrix} 4 & -1 & 0 & -1 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ -1 & 0 & -1 & 4 \end{pmatrix}, \tag{H.1}$$

which is generated, for independent Gaussian innovations $\varepsilon_1, \dots, \varepsilon_4$ with variances $7/24$, $15/56$, $4/15$, and $1/4$, by the acyclic linear model

$$\begin{aligned} X_1 &= a + \varepsilon_1, & X_2 &= \tfrac{2}{7} X_1 + \varepsilon_2, \\ X_3 &= \tfrac{1}{15} X_1 + \tfrac{4}{15} X_2 + \varepsilon_3, & X_4 &= \tfrac{1}{4} X_1 + \tfrac{1}{4} X_3 + \varepsilon_4. \end{aligned} \tag{H.2}$$

Gaussian conditional independence is read off the zero pattern of Q (Lauritzen, 1996, Chapter 5), so G is the chordless cycle $1{-}2{-}3{-}4{-}1$ and every intimate set is empty, whereas the moral graph of the generating structure, with edges $1 \to 2$, $1 \to 3$, $2 \to 3$, $1 \to 4$, $3 \to 4$, contains the chord $1{-}3$. The chord is absent from G because

$$Q_{13} = \underbrace{-\frac{1/15}{4/15}}_{\text{edge } 1\to 3} + \underbrace{\frac{(1/4)(1/4)}{1/4}}_{\text{common child } 4} = -\tfrac{1}{4} + \tfrac{1}{4} = 0, \tag{H.3}$$

so $X_1 \perp X_3 \mid (X_2, X_4, c)$ although $1 \to 3$ is a structural edge; the sink 4 has parents $\{1, 3\}$, adjacent in $M(D)$ but not in G, which is where the sink argument of Appendix H.1 needs $G = M(D)$. The cancellation is an equality constraint in the structural parameterization: replacing the coefficient $1/15$ of X_1 in the third equation by $1/15 + r$ gives

$$Q_{13} = -\tfrac{15}{4}\, r, \tag{H.4}$$

and any nonzero r restores the chord and turns the neighbor sets of nodes 2 and 4 into cliques. The cancellation stays exact within a parameterization that fixes the zero pattern of the precision matrix, which the family below does while also supplying the history variation that this fixed covariance lacks.



Figure 6: **Generating DAG and union graph of the fixed four-cycle.** Left: the DAG D of Eq. (H.2), whose moral graph contains the chord 1–3. Right: the union graph G is the chordless cycle; the chord (dashed) is absent because the terms from the edge $1 \to 3$ and from the common child 4 cancel in Eq. (H.3). Every intimate set of G is empty.

H.3 The history-dependent kernel family

We now vary the kernel with history while keeping the four-cycle's precision zeros exact. Let $n_d = 4$, let the dynamic domain be $\mathbb{R}^4$ and the action domain $(-1, 1)$, and take $n_s = 0$. Write $t_m = \tanh u_m$,

$$F_r(u) = \frac{1}{r} \sum_{m=1}^{4} t_m^r, \quad r = 1, \dots, 12, \qquad \delta = \tfrac{1}{100},$$

and enumerate the edges $e_1 = \{1, 2\}$, $e_2 = \{2, 3\}$, $e_3 = \{3, 4\}$, $e_4 = \{4, 1\}$. Define the symmetric matrix $Q(u)$ and the vector $r(u)$ by

$$\begin{aligned} Q_{ii}(u) &= 4 + \delta F_{4+i}(u), & Q_{e_j}(u) &= -1 + \delta F_{8+j}(u), \\ Q_{13}(u) &= Q_{24}(u) = 0, & r_i(u) &= \delta F_i(u). \end{aligned} \tag{H.5}$$

Since $|F_r| \le 4/r$, the diagonal entries exceed 3.96 and each nonzero off-diagonal entry has magnitude at most 1.04, with two per row; $Q(u)$ is uniformly positive definite with spectrum in $[1.88, 6.12]$, the four cycle entries stay strictly negative, and G is the same chordless four-cycle at every history. Let $\Sigma(u) = Q(u)^{-1}$, $\kappa(u) = 1/\Sigma_{11}(u)$, and $\alpha(u, a) = r(u) + a\,\kappa(u)\,e_1$, and take the Gaussian transition kernel

$$K_a(x \mid u) \propto \exp\Big\{-\tfrac{1}{2}x^{\mathsf{T}}Q(u)x + \alpha(u, a)^{\mathsf{T}}x\Big\}, \tag{H.6}$$

whose mean and covariance are

$$\mu(u, a) = \Sigma(u)r(u) + a\,\frac{\Sigma(u)e_1}{\Sigma_{11}(u)}, \qquad \mathrm{Cov}(X \mid u, a) = \Sigma(u). \tag{H.7}$$

H.4 Independent innovations, and the action enters only the root

Relative to $a = 0$, the kernel (H.6) is multiplied by $\exp(a\kappa(u)x_1)$ and renormalized. Factorizing in the order $1, 2, 3, 4$, the conditional kernels $p(x_i \mid x_1, \dots, x_{i-1}, u, a)$ for $i \ge 2$ do not depend on a, because the tilt cancels between numerator and denominator, while $X_1 \mid u, a \sim N(\mu_{0,1}(u) + a, \Sigma_{11}(u))$. Realizing each Gaussian conditional with an independent standard normal innovation gives time-homogeneous acyclic structural equations whose coefficients depend smoothly on the history; the innovations remain independent and identically distributed, and only the root's equation reads the action. The nonzero structural pattern can be taken as in Eq. (H.2): eliminating later nodes produces only the fill-in $1-3$, all precision edges are negative, and the corresponding regression coefficients are positive.

H.5 Global history variation with $q = 12$

Here $q = 2n_d + |\mathcal{E}(G)| = 8 + 4 = 12$. Fix a history variable m, a current point x, the other history variables u_{-m}, and an action a, and write $t = \tanh u_m$. With $\ell = \log K_a(x \mid u)$, differentiate the vector of diagonal curvatures, scores, and edge cross-curvatures in u_m. The entries of the resulting vector w_m are

$$\begin{aligned} \ell_{x_i x_i u_m} &= -\partial_{u_m} Q_{ii}, \\ \ell_{x_i u_m} &= \partial_{u_m}\alpha_i - \sum_j x_j\,\partial_{u_m} Q_{ij}, \\ \ell_{x_i x_j u_m} &= -\partial_{u_m} Q_{ij} \quad (\{i, j\} \in \mathcal{E}(G)). \end{aligned}$$

Subtracting x_j times the corresponding curvature and cross-curvature rows from the score rows removes every term $x_j\partial_{u_m}Q_{ij}$; this row operation is invertible and depends only on the fixed x, not on t. Removing the positive factor $dt/du_m = 1 - t^2$ from each column and reordering rows and signs leaves the function vector

$$v(t) = \big(\delta + a\,\kappa_t(t),\ \delta t,\ \delta t^2,\ \dots,\ \delta t^{11}\big)^{\mathsf{T}}, \tag{H.8}$$

where κ_t denotes the derivative of κ with respect to t. Because Q uses only $F_5, \dots, F_{12}$, $\partial_t Q(0) = 0$ and hence $\kappa_t(0) = 0$, so $v(0) = (\delta, 0, \dots, 0)^{\mathsf{T}}$. Choose $t_b = b/12$ for $b = 0, \dots, 11$, that is, $u_m = \operatorname{artanh}(t_b)$. Expanding the determinant of the 12×12 matrix along the column $b = 0$ removes every term containing $a\kappa_t$; the remaining minor is a Vandermonde matrix in $t_1, \dots, t_{11}$, and, with the row operations and signs just described, the derivative evaluation matrix W_m satisfies

$$\det W_m = \delta^{12}\Big[\prod_{b=0}^{11}(1 - t_b^2)\Big]\Big[\prod_{b=1}^{11} t_b\Big]\Big[\prod_{1 \le i < j \le 11}(t_j - t_i)\Big] > 0. \tag{H.9}$$

The expression does not depend on x, u_{-m}, a, or m, so full derivative span holds at every action and complementary history with one common set of history values, and Lemma 5(ii) yields Assumption 2.

