TRACE-WM: LEARNING PHYSICAL CONCURRENCY FOR ASYNCHRONOUS WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

World models trained from asynchronous interaction logs can overfit arbitrary action serializations: physically interchangeable actions appear as different futures, while global order invariance can erase interactions for which order genuinely matters. We study whether a world model can instead infer where temporal order is physically irrelevant. We formulate state-dependent action swappability through a stochastic two-step diamond discrepancy and introduce *Active Diamond Discovery* (ADD), an active probing framework for uncertain likely-safe diamonds, together with TRACE-WM, which imposes invariance only on relation-admitted reorderings. In a fresh pre-specified 16-seed comparison, a local-uncertainty ADD acquisition discovers 96.19 safe cross-robot probes versus 54.19 for Random, a 77.5% gain with 16/16 paired wins, and exceeds the original variance-reduction instantiation by 28.0%. In a separate 16-seed matched held-out-state evaluation, TRACE-WM reduces balanced distributional prediction error by 34.0% relative to Ordered and dependent order-effect error by 57.2% relative to Global invariance. A size-matched, common-initialization control shows that ADD-selected relation elements are modestly more useful downstream than Random-selected elements, while extra imperfect coverage can offset this benefit. The discovery advantage also persists under a discontinuous hard-mutex interaction. Relation extrapolation remains imperfect, so the learned relation is treated as an empirical inference rather than a universal safety certificate.

1 INTRODUCTION

World models compress interaction experience into predictive dynamics that can support imagination, planning, and control (Ha and Schmidhuber, 2018; Hafner et al., 2025). In asynchronous physical systems, however, the data contain an additional source of complexity that is easy to overlook: the same physical outcome may be produced by many different serializations of actions. Two robots can often execute independent motions in either order, two processes may update disjoint resources, and distributed agents may produce interleavings whose differences are artifacts of logging rather than differences in the underlying physics. A conventional sequence model is nevertheless trained to distinguish these serializations unless it is explicitly given a structural reason not to.

One response is to impose order invariance globally. This can improve generalization across unseen interleavings, but it is physically wrong whenever actions interact through contact, congestion, shared resources, or state-dependent safety constraints. The resulting tension is fundamental: *an ordered model distinguishes too much, while a globally invariant model can distinguish too little*. We ask whether a world model can instead learn the boundary between these regimes from interaction. Asynchronous physical systems combine distributed event streams, observation delays, and shared-resource interactions, making serialization a physical as well as statistical issue.

We formulate this boundary through a local stochastic *diamond discrepancy*. For state s and actions a, b, compare the two-step endpoint distributions obtained from ab and ba,

$$\Delta(s; a, b) = D(P_b P_a(\cdot \mid s),\, P_a P_b(\cdot \mid s))\,, \tag{1}$$

where D is a task-relevant distributional discrepancy. A pair is locally swappable when $\Delta(s; a, b) \leq \epsilon$. Crucially, this relation is state dependent: actions may commute far from an interaction zone and become strongly order dependent after the same agents approach a shared resource.

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.

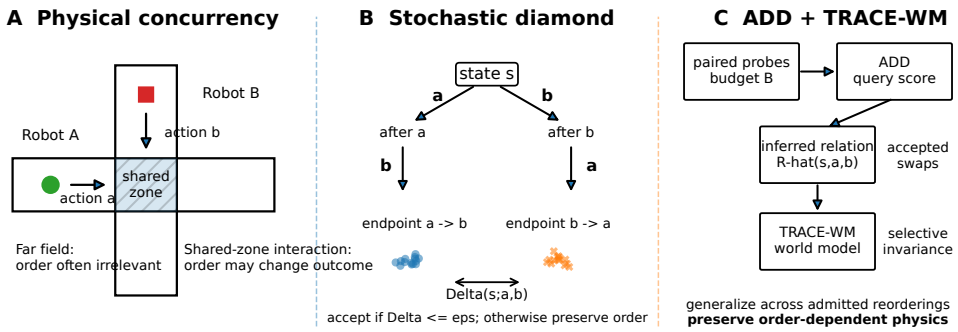


Figure 1: **Engineering view of TRACE-WM.** (A) Two asynchronous robots can be order-insensitive in the far field but order-sensitive near a shared interaction zone. (B) We test this locally with a stochastic two-step diamond: execute ab and ba from the same state and compare their endpoint distributions through $\Delta(s; a, b)$. (C) ADD allocates a limited paired-intervention budget to infer a selective swappability relation, and TRACE-WM imposes invariance only on relation-admitted reorderings.

Figure 1 summarizes the physical setting and learning pipeline. Our approach has two components. **Active Diamond Discovery (ADD)** treats physical interventions as a limited resource and actively selects which state–action diamonds to probe. Rather than reconstructing the entire transition model, ADD targets uncertainty around likely-safe but unresolved diamonds under a fixed intervention budget. **TRACE-WM** then uses the inferred relation as a selective invariance: only action reorderings admitted by that relation are tied together during world-model training. The resulting model generalizes across physically justified interleavings without forcing genuinely order-sensitive interactions to collapse.