H.6 STATIONARITY, ROLE LABELS, AND MINIMUM CANDIDATES

Stationarity, support, and behavior policy. Take the behavior policy

$$p(a \mid u) = \frac{\exp(a \tanh u_1)}{\int_{-1}^{1} \exp(b \tanh u_1)\, db}, \qquad -1 < a < 1, \tag{H.10}$$

a common, smooth, everywhere positive action density that uses dynamic factor 1 only and can be generated from independent uniform private noise by the inverse distribution function. The spectrum of $Q(u)$ lies in $[1.88, 6.12]$, and the mean in Eq. (H.7) satisfies $\|\mu(u,a)\| \le \|\Sigma r\| + |a|\,\|\Sigma e_1\|/\Sigma_{11} \le 0.0254 + 6.12/1.88 < 4$, using $\|r\| \le \delta\big(\sum_{i\le 4} 16/i^2\big)^{1/2} < 0.048$, $\|\Sigma\| \le 1/1.88$, and $\Sigma_{11} \ge 1/6.12$. Lemma 4 applies with $\lambda = 1.88$, $\Lambda = 6.12$, and $M = 4$; writing ρ for the invariant law, at any fixed action $p(u \mid a) \propto p(a \mid u)\rho(u)$ is positive on $\mathbb{R}^4$, so the full range of history values is available. The identity observation map satisfies Assumption 1.

Role sets and mediated responses. The policy of Eq. (H.10) uses factor 1 only; since $\tanh$ is strictly monotone and the full state density is positive, the conditional action distribution varies with u_1 at any fixed value of the other coordinates, so read-faithfulness holds and $S_{\mathrm{read}} = \{1\}$. Writing $Q = D_Q - W$ with D_Q diagonal and $W \ge 0$, strict diagonal dominance and the connected cycle make the Neumann series $Q^{-1} = \sum_{k\ge 0} (D_Q^{-1} W)^k D_Q^{-1}$ convergent with every entry of Σ strictly positive, hence

$$\partial_a \mu_i(u,a) = \frac{\Sigma_{i1}(u)}{\Sigma_{11}(u)} > 0 \qquad \text{for every } i \text{ and every } u. \tag{H.11}$$

Every Gaussian marginal responds to the action, so $S_{\mathrm{effect}} = \{1,2,3,4\}$ and $S_{\mathrm{read}} \cap S_{\mathrm{effect}} = \{1\}$. Since only the root's structural kernel reads the action, the responses of factors 2–4 are transmitted through contemporaneous dependencies. Theorem 1 identifies this effect membership and nothing about its transmission. Each marginal has action-independent support $\mathbb{R}$, positive action-independent variance $\Sigma_{ii}(u)$, and a mean affine in a, verifying the effect part of Assumption 3. Read-faithfulness holds while faithfulness of the contemporaneous structure fails.

Minimum candidates and full correspondence. Lemma 2 gives an all-candidate lower bound of four edges for this process. The identity encoder belongs to the class in Assumption 1 and has four edges, so it attains the minimum. For every selected candidate, $U_\Psi = U_{\mathrm{sep}} = [4]$ gives full scalar correspondence by Lemma 3 and both complete set equalities by Theorem 1.

I RELATED WORK

Multiview representation learning characterizes when contrastive or alignment objectives recover shared, invariant, or selected latent information from paired observations (Radford et al., 2021; Zimmermann et al., 2021; von Kügelgen et al., 2021; Daunhawer et al., 2023; Yao et al., 2024). In the static, invertible-readout regime, mask-based alignment can block-identify the modality-selected visual subspace (Xie et al., 2025). Our setting uses a lossy readout that precludes exact cross-modal alignment, placing us outside this regime; we use temporal structure to derive partial variable-level role guarantees.

Temporal component identification uses a varying context to resolve nonlinear representations. Auxiliary-variable ICA develops this principle through segment labels, general contexts, and conditional latent models (Hyvärinen & Morioka, 2016; Hyvärinen et al., 2019; Khemakhem et al., 2020), while temporally disentangled representation learning applies it to lagged dynamics (Yao et al., 2022). Lippe et al. (2022) and, with instantaneous dependencies, Lippe et al. (2023) require per-variable intervention targets, whereas Li et al. (2025) obtain component-wise identification and Markov-network isomorphism without those labels. Unlike a per-variable intervention target, our logged action serves as an observed conditioning context, with histories and actions jointly supplying identifying variation. Theorem 1 adapts graph-based identification to give role-specific partial guarantees for this conditional law.

Latent-action world models infer action codes from unlabeled video for interactive generation and downstream control (Schmidt & Jiang, 2024; Bruce et al., 2024; Ye et al., 2025). Empirical and

analytical studies show that such codes can mix action-related signals with distractors or exogenous variation (Nikulin et al., 2025; Zhang et al., 2025), while factored dynamics decomposes multi-entity action spaces (Wang et al., 2026). In this lineage, Lachapelle (2025) identifies an unobserved action code through entropy-regularized inverse dynamics; we take the action as observed and identify read and effect labels on graph-determined subsets of dynamic visual variables. Mechanism-sparsity identification (Lachapelle et al., 2026) also treats actions as observed, with partial and component guarantees and an action-influence graph, but its transition model excludes instantaneous latent effects, which the union graph admits (Appendices E and H). Delay-aware control modifies the decision process to accommodate delayed execution or observation (Chen et al., 2021; Wikman et al., 2026); our setting treats action–frame timing as a modeling premise (Section 2).

Action-conditioned world models use logged controls for predictive simulation and planning at scale (Hafner et al., 2025; Zhu et al., 2025; Wang et al., 2025). Concurrent work by Zhang et al. (2026) studies identifiability in controlled world models, assuming linear-Gaussian dynamics and invertible observations; under policy-dependent excitation, they identify the full state and controlled conditional mean up to a shared orthogonal transform. Adjacent work spans identifiable action- and reward-defined state blocks (Liu et al., 2023), task-relevance objectives that remove distractors (Wang et al., 2022), and inverse-dynamics biases toward controllable factors (Ivashkov et al., 2026). We take a different route: history–action variation and a minimum-edge union graph restrict nonlinear mixing, yielding effect labels for individually recovered variables and read labels for a separated subset. These targets guide the variable-level audit of a fixed encoder.

J Method and audit implementation details

This appendix specifies the controlled implementation of Section 4: the model, every training term with its coefficient, the two audit estimators, the configurations used in the three studies, and the ideal read-subset criterion under full correspondence. The controlled observation x_τ is a nonlinear mixing of the latent state rather than an image, redrawn by data seed and standardized on training frames before encoding; the process is specified in Appendix K.1. Throughout, a squared norm denotes a mean over the batch and over the coordinates of its argument. Model quantities are written for one transition pair, whose earlier step is $\tau - 1$.

J.1 The controlled model

Encoder f_θ and decoder h_θ are mirrored multilayer perceptrons with two $\tanh$ hidden layers of width 64, and the learned representation $\hat{z}_\tau = f_\theta(x_\tau) = (\hat{s}_\tau, \hat{d}_\tau)$ has $n_z = n_s + n_d$ coordinates, $n_s = 1$ and $n_d = 5$ in the reported runs. Two gate vectors $\beta^{\mathrm{r}} = \sigma(\ell^{\mathrm{r}})$ and $\beta^{\mathrm{e}} = \sigma(\ell^{\mathrm{e}})$ in $(0,1)^{n_d}$ are free parameters initialized at logit zero. The read head r_η is a $\tanh$ multilayer perceptron with one hidden layer of width 32, and the transition base b_ϕ has the same shape but takes the complete representation:

$$\hat{a}_{\tau-1} = r_\eta(\hat{d}_{\tau-1} \odot \beta^{\mathrm{r}}), \qquad \breve{a}_{\tau-1} = a_{\tau-1} - \mathrm{sg}(\hat{a}_{\tau-1}), \qquad \tilde{d}_\tau = b_\phi(\hat{z}_{\tau-1}) + \beta^{\mathrm{e}} \odot (W_{\mathrm{e}} \breve{a}_{\tau-1}),$$

with $W_{\mathrm{e}} \in \mathbb{R}^{n_d \times n_a}$ and no bias, so that a zero residual action contributes nothing to the predicted mean. The read head is evaluated once per pair, and the same $\hat{a}_{\tau-1}$ serves as the regression prediction and as the detached subtrahend. The stop-gradient blocks the path from the transition and next-observation terms back through $\hat{a}_{\tau-1}$, so the read head's own parameters receive gradient from the read regression alone; those terms still train the shared encoder, and so can change the read head's input. The read gate β^{r} is shaped in addition by the alignment and gate-sparsity terms. The predicted next observation is $h_\theta\Big([\hat{s}_{\tau-1}, \tilde{d}_\tau]\Big)$, so the static block is carried forward unchanged.

At a fixed current representation, the read prediction is constant with respect to the action. Subtracting it therefore leaves the predicted mean's action derivative equal to $\mathrm{diag}(\beta^{\mathrm{e}}) W_{\mathrm{e}}$. This coefficient describes the fitted mean response; it does not distinguish direct from mediated responses.

J.2 Training terms and coefficients

The fit terms are squared errors on the observation and its one-step prediction,

$$\mathcal{L}_{\mathrm{rec}} = \tfrac{1}{2} \left\| h_\theta(\hat{z}_{\tau-1}) - x_{\tau-1} \right\|^2 + \tfrac{1}{2} \left\| h_\theta(\hat{z}_\tau) - x_\tau \right\|^2, \qquad \mathcal{L}_{\mathrm{pred}} = \left\| h_\theta\Big([\hat{s}_{\tau-1}, \tilde{d}_\tau]\Big) - x_\tau \right\|^2,$$

Table 4: Terms of the controlled objective and their coefficients, grouped as in Section 4.1. These are the configuration values of our model's full setting; the same-architecture ablation arms set individual coefficients to zero, as listed in Appendix K.1. The groups carry no weight of their own.

Group	Terms (coefficient)
$\mathcal{J}_{\text{fit}}$	$\mathcal{L}_{\text{rec}}$ (0.5), $\mathcal{L}_{\text{trans}}$ (1.0), $\mathcal{L}_{\text{pred}}$ (1.0)
$\mathcal{J}_{\text{act}}$	$\mathcal{L}_{\text{read}}$ (1.0), $\mathcal{L}_{\text{align}}$ (1.5)
$\mathcal{J}_{\text{stat}}$	$\mathcal{L}_{\text{stat}}$ (0.5)
$\mathcal{J}_{\text{reg}}$	$\mathcal{L}_{\text{var}}$ (0.05), $\mathcal{L}_{\text{cov}}$ (0.02), $\mathcal{L}_{\text{gate}}$ (0.002)

together with a Gaussian negative log-likelihood for the transition, using a learnable diagonal log-variance $\varsigma \in \mathbb{R}^{n_d}$ clamped to $[-5, 3]$:

$$\mathcal{L}_{\text{trans}} = \frac{1}{2n_d} \sum_{k=1}^{n_d} \left[e^{-\varsigma_k} \left(\hat{d}_{\tau,k} - \tilde{d}_{\tau,k} \right)^2 + \varsigma_k \right].$$

Its target $\hat{d}_\tau$ is the encoder output at step τ and is not detached, so this term also shapes the encoder. The two action terms are the read regression $\mathcal{L}_{\text{read}} = \|\hat{a}_{\tau-1} - a_{\tau-1}\|^2$ and an alignment term $\mathcal{L}_{\text{align}}$ that sends the gated dynamic block and the action into $\mathbb{R}^{32}$ through affine maps, $q^v = \text{norm}(W_v(\hat{d}_{\tau-1} \odot \beta^{\text{r}}) + b_v)$ and $q^a = \text{norm}(W_a a_{\tau-1} + b_a)$, then applies symmetric InfoNCE over the batch at temperature 0.1 with the paired step as the positive. Static consistency is $\mathcal{L}_{\text{stat}} = \|\hat{s}_{\tau-1} - \hat{s}_\tau\|^2$, with both sides differentiated. The regularizers are computed on the batch concatenation of $\hat{z}_{\tau-1}$ and $\hat{z}_\tau$,

$$\mathcal{L}_{\text{var}} = \frac{1}{n_z} \sum_j \left(\sqrt{\hat{\sigma}_j^2 + 10^{-4}} - 1 \right)^2, \qquad \mathcal{L}_{\text{cov}} = \frac{1}{n_z^2} \sum_{j \neq j'} \hat{\Sigma}_{jj'}^2,$$

$$\mathcal{L}_{\text{gate}} = \frac{1}{n_d} \sum_k \beta_k^{\text{r}} + \frac{1}{n_d} \sum_k \beta_k^{\text{e}},$$

where $\hat{\sigma}_j^2$ is the second moment of coordinate j about the batch mean and $\hat{\Sigma}$ the sample covariance, normalized by n and $n-1$ respectively.

Optimization uses AdamW at learning rate 1.5×10^{-3}, weight decay 10^{-4}, batch size 16, gradient-norm clipping at 5, and 200 epochs. The variance and covariance penalties discourage collapse and linear redundancy among the coordinates.

Two differences from the ideal criteria explain why role membership is evaluated on the frozen coordinates rather than read from the fitted heads. First, the graph in the minimum-edge criterion is defined by the actual encoded conditional law, whereas the fitted Gaussian density has a diagonal, state- and action-independent covariance, and the two gates weight predictor inputs and mean-response outputs. Second, the read head predicts the action mean with MSE and sigmoid weights on gated dynamic inputs, whereas the ideal hard-subset criterion of Appendix J.5 (Corollary 1 and Eq. (J.5)) selects the smallest subset that preserves the conditional action law; that appendix also gives a policy whose dependence mean prediction misses.

J.3 Audit estimators

Read. The read score uses the Frenzel–Pompe conditional variant of the KSG estimator with $k = 5$ nearest neighbors under the Chebyshev metric, evaluated on the frozen standardized representation with per-dimension tie-breaking jitter. The operating threshold is $\delta_r = 0.02$, a configuration constant rather than a per-run calibration.

Effect. The three feature maps are built from standardized inputs:

Table 5: Readouts used in the experiments. Thresholds and normalizations are specific to each experiment. Controlled role scores are computed on every learned coordinate; matching to simulator factors is used only for recovery metrics. Encoded response magnitude is a continuous diagnostic.