This differs from partial-order and trace-theoretic methods that exploit a supplied independence relation (Fan et al., 2018), from active automata learning where rewrite structure is supplied as advice (Fica and Otop, 2025), and from Commutative-MDP work that assumes or exploits supplied action commutativity (Clark, 2025). A public implementation associated with a concurrent non-archival TMLR submission likewise gates value updates using an environment-provided commutativity oracle (Clark, 2026). It also complements compositional multi-agent world models such as COMBO (Zhang et al., 2025): our object of interest is not merely factorization over agents, but the learned physical equivalence of temporal interleavings.

Our experiments deliberately include strong negative controls. Global invariance can achieve lower raw prediction error on some metrics, and a full-dynamics learner can obtain broader relation coverage. However, these gains can be purchased by contaminating the quotient with false swaps and erasing real order effects. The key empirical question is therefore not simply how many traces can be merged, but how much physically faithful structure can be removed from the learning and planning problem.

Contributions. We make four contributions:

1. We introduce a stochastic, state-dependent physical concurrency formulation based on local two-step diamond discrepancies and use it to define error-controlled local trace rewrites rather than assuming a global equivalence relation.
2. We propose **ADD**, an active intervention framework that prioritizes likely-safe uncertainty near the swappability boundary while controlling pair-level allocation. A fresh pre-specified 16-seed confirmation of the simpler local-uncertainty acquisition discovers 96.19 genuinely safe cross-robot probes versus 54.19 for Random, a 77.5% gain with 16/16 paired wins (exact one-sided $p = 1.53 \times 10^{-5}$). The original exact variance-reduction instantiation is weaker on the same fresh seeds, so we do not attribute ADD's discovery gain to exact global variance reduction.
3. We propose **TRACE-WM**, a probabilistic world-model training method that enforces invariance only on relation-admitted reorderings. In a 16-seed matched evaluation, TRACE-WM with the frozen ADD relation reduces balanced distributional error by 34.0% relative to Ordered and

dependent order-effect error by 57.2% relative to Global invariance, with 16/16 desired-direction outcomes in both comparisons. A size-matched common-initialization control further shows that ADD-selected relation elements are modestly more useful downstream than Random-selected elements at comparable relation size.

4. We introduce **Graph-TRACE**, a permutation-invariant residual extension for agent-count shift. With only 40 multi-agent calibration probes, it increases relation recall by $5.43\times$ on three robots and $9.65\times$ on four robots relative to pair-only zero-shot transfer, while maintaining a substantially safer quotient than scratch contextual learning or global invariance.

2 Related Work

World models and compositional dynamics. World models learn predictive dynamics for planning, control, and representation learning (Ha and Schmidhuber, 2018; Hafner et al., 2019; 2020; 2021; 2025; Schrittwieser et al., 2020; Micheli et al., 2023; Hansen et al., 2024). Recent work has also emphasized compositional or decision-relevant structure. COMBO learns a compositional world model for embodied multi-agent cooperation (Zhang et al., 2025), SeqWM decomposes multi-robot dynamics into sequential agent-wise models (Zhao et al., 2026), WorldGym evaluates policies through world-model rollouts (Quevedo et al., 2026), ExoPredicator models concurrent exogenous processes for planning (Liang et al., 2026), and Sparse Imagination studies the computation–control tradeoff of model-based planning (Chun et al., 2026). These methods do not decide whether two temporal serializations are physically equivalent at a particular state.

Partial orders, traces, and approximate commutativity. Classical trace theory and partial-order reduction avoid redundant interleavings by exploiting an independence relation supplied by the system model (Diekert and Rozenberg, 1995; Valmari, 1991; Peled, 1993; Godefroid, 1996; Flanagan and Godefroid, 2005; Abdulla et al., 2014). Approximate partial-order reduction extends this idea to metric transition systems through approximate independence and commutativity (Fan et al., 2018). TRACE-WM differs in where the relation comes from: the state-dependent relation is inferred from stochastic physical interventions rather than assumed from a specification.

MDP homomorphisms, abstraction, and bisimulation. MDP minimization, homomorphisms, and bisimulation quotient states or state–action pairs under exact or approximate behavioral equivalence (Givan et al., 2003; Ravindran and Barto, 2002; 2004; Li et al., 2006; Ferns et al., 2004; Zhang et al., 2021). This literature establishes principled abstraction and value-preservation criteria once the relevant equivalence is given or learned over the MDP representation. TRACE-WM quotients a different object: a state-dependent relation over temporal interleavings, inferred from paired stochastic interventions and used as a selective world-model invariance rather than as direct state aggregation.

Active discovery and equivalence testing. Active automata learning with advice uses trusted rewrite systems to reduce query complexity (Fica and Otop, 2025). Clark's Commutative-MDP thesis formalizes order-invariant MDP structure (Clark, 2025); a public implementation associated with concurrent non-archival TMLR work extends this line to state-conditioned action commutativity but gates value updates with an environment-level `is_commutative` oracle (Clark, 2026). More broadly, active learning and Gaussian-process experimental design select observations to reduce predictive uncertainty or information loss (Cohn et al., 1996; Krause et al., 2008; Srinivas et al., 2010). ADD instead targets uncertainty reduction specifically around likely-safe unresolved diamonds. Statistical certification is also distinct from ordinary non-equivalence detection: tolerant distribution and equivalence testing can require substantially more samples than their non-tolerant counterparts (Canonne et al., 2022).

Learned invariance and multi-agent context. Data-dependent invariance can be learned rather than hard coded (Benton et al., 2020), while graph networks provide a general relational inductive bias for multi-entity systems (Battaglia et al., 2018). Graph-TRACE uses a permutation-invariant set function (Zaheer et al., 2017), but we do not claim the aggregation architecture as novel. Its role is narrower: it models the residual change in pairwise swappability caused by surrounding agents while retaining the frozen two-agent physical-concurrency model as a prior mean.

3 PROBLEM FORMULATION

Consider a controlled stochastic system with state space $\mathcal{S}$, action set $\mathcal{A}$, and action-conditioned Markov kernels $P_a(\cdot \mid s)$. For a sequence $\tau = (a_1, \ldots, a_H)$, let $P_\tau = P_{a_H} \circ \cdots \circ P_{a_1}$. We are interested in asynchronous systems where multiple action sequences may differ only by the ordering of physically independent local events.

3.1 STOCHASTIC DIAMOND DISCREPANCY

For adjacent actions a, b, define the two serialized endpoint laws

$$Q_{ab}(\cdot \mid s) = P_b P_a(\cdot \mid s), \qquad Q_{ba}(\cdot \mid s) = P_a P_b(\cdot \mid s), \tag{2}$$

and the local diamond discrepancy

$$\Delta(s; a, b) = D\big(Q_{ab}(\cdot \mid s), Q_{ba}(\cdot \mid s)\big). \tag{3}$$

For tolerance $\epsilon > 0$, the state-dependent safe relation is

$$\mathcal{R}_\epsilon = \{(s, a, b) : \Delta(s; a, b) \le \epsilon\}. \tag{4}$$

In our stochastic benchmark, we estimate Δ with a common-random-number coupling: both serializations reuse the same exogenous disturbance realization and we measure the normalized endpoint distance. The expected coupled distance is a valid upper bound on Wasserstein-1 under that coupling, while remaining directly interpretable in the physical state coordinates.

We deliberately call Eq. (4) a *local trace relation*, not a global equivalence relation. State dependence means that a rewrite valid at s need not remain valid after an earlier rewrite changes the reached state. For a trace τ and reached state s_t, we therefore permit the adjacent rewrite

$$(\ldots, a, b, \ldots) \rightsquigarrow (\ldots, b, a, \ldots) \quad \text{only if} \quad (s_t, a, b) \in \mathcal{R}_\epsilon. \tag{5}$$

3.2 LEARNING OBJECTIVE

The learner receives a budget B of paired physical interventions. Each query chooses a candidate $x = (s, a, b)$ and returns a stochastic discrepancy observation y. The goal is not global system identification; it is to recover a high-purity subset $\widehat{\mathcal{R}} \subseteq \mathcal{R}_\epsilon$ that is large enough to support useful trace sharing under a limited intervention budget.

The learned relation is then used to train a probabilistic world model p_θ from asynchronous logs. Let $\mathcal{D}$ denote the observed ordered transitions and $\mathcal{A}(\mathcal{D}, \widehat{\mathcal{R}})$ the selectively augmented data formed by adding reverse serializations only when the learned relation declares the diamond safe. The TRACE-WM objective is

$$\theta^* = \arg\min_\theta \; \mathcal{L}_{\mathrm{NLL}}\Big(\mathcal{D} \cup \mathcal{A}(\mathcal{D}, \widehat{\mathcal{R}}); \theta\Big). \tag{6}$$

The ordered baseline uses no augmentation; global invariance augments every pair; and the oracle baseline augments exactly the ground-truth safe relation. This isolates whether learned selective invariance occupies a useful middle ground between memorizing serialization and erasing order.

4 METHOD

4.1 ACTIVE DIAMOND DISCOVERY

For each unordered action pair, ADD models the discrepancy surface with a Gaussian process (GP). At round t, the posterior at candidate u has mean $\mu_t(u)$, standard deviation $\sigma_t(u)$, and conservative upper score

$$U_t(u) = \mu_t(u) + \beta \sigma_t(u). \tag{7}$$

A candidate is admitted to the learned relation only when $U_t(u) \le \epsilon$. The multiplier $\beta = 2.5$ was selected only on development seeds 1600–1603 from $\{1.5, 2.0, 2.5\}$ and was frozen before all confirmatory two-robot evaluations.