Experiment / object	Readout	Setting / definition
Controlled Read set	KSG conditional mutual information on frozen coordinates	$k = 5$, $\delta_r = .02$; fixed empirical cutoff
Controlled Effect set	Cross-fitted conditional-feature dependence	$\delta_e = .077721$; calibration in Appendix K.1
Encoded mean response	Across-state RMS paired-action contrast, normalized by training SD	Continuous; Figure 4b and Appendix K.3
Response detection	Unconditional distribution, conditional mean, and conditional distribution statistics	Separate null cutoff for each checkpoint/statistic; Appendix K.3
Fixed RGB roles and utility	Conditional-association set estimates; action-readout and response NMSE	Re-estimated in each mechanism setting; Appendix L.1
Robot-log decoding	All-coordinate gain over state-only R^2; selected-minus-random-coordinate R^2	Development-selected sets and probes; Appendix M

Input	Features
State $c = \hat{z}_{\tau-1}$	$[\,1,\ c,\ c \odot c,\ \sin(Wc+\varphi),\ \cos(Wc+\varphi)\,]$ with 12 random Fourier directions
Action $a = a_{\tau-1}$	$\Phi(a) = [\,a,\ a \odot a,\ \sin(W'a+\varphi'),\ \cos(W'a+\varphi')\,]$ with 6 random directions
Next coordinate $y = \hat{z}_{\tau,j}$	$\Psi(y) = [\,y,\ y^2,\ (\sin\omega y, \cos\omega y)_{\omega\in\{0.5,1,2\}}\,]$: the fixed pair (y, y^2) and six terms from three fixed frequencies, eight entries

The random directions are regenerated from one fixed seed on every call, so they are identical across runs at a given conditioning width. Nuisance regressions are ridge fits with penalty 10^{-3} on two folds formed by index parity, predicting $\Phi(a_{\tau-1})$ and each $\Psi(\hat{z}_{\tau,j})$ from the features of $c_{\tau-1}$. With $\mathcal{T}$ indexing the $N = |\mathcal{T}|$ sampled transitions, $b(\tau)$ the fold containing τ, and $\hat{\mu}^{(-b(\tau))}$ the fit on the other fold, the out-of-fold residuals and the score of Section 4.2 are

$$u_\tau = \Phi(a_{\tau-1}) - \hat{\mu}_a^{(-b(\tau))}(c_{\tau-1}), \qquad v_{\tau,j} = \Psi(\hat{z}_{\tau,j}) - \hat{\mu}_j^{(-b(\tau))}(c_{\tau-1}),$$

$$\hat{e}_j = \max_{p,q} \left| N^{-1} \textstyle\sum_{\tau\in\mathcal{T}} \bar{u}_{\tau,p}\, \bar{v}_{\tau,j,q} \right|, \qquad \hat{S}^{\text{est}}_{\text{effect}} = \{ j \in [n_z] : \hat{e}_j > \delta_e \},$$

where bars denote per-feature standardization of the residuals and $[n_z] = \{1, \ldots, n_z\}$. Because the standardization uses the uncorrected second moment, each entry of that maximum is a Pearson correlation. In the oracle coordinates of the controlled process the transition is Gaussian given the preceding state and action, so mean and variance determine the conditional marginal law and the Fourier features extend coverage after component-wise warps; on learned coordinates the score remains a scoped dependence statistic rather than a universal conditional-independence test. With exact conditional feature means and finite feature second moments, the population residual cross moment equals the average conditional covariance, so contributions with opposite signs at different states can cancel; a small score can also reflect limited feature coverage, and imperfect nuisance fits can leave state-related dependence in the residuals. The scores do not test whether a marginal action derivative vanishes everywhere. The frozen value of δ_e appears in Table 5 and its construction in Appendix K.1.

Audit sample. Both scores are computed on a frozen encoder, with representation standardization fitted on training sequences only. The read score samples current-frame records; the effect score samples paired records that have a successor. Both draw from the full record rather than a held-out split, which is used by the prediction probe alone, so the reported set estimates are pooled-sample diagnostics on a frozen representation; their generalization to unseen sequences is not evaluated.

Table 5 lists each readout used in the experiments, with its setting and where it is defined.

J.4 Study-specific instantiations

Table 6 distinguishes representation learning from set estimation and subsequent probe fitting. The fixed RGB study fits sets and prediction heads on a common representation; the controlled and robot-log studies also learn representations.

Table 6: Representation, fitting, and evaluation in the three studies. Study-specific settings are given at the indicated appendix entries.

Study	Representation	Fitting	Evaluation units
Controlled	6 learned coordinates, evaluated against 1 static and 5 dynamic factors	Table 4; sequence-level 75/25 train/test split; common post-hoc role scores	Five data seeds per method/DGP; pooled samples for set estimation and held-out prediction (Appendix K)
Fixed RGB	Six position/velocity coordinates from three objects, with scene, color, and single-axis-motion priors	Calibration for set estimation in each of four settings; prediction heads with 128–4096 fit windows per setting	Five seeds, each averaging four settings; one pooled matching per seed (Appendix L.1)
BridgeData V2	Frozen visual features projected to 32 coordinates; probes read both frames and current proprioceptive state	Representation-specific objectives; development-selected sets and ridge probes; 50 equal-size random subsets	Five seeds per method on a common episode split (Appendix M)

Simulator labels do not enter the controlled representation-training losses; they determine recovery scores only. The RGB correspondence uses simulator coordinates for evaluation, whereas BridgeData has no factor-level membership labels.

BridgeData representation adapters. The adapters project frozen CLIP ViT-B/32 features from two frames, except for the single-frame Static baseline. Our model separates state dynamics from a gated action term; Temporal uses a joint transition. CPC uses a contrastive objective, IDOL adds a transition-Jacobian penalty, and the variational baselines use action-conditioned (iVAE), lagged (TDRL), or standard-normal (β-VAE) priors. The Random encoder stays at initialization. Default adapters use hidden width 128 and 6,000 AdamW steps; the tuned model uses width 192 and 8,000 steps, selected by development-set decoding. Its gate-sparsity and future-difference penalties are zero. Appendix M specifies the common set-selection and decoding procedures.

J.5 Ideal read-subset selection under full correspondence

Under full correspondence, Eq. (C.7) recovers every dynamic coordinate up to a permutation π and static-dependent scalar maps, and an ideal read-subset criterion follows. The criterion retains the entire recovered static block and asks which dynamic subsets preserve the conditional action law. A dynamic subset $J \subseteq [n_d]$ is *read-sufficient* when $P_{A|\hat{S},\hat{D}} = P_{A|\hat{S},\hat{D}_J}$ almost surely; its omission gap is

$$\Delta_{\mathrm{read}}(J) := I(A; \hat{D}_{-J} \mid \hat{S}, \hat{D}_J) = \mathbb{E}\, D_{\mathrm{KL}}\Big(P_{A|\hat{S},\hat{D}} \,\Big\|\, P_{A|\hat{S},\hat{D}_J}\Big), \tag{J.1}$$

and read sufficiency is equivalent to $\Delta_{\mathrm{read}}(J) = 0$.

Corollary 1 (Minimum read-sufficient subset)**.** *Suppose the private-noise policy model, the full component correspondence in Eq.* (C.7)*, and read-faithfulness hold. With $S^\star = \pi(S_{\mathrm{read}})$,*

$$\{J : \Delta_{\mathrm{read}}(J) = 0\} = \{J : S^\star \subseteq J\}. \tag{J.2}$$

Thus $S^\star$ is the unique inclusion-minimal and unique minimum-cardinality read-sufficient dynamic subset.

Proof. Use the fixed-time variables (S, D, A) and $S_0 = S_{\mathrm{read}}$ of Appendix C.2, with $\hat{S} = b(S)$ and $\hat{D}_{\pi(k)} = T_{k,S}(D_k)$; since $\hat{S}$ determines S, each $T_{k,S}$ is a fixed invertible map given $\hat{S}$, and regular conditional laws exist on the open Euclidean domains. By Eq. (C.4) the action law is measurable with respect to (S, D_{S_0}), hence with respect to $(\hat{S}, \hat{D}_{S^\star})$ after the invertible recoding, so $S^\star$ is sufficient, and by the tower property so is every superset of $S^\star$. Conversely, let J be sufficient and $j \notin J$. For every action event B, sufficiency makes $\mathbb{E}[\mathbf{1}_{\{A\in B\}} \mid \hat{S}, \hat{D}]$ measurable with respect to $(\hat{S}, \hat{D}_J)$, hence with respect to $(\hat{S}, \hat{D}_{-j})$, and the tower property identifies it with $\mathbb{E}[\mathbf{1}_{\{A\in B\}} \mid \hat{S}, \hat{D}_{-j}]$; a countable determining family of events gives $A \perp \hat{D}_j \mid (\hat{S}, \hat{D}_{-j})$. If $j = \pi(k)$ with $k \in S_{\mathrm{read}}$, this contradicts read-faithfulness through Eq. (C.5). Hence every sufficient J contains $S^\star$, which proves Eq. (J.2) and both uniqueness statements, including the case $S^\star = \varnothing$. □