Algorithm 1 Active Diamond Discovery (ADD)

1: Initialize each action pair with a small stratified probe set.
2: **while** total probes $< B$ **do**
3: Fit pair-conditioned GP posteriors (μ_t, σ_t).
4: Form unresolved likely-safe sets $\mathcal{U}_t^{ab}$.
5: **for** each available query x **do**
6: Score local boundary uncertainty using Eq. (9).
7: **end for**
8: Probe the highest-scoring candidates subject to pair caps.
9: **end while**
10: Return $\widehat{\mathcal{R}} = \{u : \mu(u) + \beta\sigma(u) \le \epsilon\}$.

ADD focuses queries on candidates that are plausibly safe but unresolved,

$$\mathcal{U}_t^{ab} = \left\{u : \; p_t^{\mathrm{safe}}(u) > 0.35, \; U_t(u) > \epsilon\right\}, \tag{8}$$

where $p_t^{\mathrm{safe}}(u) = \Phi((\epsilon - \mu_t(u))/\sigma_t(u))$. The empirically supported local acquisition is

$$\alpha_t^{\mathrm{local}}(x) = \frac{p_t^{\mathrm{safe}}(x)\,\sigma_t(x)}{\sqrt{n_{ab}}} \exp\!\left(-\frac{(U_t(x) - \epsilon)_+}{\tau}\right), \tag{9}$$

where n_{ab} is the number of probes already assigned to pair (a, b). The score combines local uncertainty, safe probability, proximity to the swappability boundary, and diminishing returns across action pairs. A stratified coverage floor and per-pair cap prevent globally easy pairs from absorbing the full intervention budget.

The original confirmatory discovery experiments used a more expensive exact one-step GP variance-reduction score over unresolved candidates. A fresh pre-specified 16-seed comparison shows that Eq. (9) discovers 96.19 safe cross-robot probes versus 75.13 for the exact variance-reduction instantiation and 54.19 for Random. We therefore treat exact global variance reduction as a historical ADD instantiation rather than as the mechanism responsible for the discovery gain. The exact score and its ablation are reported in Appendix G; downstream TRACE-WM results retain the frozen relation used in their original matched protocol rather than being recomputed after observing this ablation.

4.2 TRACE-WM: SELECTIVE INVARIANCE FOR PROBABILISTIC WORLD MODELS

TRACE-WM is agnostic to a particular world-model architecture; the required interface is a predictive distribution conditioned on state and action sequence. In our controlled benchmark we use a Gaussian endpoint model that predicts mean and diagonal variance for a two-step action sequence. The model is trained by Gaussian negative log likelihood.

For a logged sample (s, a, b, s'), if $(s, a, b) \in \widehat{\mathcal{R}}$, TRACE-WM adds the reverse serialization (s, b, a, s') to the training set. No reverse sample is added for unresolved or predicted-dependent diamonds. Thus the learned relation acts as a data-dependent selective invariance rather than a universal symmetry constraint. This implementation is equivalent to tying the two serializations through shared target evidence while preserving unconstrained capacity on dependent interactions.

For evaluation, let the predicted endpoint Gaussians under ab and ba be $\mathcal{N}(\hat{m}_{ab}, \hat{\Sigma}_{ab})$ and $\mathcal{N}(\hat{m}_{ba}, \hat{\Sigma}_{ba})$. We use the diagonal-Gaussian 2-Wasserstein distance for distributional prediction error and separately measure the error in the predicted *order effect*, i.e., the discrepancy between the two predicted endpoint laws. The latter is critical: a globally invariant model can score well on average prediction while incorrectly driving all order effects toward zero.

4.3 GRAPH-TRACE FOR AGENT-COUNT SHIFT

Pairwise swappability can become too conservative when additional agents change local congestion. Graph-TRACE retains the frozen two-agent ADD model as a prior mean and learns only a contextual residual from a small multi-agent calibration set,

$$\widehat{\Delta}_N(i, j) = \widehat{\Delta}_2(i, j) + g_\phi\big(q_i, q_j, \mathrm{Agg}_{k \neq i,j}\, \psi(q_k)\big). \tag{10}$$

The aggregation is permutation invariant and summarizes surrounding-agent occupancy/centrality at the physical interaction radii. We emphasize that permutation-invariant aggregation itself is standard (Zaheer et al., 2017); the contribution is its use as a residual correction to a learned pairwise physical-concurrency prior. In our experiments, Graph-TRACE uses only 40 stochastic multi-agent calibration probes and is evaluated on disjoint held-out states.

4.4 Local rewrites, confidence, and composition

The learned rule $U_t(s, a, b) \leq \epsilon$ is a conservative *model score*, not a universal deployment-safety certificate. Our physical claims are therefore empirical: we report finite-benchmark precision on held-out states and a separate independent-confirmation study under an explicit Gaussian observation model (Appendix C). This distinction prevents GP posterior uncertainty from being misreported as a distribution-free guarantee.

A local rewrite also need not remain harmless after an arbitrary suffix. If each suffix transition kernel is W_1-Lipschitz, $W_1(P_c\mu, P_c\nu) \leq L_c W_1(\mu, \nu)$, then a local swap with discrepancy at most ϵ_j followed by actions $c_{j+1:T}$ contributes at most

$$\epsilon_j \prod_{k=j+1}^{T} L_{c_k} \tag{11}$$

to terminal discrepancy. By the triangle inequality, multiple accepted local rewrites accumulate additively after the corresponding suffix amplification factors. We use this only as an interpretive bound; our benchmark does not assume known global contraction constants, and the quotient checks swaps at reached states. A derivation and empirical composition audit appear in Appendix H.

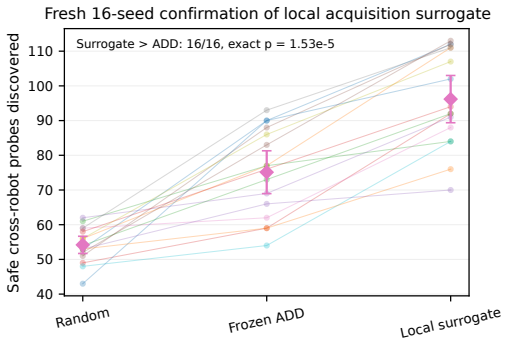
5 Experiments

Benchmark. Two asynchronous robots move on perpendicular lanes through a shared interaction zone. Stochasticity includes actuation variation, position jitter, congestion-dependent communication/safety delay, and event-keyed common random numbers across the two serializations. The benchmark contains 200 initial states and six unordered action pairs (1,200 diamonds). At $\epsilon = 0.04$, two same-robot pairs are globally safe while the four cross-robot pairs are approximately 46.5–47.0% safe, creating a genuinely state-dependent concurrency boundary.

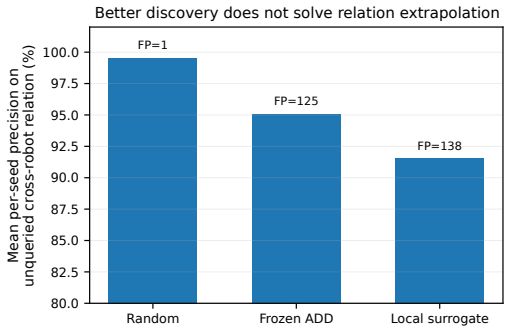
Baselines and protocol. We compare Random probing, ADD, an active full-dynamics learner from which swappability is derived post hoc, Ordered world-model training, Global invariance, Random-derived selective invariance, FullDyn-derived selective invariance, and Oracle selective invariance. Multi-agent transfer additionally compares pair-only zero-shot transfer, ScratchGraph40, Graph-TRACE-40, and Global invariance. Main comparisons use paired seeds; plots report seed-level means with 95% confidence intervals unless otherwise noted. The matched TRACE-WM experiments use the frozen original ADD relation for protocol continuity, while the fresh local-uncertainty ADD acquisition is evaluated separately as a discovery result.

6 Results and Analysis

Active discovery. Figure 2 summarizes the fresh discovery comparison. On fresh pre-specified seeds 8000–8015, the local-uncertainty ADD acquisition finds 96.19 genuinely safe cross-robot probes on average versus 75.13 for the original exact variance-reduction instantiation and 54.19 for Random. The local score beats both baselines in 16/16 seeds (exact one-sided $p = 1.53 \times 10^{-5}$ in both comparisons). This strengthens the active-discovery result while narrowing the mechanism claim: exact global variance reduction is not required. The tradeoff is relation extrapolation, where mean per-seed precision on unqueried cross-robot elements is 91.5% for the local score, 95.1% for the original ADD instantiation, and 99.5% for Random (Appendix G). On a separate discontinuous hard-mutex benchmark with no retuning of ϵ, the frozen original ADD instantiation also retains a 33.2% discovery advantage over Random (70.19 versus 52.69 safe probes; 15 wins, one tie; $p = 3.26 \times 10^{-4}$).

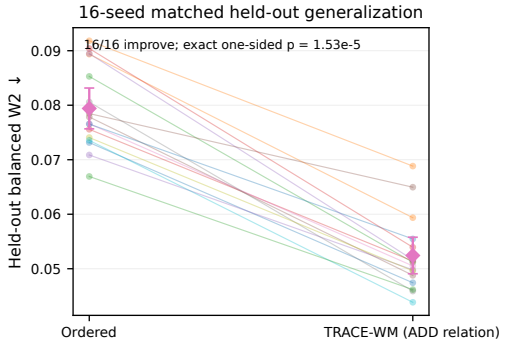


(a) Safe-probe discovery on fresh seeds.

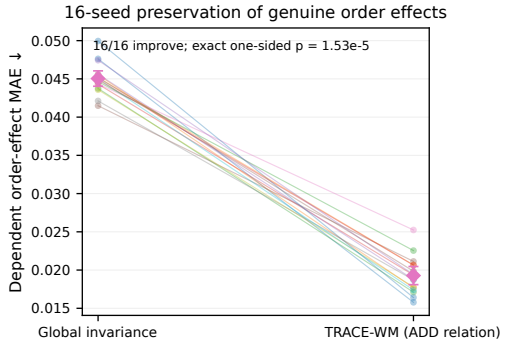


(b) Precision on unqueried cross-robot relation elements.