The same premises give an ideal conditional-KL selection. Write $X = (\hat{S}, \hat{D})$ and $X_J = (\hat{S}, \hat{D}_J)$, assume a common sigma-finite dominating measure for the action conditional densities

and $\Delta_{\mathrm{read}}(J) < \infty$ for every subset, and for every hard subset J, let $\mathcal{Q}_J$ be a conditional-density class reading only X_J and containing the true law $P_{A|X_J}$. Define

$$\mathcal{E}_{\log}(J) = \inf_{Q \in \mathcal{Q}_J} \mathbb{E}\, D_{\mathrm{KL}}(P_{A|X} \| Q_{A|X_J}). \tag{J.3}$$

The relative-entropy chain rule gives

$$\mathbb{E}\, D_{\mathrm{KL}}(P_{A|X} \| Q_{A|X_J}) = \Delta_{\mathrm{read}}(J) + \mathbb{E}\, D_{\mathrm{KL}}(P_{A|X_J} \| Q_{A|X_J}), \tag{J.4}$$

whose second term is nonnegative and vanishes at the admissible true conditional law. Thus $\mathcal{E}_{\log}(J) = \Delta_{\mathrm{read}}(J)$ and

$$\operatorname*{arg\,min}_{J : \mathcal{E}_{\log}(J) = 0} |J| = \{S^\star\}. \tag{J.5}$$

Well-specification is required separately for every subset; a fixed neural readout family, an MSE objective, or a sample-dependent sigmoid mask need not satisfy this premise. For example, $A = DE$ with $D \in (1, 2)$ and independent centered Gaussian E has a constant conditional mean but a D-dependent conditional variance, so mean preservation would miss this policy dependence.

K Additional Controlled Numerical Results

Appendix K.1 specifies the main comparison: its process, baselines, evaluation, and Effect threshold. Appendix K.2 gives the complete results behind Table 1, and Appendix K.3 the protocols behind Figure 4. Appendix K.4 repeats the comparison with the temporal backbone on three more topologies and reports a matched ablation of the read objective. Appendix K.5 asks which Effect errors a larger audit sample removes and which remain on aligned learned coordinates.

K.1 Setup, baselines, and evaluation of the main comparison

Shared-action process. The state has five dynamic latents and one sequence-static latent, and the action is two dimensional. With dynamic state d_t and logged action a_t, the shared-action process is

$$d_0 = \sigma_z \eta_0, \tag{K.1}$$

$$a_t = \tanh(W^\top d_t) + \sigma_a \varepsilon_t, \tag{K.2}$$

$$d_{t+1} = \lambda \tanh(B d_t) + \sigma_z \eta_{t+1} + \gamma C a_t. \tag{K.3}$$

The rows of W are nonzero only on the read set $\{0, 2\}$ and the rows of C only on the effect set $\{1, 3\}$. The same realized action is therefore both a noisy current-state readout and an input to a distinct part of the next-state transition. The mixing map, action matrices, and noise are redrawn by data seed; the true sets are fixed. The shared-action process, its instantaneous variant, and the three topology generators below all use fixed innovation covariances and therefore fail Assumption 2, as Section 5.1 states.

Table 7: Specification of the shared-action process used for the recovery study in Section 5.1.

Quantity	Shared-action process
Dynamic / static latents	$5/1$
Action channels	one shared, dim. 2
Read / effect set	$\{0, 2\}/\{1, 3\}$
Sequence length / sequences	$80/32$
Readout noise / effect gain	$0.55/0.80$
Autoregression / innovation	$0.45/0.70$
Mixing depth / slope	$1/0.80$
Design	5 seeds $\times$ 1 final checkpoint
Runs / evaluation samples	5 per method/DGP / 2400

Compared methods. All eight learned methods in Table 1 share the data split, representation dimension, matching rule, and post-hoc evaluator; each method is trained on five data seeds per DGP and evaluated at its final checkpoint. Our model and its two 200-epoch same-architecture controls differ by removing the interface objectives (temporal backbone) or the predictive objectives (static alignment). TDRL, IDOL, SlowVAE, IDOL+A, and Mechanism Sparsity use fixed 100-epoch adapters; the two action-aware adapters are specified below.

Action-aware comparison methods. Table 1 includes two additional models that condition their transition on the logged action. IDOL+A augments the IDOL transition prior (Li et al., 2025) with the aligned continuous action, retaining its latent Jacobian sparsity penalty; it is our action-conditioned variant, not the original method. The Mechanism Sparsity adapter uses the released core model (Lachapelle et al., 2026), jointly learns temporal and action masks, and applies penalty regularization. Its regularization coefficients are selected from three fixed candidates using validation negative ELBO on separate development seeds, without latent or membership labels. Both encoders receive observations only and return six coordinates.

The two models train for 100 epochs with batch size 64 on the same five seeds per DGP and use the final checkpoint; observation and action normalization uses training sequences only. IDOL+A uses 384 length-six training windows and 600 optimizer steps per run; Mechanism Sparsity uses 1,896 training transitions and 2,900 steps. Their parameter counts are 312,864 and 8,968,349. Evaluation reuses the data splits, full-space matching, pooled samples for set estimation, and frozen thresholds of Table 1. Both DGPs fall outside Mechanism Sparsity's identification assumptions: its sufficient-influence and graph conditions and, under instantaneous dependencies, its factorized-transition model.

Evaluation and uncertainty. Post-hoc readouts measure set recovery, while a common ridge probe from current coordinates and action to the next observation measures predictive fit; Table 5 in Appendix J.3 lists each readout with its object, its setting, and where it is defined. Learned coordinates are matched one-to-one to simulator variables by Hungarian assignment, and the mean correlation coefficient (MCC) over dynamic variables measures this variable matching; simulator membership labels are used for evaluation, not for training. An Effect F_1 of $.500$ can result from selecting all six coordinates when the true set has two members, so the Effect columns measure thresholded set recovery rather than classification accuracy.

Throughout this appendix, tabulated absolute results are means with sample SDs over the stated seeds; paired differences carry 95% paired percentile-bootstrap intervals from 20,000 draws, without multiplicity adjustment; exact recovery is a count of successful seeds; and figure bands are defined in their captions.

How δ_e is constructed. The Effect threshold comes from a synthetic null in oracle coordinates, generated separately from the runs it is later applied to and frozen before them. The null has two arms, one with no action effect and one with the action permuted across samples, and is drawn for six calibration seeds disjoint from the evaluation seeds, giving twelve null datasets in each $(N, \text{noise}, \text{warp})$ cell. Within a dataset we take the largest score over the output coordinates; the cell threshold is the 95th percentile of these maxima across the twelve datasets. At the reported sample size $N = 2400$ and innovation scale 0.7 we take the largest cell threshold over the three warps (identity, affine, tanh), which gives $\delta_e = .077721$. Calibration uses five oracle dynamic coordinates, whereas the common evaluator scores all six learned coordinates, so the transferred cutoff does not fix the family-wise rate of null selections on the six learned coordinates at 5%.

K.2 Complete results behind Table 1

Table 8 lists every paired contrast for Table 1 in both processes. The instantaneous process adds a nonempty acyclic same-time dependence to the lag-only process and keeps the matched-seed evaluator. Our model's MCC and Read advantages carry over to it, whereas the Effect contrasts are smaller and fewer of their intervals exclude zero. Table 9 then applies a stricter endpoint: both sets must be exactly correct under the same one-to-one assignment, with a static selection counting as a false positive. No learned method attains this endpoint in either process.

Table 8: **Paired contrasts behind Table 1.** Each contrast pairs the same five data seeds within a process. ΔMSE is comparator minus our model, in units of 10^{-3}; ΔMCC, ΔRead, and ΔEffect are our model minus comparator. Positive values favor our model. Brackets give 95% paired percentile-bootstrap intervals. The oracle ceilings are the last row of Table 1.