Figure 2: **Discovery versus relation extrapolation.** On fresh pre-specified seeds, the local-uncertainty ADD acquisition discovers 96.19 safe cross-robot probes versus 75.13 for the original exact variance-reduction instantiation and 54.19 for Random. The stronger discovery score does not imply safer extrapolation: mean per-seed relation precision is 91.5%, 95.1%, and 99.5%, respectively.



(a) Held-out distribution prediction error.



(b) Error in genuine order-dependent effects.

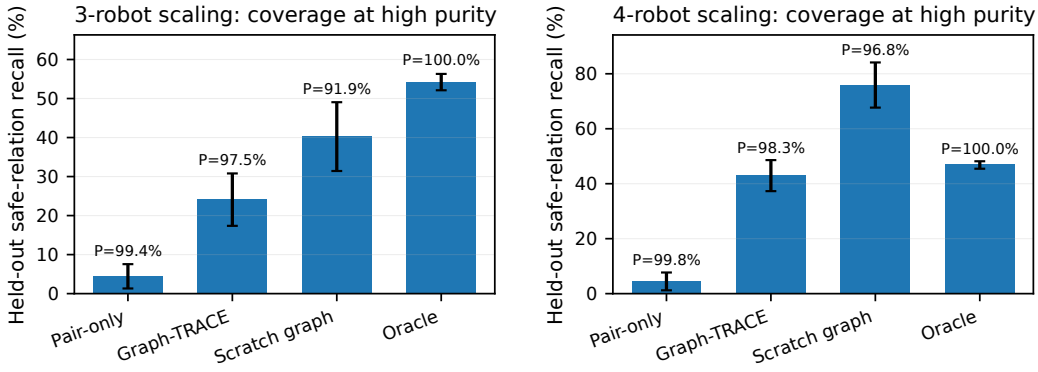
Figure 3: **The central TRACE-WM result on 16 matched seeds.** Lower is better in both panels. TRACE-WM with the ADD relation reduces held-out balanced distribution error by 34.0% relative to Ordered and dependent order-effect error by 57.2% relative to Global invariance. Every paired seed moves in the desired direction in both comparisons.

Matched held-out generalization. When held-out states are excluded from both discovery and world-model training, the 16-seed rerun reduces balanced distributional W2 from 0.07941 (Ordered) to 0.05243 for TRACE-WM with the frozen ADD relation, a 33.97% reduction with 16/16 seed-level improvements (exact one-sided Wilcoxon $p = 1.53 \times 10^{-5}$). The original unmatched comparison gives 0.05285 for Random-derived TRACE-WM and does not establish ADD superiority ($p = 0.264$). To isolate relation composition, we then hold world-model initialization fixed and subsample ADD to the Random relation size. Size-matched ADD obtains W2 0.052013 versus 0.052595 for Random (13/16 seeds, exact $p = 0.0327$) and dependent order-effect MAE 0.018385 versus 0.018918 (12/16, $p = 0.0222$). Full ADD obtains W2 0.052318 and does not improve over the size-matched subset. Thus ADD-selected elements are modestly more useful at comparable relation size, while extra imperfect coverage can offset that benefit. Figure 3 and Table 1 report the original matched comparison; Appendix G gives the composition control and relation-size sweep.

Why not global invariance? Global invariance attains lower raw balanced W2 than ADD in the matched experiment, but its dependent order-effect MAE is 0.04504 versus 0.01928 for TRACE-WM with ADD. Selective TRACE-WM therefore reduces dependent order-effect error by 57.2% relative to Global invariance, again with 16/16 paired improvements (exact one-sided $p = 1.53 \times 10^{-5}$).

Table 1: **Matched held-out-state comparison** across 16 seeds. Lower is better. Relation size and precision are means over held-out relations. For ADD, mean per-seed precision is 97.3%; pooled precision is 95.9%, with two low-precision seeds discussed in Appendix G.

Method	Balanced W2 ↓	Order MAE ↓	$\lvert\widehat{\mathcal{R}}\rvert$	Precision (%)
Ordered	0.079410	0.02038	0	–
Random-rel	0.052852	**0.01844**	37.9	**99.4**
TRACE-WM (ADD)	0.052431	0.01928	64.7	97.3
FullDyn-rel	0.043900	0.02591	116.9	82.1
Global	0.045807	0.04504	240.0	62.0
Oracle	**0.041317**	0.02105	148.7	100.0



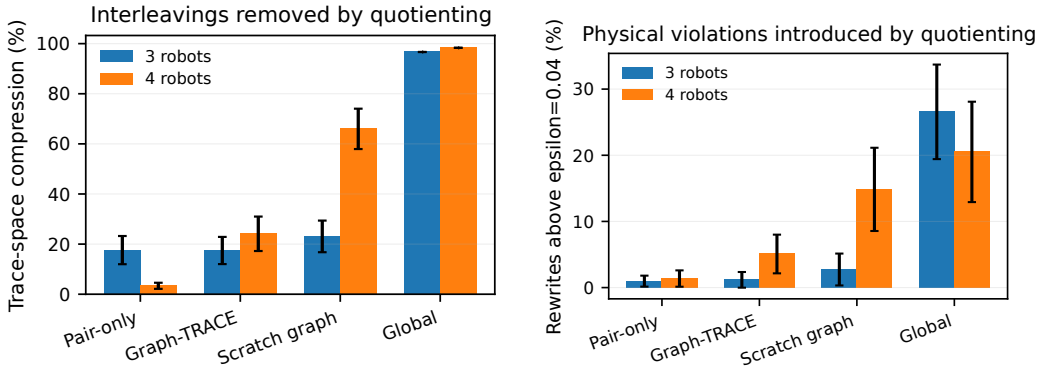
(a) Three robots. Bar height is recall; text above each bar is precision.

(b) Four robots.

Figure 4: **Graph-TRACE under agent-count shift.** Pair-only transfer is extremely pure but too conservative. With only 40 multi-agent calibration probes, Graph-TRACE increases held-out safe-relation recall by 5.43× at three robots and 9.65× at four robots while maintaining mean precision of 97.5% and 98.3%, respectively. Scratch graph learning obtains more coverage but with lower purity. Oracle local is restricted to locally certifiable rewrites, while recall for every method is measured against the full ground-truth safe set.

This is the central tradeoff of the paper: aggressive quotienting can improve average fit by merging physically distinct futures.

Agent-count shift. Graph-TRACE increases relation recall by $5.43\times$ on three robots and $9.65\times$ on four robots relative to pair-only zero-shot transfer with the same frozen two-agent prior and only 40 multi-agent probes (Fig. 4). On four-robot trace quotienting, compression rises from 3.36% to 24.11% while mean physical rewrite error remains 0.00774 (Fig. 5). Three-robot compression does not improve materially, which we report as a limitation rather than a solved scaling result.



(a) Compression: fraction of legal interleavings removed.

(b) Risk: fraction of rewrites whose physical discrepancy exceeds $\epsilon = 0.04$.

Figure 5: **Engineering compression–risk tradeoff.** Global invariance removes almost the entire trace space, but 20–27% of audited rewrites exceed the physical tolerance. Selective quotients compress less aggressively while keeping the violation rate small. Graph-TRACE is particularly useful at four robots, where it raises compression from 3.36% (pair-only) to 24.11% while remaining far safer than ScratchGraph or Global.

7 Limitations and Scope

Our experiments use controlled stochastic shared-intersection benchmarks rather than a high-fidelity simulator or physical robot system. The world model is a two-step Gaussian endpoint predictor on a two-dimensional physical state, not a latent visual sequence model; scalability to long-horizon, high-dimensional world models remains untested. The hard-mutex variant shows that ADD's discovery advantage is not restricted to a smooth discrepancy surface, but the learned relation remains an empirical inference rather than a deployment-safety certificate. In the expanded 16-seed matched evaluation, two original ADD runs have held-out relation precision of 76.1% and 80.4%, concentrated in one under-probed cross-robot pair. Fresh support-gating studies reduce catastrophic extrapolation but still fail pre-specified precision and coverage criteria, and the stronger local-uncertainty acquisition improves discovery while lowering relation precision (Appendix G). State-dependent local rewrites are also not automatically transitive: composing many swaps can accumulate error and can move a trace into regions where a previously learned local relation no longer applies. Finally, Graph-TRACE improves coverage under agent-count shift but does not dominate scratch contextual learning in recall. We therefore separate intervention-efficient discovery, imperfect relation inference, and downstream selective invariance rather than treating them as one solved problem.

8 Conclusion

We introduced TRACE-WM, a world-modeling approach that infers where temporal order is physically irrelevant instead of assuming either full serialization sensitivity or global order invariance. ADD actively allocates stochastic paired interventions near uncertain swappability boundaries, and a fresh 16-seed confirmation shows that a simpler local-uncertainty score improves safe discovery beyond the original exact variance-reduction instantiation. A size-matched common-initialization control further shows that ADD-selected relation elements are modestly more useful downstream than Random-selected elements, while a hard-mutex benchmark demonstrates that the active-discovery advantage survives a discontinuous interaction. TRACE-WM converts the inferred relation into selective training invariance and, across 16 matched held-out seeds, preserves genuine order-dependent effects substantially better than Global invariance. Reliable out-of-sample certification of relation elements and scaling to high-dimensional long-horizon world models remain open problems.

Reproducibility Statement

The anonymous supplementary package contains the benchmark dynamics, frozen outputs, random seeds, discovery budgets, GP hyperparameters, relation thresholds, ADD and TRACE-WM implementations, Graph-TRACE calibration code, and analysis artifacts included with this submission. The appendix reports complete hyperparameters, negative development results that motivated ADD, seed-level tables, statistical-test definitions, and the exact construction of stochastic common-random-number paired probes.