Comparator	ΔMSE ($\times 10^{-3}$)	ΔMCC	ΔRead	ΔEffect
Lag-only				
TDRL	$+8.62$ $[-1.11, 25.60]$	$+.1027$ $[-.0090, .2423]$	$+.231$ $[.100, .349]$	$+.089$ $[.041, .140]$
IDOL	$+7.82$ $[2.45, 13.60]$	$+.0576$ $[.0176, .1072]$	$+.167$ $[-.027, .367]$	$+.136$ $[.071, .228]$
IDOL+A	$+7.71$ $[2.48, 13.67]$	$+.0614$ $[.0213, .1126]$	$+.212$ $[.027, .391]$	$+.136$ $[.071, .228]$
Mechanism Sparsity	$+.63$ $[-.43, 1.49]$	$+.0958$ $[.0145, .1623]$	$+.246$ $[.120, .357]$	$+.108$ $[.043, .183]$
SlowVAE	$+368.95$ $[248.69, 493.99]$	$+.6129$ $[.5502, .6879]$	$+.853$ $[.747, .960]$	$+.136$ $[.071, .228]$
Temporal backbone	-1.21 $[-2.27, -.16]$	$+.0457$ $[.0005, .1012]$	$+.198$ $[.026, .350]$	$+.089$ $[-.010, .172]$
Static alignment	$+43.10$ $[13.33, 90.72]$	$+.2282$ $[.1591, .3050]$	$+.225$ $[.112, .312]$	$+.033$ $[.000, .071]$
Instantaneous				
TDRL	$+8.04$ $[1.28, 18.90]$	$+.0832$ $[.0100, .1700]$	$+.293$ $[.213, .333]$	$+.021$ $[-.109, .119]$
IDOL	$+7.72$ $[1.39, 15.30]$	$+.0590$ $[.0018, .1063]$	$+.404$ $[.251, .531]$	$+.062$ $[-.024, .148]$
IDOL+A	$+8.06$ $[1.87, 15.23]$	$+.0616$ $[.0038, .1079]$	$+.431$ $[.266, .571]$	$+.062$ $[-.024, .148]$
Mechanism Sparsity	$+3.56$ $[.11, 8.00]$	$+.0881$ $[.0498, .1312]$	$+.399$ $[.251, .527]$	$+.067$ $[.000, .133]$
SlowVAE	$+346.52$ $[219.18, 445.45]$	$+.5919$ $[.5524, .6242]$	$+.960$ $[.880, 1.000]$	$+.095$ $[.043, .148]$
Temporal backbone	$-.47$ $[-2.01, 1.65]$	$+.0650$ $[.0117, .1078]$	$+.360$ $[.247, .467]$	$+.035$ $[-.053, .114]$
Static alignment	$+37.31$ $[14.06, 72.22]$	$+.2025$ $[.1486, .2452]$	$+.377$ $[.251, .497]$	$-.046$ $[-.149, .038]$

Table 9: **Exact set recovery under one shared full-space assignment.** Entries are successful seeds out of five. Joint requires simultaneous Read and Effect exactness. The seven-comparator row applies separately to TDRL, IDOL, IDOL+A, Mechanism Sparsity, SlowVAE, Temporal backbone, and Static alignment. Oracle is an evaluation ceiling. For a $0/5$ count the 95% Wilson upper bound is .434.

	Lag-only			Instantaneous		
Method	Read	Effect	Joint	Read	Effect	Joint
Ours	2/5	0/5	0/5	4/5	0/5	0/5
Seven comparators (each)	0/5	0/5	0/5	0/5	0/5	0/5
Oracle coordinates	5/5	5/5	5/5	5/5	4/5	4/5

K.3 Protocols behind Figure 4

The three panels of Figure 4 examine a prediction-preserving coordinate change, a direct response of the frozen encoder, and thresholded detection of different response laws. All use the five lag-only systems with seeds 15–19.

A: refitting after coordinate mixing. For each system we leave the true static factor unchanged and multiply the five true dynamic factors by a seed-fixed dense orthogonal matrix. The same ridge probe ($\lambda = 10^{-3}$) is refitted from current coordinates and action to the next observation in each representation, using the same sequence split and training-only normalization. Role scores use the common evaluator, unchanged simulator labels, and a full-space Hungarian assignment for each representation. Points and intervals are computed from the five paired systems; the relative MSE gap divides the absolute paired difference by the unmixed MSE.

B: direct encoded action responses. We compare our model, the temporal backbone, TDRL, IDOL, and Mechanism Sparsity using their 25 frozen Table 1 checkpoints. At each of 32 true current states, we perturb each action component separately by $\pm\sigma_a = \pm 0.55$ around its sampled logged value. The two next-state draws share the current state and innovation; 1,024 antithetic innovation draws estimate the difference between their encoded means. For each learned coordinate, we take the RMS of this mean contrast across the 32 states, then the maximum over the two action directions. Training latent SDs normalize the response, and the original full-space matching assigns factor labels; neither is refitted. The panel averages these continuous magnitudes across five systems. A response outside the true effect factors can therefore reveal mixing in the learned coordinates without involving a set-selection cutoff.

C: response laws and three statistics. The encoders of our model and the backbone remain frozen after training on the original mean-response process. New samples share current state, action, and innovation across three one-step laws. Writing $\mu_t = \lambda \tanh(Bd_t)$ and $v_t = \gamma C a_t$, their dynamic targets are

$$\begin{aligned} &\text{no effect:} && \mu_t + \sigma_z \eta, \\ &\text{mean change:} && \mu_t + v_t + \sigma_z \eta, \\ &\text{variance change:} && \mu_t + \sigma_z \exp\big(.6 \tanh(v_t)\big) \odot \eta. \end{aligned}$$

Static coordinates are copied forward. The variance law preserves the true conditional mean, although nonlinear encoding can change a learned coordinate's conditional mean.

Each statistic takes the maximum absolute empirical correlation between action and next-coordinate features, using linear, squared, and Fourier features. The unconditional statistic uses these features directly. The conditional statistics residualize both sides against features of all six current encoded coordinates by two-fold cross-fitted ridge regression ($\lambda = 10^{-3}$). Conditional mean retains only the linear target feature but all action features; conditional distribution retains the full target bank. Neither conditions on next-step coordinates.

Calibration, denominators, and null selections. For each checkpoint and statistic, 39 independent $N = 4800$ structural-null datasets give maxima over all six coordinates. The 38th ordered maximum ($\lceil 40 \times .95 \rceil$) defines the cutoff, frozen before evaluation; these cutoffs are separate from the main comparison's Effect threshold. Evaluation uses three new $N = 4800$ pools per system. Target selection is the fraction of the two true effect factors selected; non-target selection uses the other four factors, including the static factor. The no-effect control instead uses all six factors as non-targets. We average repeats within system, then report five-system means. For our model, the no-effect non-target rates are 1.1%, 1.1%, and 0% for unconditional, conditional-mean, and conditional-distribution scores, respectively. Under mean changes the same cutoffs select many non-targets (Figure 4c). A cutoff calibrated where no factor responds therefore does not control non-target selections when learned coordinates mix responding and nonresponding factors.

K.4 Topology sensitivity and objective ablation

Set recovery across instantaneous topologies. We repeat the comparison on three instantaneous topologies: chain, sparse-random, and a denser DAG. The sweep uses $d = 5$, seeds 15–19, and the

same evaluator, data scale, training budget, and tuning rules as the instantaneous process of Table 1. Table 10 gives the paired contrasts with the temporal backbone that Section 5.1 cites, and Table 11 the absolute results of six methods. In all three topologies our model has higher mean dynamic MCC, Read F_1, and Effect F_1 than the backbone, and its Read gains (.231–.337) exceed its Effect gains (.052–.089); the backbone keeps slightly lower probe MSE.

Table 10: **Paired contrasts with the temporal backbone across three instantaneous topologies.** Each row pairs the same five seeds 15–19; brackets are 95% paired percentile-bootstrap intervals. ΔMSE is backbone minus our model, in units of 10^{-3}; the other differences are our model minus backbone, so positive values favor our model.

Topology	ΔMSE ($\times 10^{-3}$)	ΔMCC	ΔRead	ΔEffect
Chain	$-.58$	$+.077$	$+.252$	$+.060$
	$[-1.88, .57]$	$[-.007, .164]$	$[.107, .400]$	$[.014, .106]$
Sparse-random	-1.48	$+.042$	$+.231$	$+.052$
	$[-2.23, -.74]$	$[-.079, .147]$	$[.040, .370]$	$[-.062, .133]$
Denser DAG	-1.28	$+.136$	$+.337$	$+.089$
	$[-3.00, .39]$	$[.074, .197]$	$[.265, .410]$	$[.029, .166]$

Table 11: **Absolute six-method results across the three additional instantaneous topologies.** Entries are mean ± sample SD over five seeds; MSE is comparable within each topology.

Method	Probe MSE ↓	Dynamic MCC ↑	Read F_1 ↑	Effect F_1 ↑
Chain				
Ours	$.644 \pm .093$	$.744 \pm .062$	$.893 \pm .153$	$.608 \pm .127$
TDRL	$.654 \pm .095$	$.613 \pm .057$	$.668 \pm .135$	$.533 \pm .075$
IDOL	$.649 \pm .096$	$.692 \pm .071$	$.538 \pm .158$	$.514 \pm .032$
SlowVAE	$.982 \pm .185$	$.151 \pm .030$	$.133 \pm .298$	$.500 \pm .000$
Temporal backbone	$.644 \pm .094$	$.668 \pm .062$	$.641 \pm .113$	$.548 \pm .073$
Static alignment	$.693 \pm .125$	$.567 \pm .034$	$.556 \pm .097$	$.629 \pm .211$
Sparse-random				
Ours	$.647 \pm .081$	$.732 \pm .081$	$.853 \pm .145$	$.581 \pm .144$
TDRL	$.655 \pm .092$	$.622 \pm .084$	$.567 \pm .091$	$.467 \pm .075$
IDOL	$.656 \pm .087$	$.669 \pm .051$	$.663 \pm .125$	$.514 \pm .032$
SlowVAE	$1.059 \pm .285$	$.130 \pm .036$	$.000 \pm .000$	$.514 \pm .032$
Temporal backbone	$.646 \pm .082$	$.690 \pm .077$	$.622 \pm .116$	$.529 \pm .039$
Static alignment	$.693 \pm .102$	$.563 \pm .035$	$.589 \pm .171$	$.531 \pm .184$
Denser DAG				
Ours	$.645 \pm .105$	$.796 \pm .060$	$.893 \pm .153$	$.617 \pm .102$
TDRL	$.649 \pm .108$	$.655 \pm .089$	$.595 \pm .071$	$.500 \pm .000$
IDOL	$.651 \pm .112$	$.673 \pm .069$	$.595 \pm .071$	$.548 \pm .073$
SlowVAE	$.978 \pm .227$	$.136 \pm .021$	$.000 \pm .000$	$.543 \pm .039$
Temporal backbone	$.644 \pm .106$	$.660 \pm .022$	$.556 \pm .097$	$.529 \pm .039$
Static alignment	$.688 \pm .138$	$.574 \pm .041$	$.570 \pm .165$	$.541 \pm .169$

Matched read-objective ablation. Table 12 tests whether the read objective accounts for the Read advantage over the backbone. Both arms are trained from scratch on the five lag-only seeds with paired data, initialization, and minibatch order, and differ only in the read-objective weight. None of the three intervals excludes zero; the lower endpoint of the Effect interval is exactly zero. The ablation therefore does not attribute the Read advantage over the backbone to this term.

Table 12: **Matched read-objective ablation.** Absolute rows are mean ± sample SD over five matched seeds of the lag-only process. For MSE, Δ is ablation minus our model, in units of 10^{-3}; for role scores, Δ is our model minus ablation. Brackets give 95% paired percentile-bootstrap intervals.

Model or contrast	Probe MSE	Read F_1	Effect F_1
Ours	$.635 \pm .089$	$.853 \pm .145$	$.636 \pm .100$
Without read objective	$.634 \pm .089$	$.853 \pm .145$	$.590 \pm .043$
Paired Δ (MSE $\times 10^{-3}$)	$-.529$	$.000$	$+.046$
95% CI	$[-1.171, .061]$	$[-.160, .160]$	$[.000, .099]$

K.5 AUDIT AND REPRESENTATION DIAGNOSTICS

Fixed-representation audit sample size. This diagnostic asks how Effect recovery and null selections change when a frozen representation is audited on more transitions. We freeze the final checkpoints of our model and the temporal backbone, training normalization, DGP and mixer parameters, one full-space assignment per representation, readout random features, and thresholds. For each of five training seeds and three new evaluation repeats, each N is a nested prefix of a pool whose rows come from independent trajectories. The matched null shares current state, action, innovation, assignment, and subset with its factual row, but removes the action term from the next-state target. Repeats are averaged within training seed before the displayed mean and interval are computed.

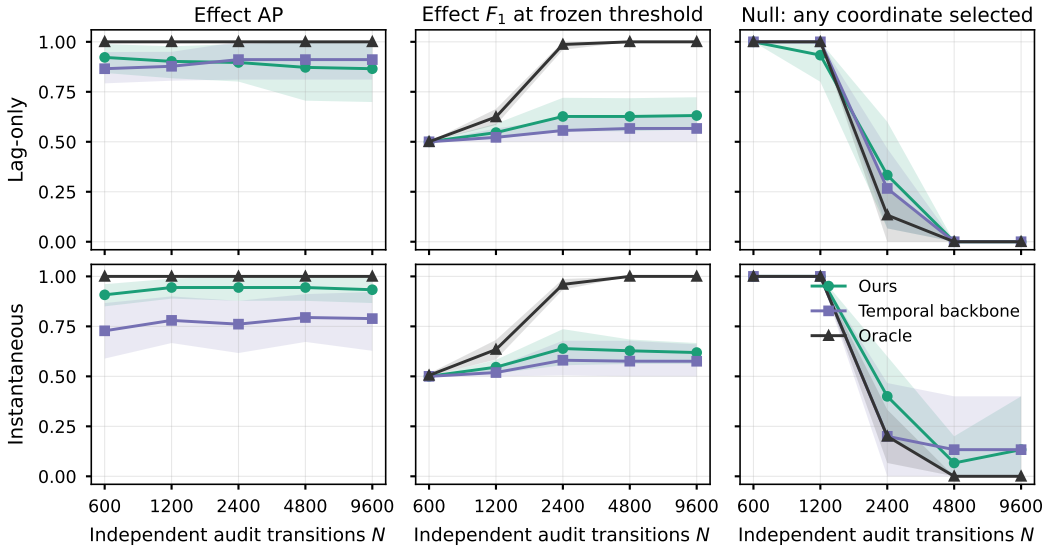

Figure 7: **Effect recovery and matched-null selections as the fixed-representation audit sample grows.** Points are means over five training seeds after averaging three evaluation repeats within seed; bands are seed-level 95% percentile-bootstrap intervals of the mean from 20,000 draws. Under the matched null, the curves give the fraction of evaluation repeats with at least one selected coordinate.

Oracle ranking is perfect throughout, and its frozen-threshold F_1 reaches one by $N = 4800$ (Figure 7). On learned coordinates, ranking stays high but frozen-threshold F_1 plateaus well below exact recovery. At $N = 9600$ our model's Effect AP is .866 for lag-only and .933 for instantaneous dynamics, yet the learned arms select 4.47–5.20 coordinates on average for a true set of two, and Effect and joint exact recovery are zero. Null selections largely vanish as N grows, with a small residual rate in the instantaneous learned arms. A larger audit sample therefore removes a low-N threshold effect but not the over-selection on learned coordinates.

Oracle and aligned learned coordinates. Table 13 applies the same readouts to the oracle coordinates and to our model's learned coordinates of the five lag-only systems, after a one-to-one Hungarian alignment used only for evaluation. On oracle coordinates both sets are recovered exactly. On the aligned learned coordinates, ranking quality stays high for both roles, but the sets selected at the frozen thresholds are less accurate, and the effect set is exact in none of the five seeds. The readouts are thus exact on the true factors; the remaining Effect error arises on the learned coordinates.

Table 13: **Readouts on oracle and on our model's aligned learned coordinates.** Lag-only process, five seeds. AP and fixed-threshold F_1 are mean $\pm$ sample SD; Exact counts the seeds whose set is recovered exactly. Alignment is a Hungarian assignment used only for evaluation.

Role	Coordinates	AP	Fixed F_1	Exact
Read	Oracle	$1.000 \pm .000$	$1.000 \pm .000$	5/5
Read	Learned, aligned	$.967 \pm .075$	$.853 \pm .145$	2/5
Effect	Oracle	$1.000 \pm .000$	$1.000 \pm .000$	5/5
Effect	Learned, aligned	$.917 \pm .118$	$.636 \pm .100$	0/5

L Physics-Simulation Details

L.1 Role changes and utility in a common fixed representation

Scene and representation. Figure 5a–b tests whether separate set estimates track changes in the policy and actuator, and whether the selected coordinates retain information for their respective prediction tasks. The fixed RGB representation has six coordinates, corresponding to the positions and velocities of objects A, B, and C. Its front end uses fixed-scene, color, and single-axis-motion priors. The five runs use distinct seeds; all set estimators and prediction tasks within a run share this representation.

Four mechanism settings. Table 14 lists the four settings; the read and effect sets change separately or together, and overlap under the actuator change. One coordinate matching, pooled across the four settings, is shared by both roles within each run. Each setting supplies its own calibration data for set estimation: the sets are re-estimated after a mechanism change.

Table 14: The four mechanism settings of the fixed RGB study. Each entry names the object, A, B, or C, that the policy reads or the actuator targets.

Setting	Policy reads	Actuator targets
Baseline	A	B
Policy change	C	B
Actuator change	A	A
Both changes	C	A

Set estimation. Our conditional-association estimator and conditional Group Lasso use the same inputs and calibration episodes. Their (Read, Effect) F1 scores are $(.967 \pm .046, .950 \pm .075)$ and $(.860 \pm .048, .683 \pm .207)$, respectively. These are means and sample SDs over five seeds, after averaging the four settings within each seed. Exact joint recovery in all four settings occurs in $2/5$ and $0/5$ seeds. Figure 5a shows our estimator's selection frequencies, including misses and false selections; the figure supplement contains both estimators' full selection counts and seed-level scores.

Learned representations from RGB input. Our model and four baselines (TDRL, IDOL, SlowVAE, and Temporal) were each trained for 100 epochs on two development seeds from the same structured RGB input. None of the ten fits reached a matched absolute Spearman correlation of at least $.8$ on all six physical state variables; per-fit correlations are in the figure supplement.

Prediction utility. Figure 5b compares the two estimated sets with all coordinates and size-matched random subsets, using the same prediction-head families. Each head uses 128, 512, 2048, or 4096 fit windows per setting; these counts exclude front-end fitting and calibration for set estimation. Read NMSE is action-prediction MSE divided by the fit action variance. Response NMSE is the error in the 0.05 s response to ± 0.5 action changes, divided by the zero-response error in the same measurement space; hence one is the zero-response reference. Random subsets are averaged within each seed and setting. Curves then average the four settings within each seed and report the

mean and sample SD across the five seeds. The selected coordinates retain information for both tasks: their errors stay below the size-matched random controls across the displayed sample sizes.

M BridgeData V2: paired-frame decoding

We examine whether visual coordinates help decode observed robot-state changes in BridgeData V2. The retrospective study uses the contiguous shards 00000–00063: 3,324 episodes and 13,293 windows, with collection-based train/dev/test splits of 2,416/501/407 episodes and 9,661/2,004/1,628 windows. Each probe receives the current seven-dimensional proprioceptive state and visual representations at both t and $t+1$ to estimate the observed state difference. The next frame is an input, so the task measures paired-frame decoding.

All methods expose 32 dynamic coordinates from frozen CLIP image features (adapters in Appendix J.4). We select effect-associated coordinates by measuring how much each action component adds to next-coordinate prediction, conditional on the current representation and the other action components. Scores use episode-grouped ridge fits on dev data. For each latent coordinate, the maximum across action components is compared with within-episode action-shift nulls from train, followed by Benjamini–Hochberg selection at $q = .05$. Decoding probes are fitted on train, their ridge penalties are selected by dev squared error, and they are refitted on train∪dev before test evaluation. Each selected set is used at both frames and compared with 50 equal-size random subsets of the same representation; each subset receives the same probe-fitting procedure.

Table 15: **Visual and selection gains on BridgeData V2.** Values are mean ± sample SD across five training seeds on the common episode split, in units of 10^{-3} R^2, the scale of Figure 5c. Visual gain is the increase in test R^2 over the state-only probe when all 32 coordinates at both frames are added. Selection gain is the selected-set R^2 minus the mean R^2 of its 50 equal-size random sets; set sizes can differ across methods. Both use total squared deviation from the train∪dev target mean as the test denominator. The first seven rows correspond to Figure 5c, where Ours denotes the tuned configuration and method names omit "-style." CPC, IDOL, iVAE, TDRL, and β-VAE are local adaptations; tuning and capacities differ. Ours w/o gate uses the original configuration.

Method	Visual gain	Selection gain
Ours (tuned)	$8.23 \pm .71$	1.85 ± 2.16
Temporal	7.37 ± 1.38	$.46 \pm .66$
TDRL-style	8.10 ± 1.63	$1.10 \pm .99$
iVAE-style	8.06 ± 1.58	$.70 \pm 1.42$
IDOL-style	7.30 ± 1.72	$.72 \pm 1.26$
CPC-style	4.12 ± 2.32	$.03 \pm .57$
Random encoder	2.62 ± 2.15	$-.14 \pm 1.39$
Ours (original)	7.75 ± 1.63	$-.41 \pm 1.20$
Ours w/o gate	6.96 ± 1.37	$.11 \pm .77$
Static alignment	6.25 ± 1.15	1.66 ± 1.23
β-VAE-style	6.13 ± 2.72	$.05 \pm 1.03$

All methods have positive mean visual gains, whereas selection gain is small and has a negative mean for the random encoder and for our original configuration (Table 15). For the tuned model, visual features raise test R^2 from .026 to .035, selection gain is positive in all five seeds, and the selected set still averages 24.8 of 32 coordinates. Both gains are small; they show that the frozen visual coordinates, and the selected subsets of them, carry information about the observed state change.

N Scope of the results, the estimators, and the studies

The identification results are population statements under the model of Section 2, Assumptions 1–3, and the minimum-edge criterion: frames and actions are synchronized, without unknown delays or timing jitter; the transition is first-order and stationary, with the current state and the logged action as its only inputs; and effect labels are one-step, reduced-form responses to the action, not mechanisms.

The estimators of Section 4 score a frozen encoder's variables with protocol-dependent thresholds. They take the assumptions and the criterion as premises rather than verifying them, read-faithfulness among them. Their finite-sample rates, and whether their threshold decisions are invariant under the allowed recodings, remain open.

The controlled study of Section 5.1 is synthetic audit evidence with full-space matching to the known factors, on processes whose fixed innovation covariances fail Assumption 2. The RGB study of Section 5.2 covers four mechanism settings of one scene over five seeds. The BridgeData study covers an already inspected 64-shard subset of one corpus without membership labels and evaluates paired-frame decoding of observed state changes (Appendix M). None provides closed-loop control, planning, or policy-return evidence.