AI Use Statement

Generative AI tools were used to help develop and refine hypotheses and the conceptual framing, provide feedback on research methodology and experimental design, assist with mathematical formulation and derivations, implement and debug code, assist with scripts used to generate the synthetic benchmark data, support quantitative analysis and interpretation of results, identify and summarize relevant literature, and draft/edit manuscript text, figures, and formatting. All reported experiments were executed against saved source code and frozen outputs. Numerical claims and statistical tests were recomputed from archived result files, and citations were checked against source records. The authors reviewed the AI-assisted code, derivations, analyses, figures, and text, and take responsibility for the final content and scientific claims.

References

Parosh Aziz Abdulla, Stavros Aronis, Bengt Jonsson, and Konstantinos Sagonas. Optimal dynamic partial order reduction. In *Proceedings of the 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, pages 373–384. ACM, 2014.

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Gregory Benton, Marc Finzi, Pavel Izmailov, and Andrew Gordon Wilson. Learning invariances in neural networks from training data. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Clement L. Canonne, Ayush Jain, Gautam Kamath, and Jerry Li. The price of tolerance in distribution testing. In *Proceedings of the 35th Conference on Learning Theory*, volume 178 of *Proceedings of Machine Learning Research*, pages 573–624. PMLR, 2022.

Junha Chun, Youngjoon Jeong, and Taesup Kim. Sparse imagination for efficient visual world model planning. In *International Conference on Learning Representations*, 2026.

Ethan M. Clark. Reinforcement learning for commutative markov decision processes. Master's thesis, Arizona State University, May 2025. M.S. thesis.

Ethan M. Clark. Commutative reinforcement learning. GitHub repository, 2026. URL `https://github.com/ethanmclark1/commutative_rl`. Public implementation associated with a concurrent non-archival TMLR submission.

David A. Cohn, Zoubin Ghahramani, and Michael I. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996.

Volker Diekert and Grzegorz Rozenberg, editors. *The Book of Traces*. World Scientific, 1995. ISBN 978-981-02-2058-7.

Chuchu Fan, Zhenqi Huang, and Sayan Mitra. Approximate partial order reduction. In *Formal Methods*, pages 588–607. Springer, 2018. doi: 10.1007/978-3-319-95582-7_35.

Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite markov decision processes. In *Conference on Uncertainty in Artificial Intelligence*, pages 162–169, 2004.

Michał Fica and Jan Otop. Active automata learning with advice. In *Proceedings of the 28th European Conference on Artificial Intelligence*, pages 1655–1662, 2025. doi: 10.3233/FAIA250992.

Cormac Flanagan and Patrice Godefroid. Dynamic partial-order reduction for model checking software. In *Proceedings of the 32nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, pages 110–121. ACM, 2005. doi: 10.1145/1040305.1040315.

Robert Givan, Thomas Dean, and Matthew Greig. Equivalence notions and model minimization in markov decision processes. *Artificial Intelligence*, 147(1–2):163–223, 2003. doi: 10.1016/S0004-3702(02)00376-4.

Patrice Godefroid. *Partial-Order Methods for the Verification of Concurrent Systems: An Approach to the State-Explosion Problem*, volume 1032 of *Lecture Notes in Computer Science*. Springer, 1996. doi: 10.1007/3-540-60761-7.

David Ha and Jürgen Schmidhuber. World models. *arXiv preprint arXiv:1803.10122*, 2018.

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*, pages 2555–2565. PMLR, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

Danijar Hafner, Jimmy Ba, Mohammad Norouzi, and Timothy Lillicrap. Mastering atari with discrete world models. In *International Conference on Learning Representations*, 2021.

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.

Nicklas Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In *International Conference on Learning Representations*, 2024.

Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.

Lihong Li, Thomas J. Walsh, and Michael L. Littman. Towards a unified theory of state abstraction for MDPs. In *Proceedings of the International Symposium on Artificial Intelligence and Mathematics*, 2006.

Yichao Liang, Dat Nguyen, Cambridge Yang, Tianyang Li, Joshua B. Tenenbaum, Carl Edward Rasmussen, Adrian Weller, Zenna Tavares, Tom Silver, and Kevin Ellis. Exopredicator: Learning abstract models of dynamic worlds for robot planning. In *International Conference on Learning Representations*, 2026.

Vincent Micheli, Eloi Alonso, and François Fleuret. Transformers are sample-efficient world models. In *International Conference on Learning Representations*, 2023.

Doron Peled. All from one, one for all: On model checking using representatives. In *Computer Aided Verification*, volume 697 of *Lecture Notes in Computer Science*, pages 409–423. Springer, 1993. doi: 10.1007/3-540-56922-7_34.

Julian Quevedo, Ansh Kumar Sharma, Yixiang Sun, Varad Suryavanshi, Percy Liang, and Sherry Yang. Worldgym: World model as an environment for policy evaluation. In *International Conference on Learning Representations*, 2026.

Balaraman Ravindran and Andrew G. Barto. Model minimization in hierarchical reinforcement learning. In *Proceedings of the Fifth Symposium on Abstraction, Reformulation and Approximation*, volume 2371 of *Lecture Notes in Computer Science*, pages 196–211. Springer, 2002.

Balaraman Ravindran and Andrew G. Barto. Approximate homomorphisms: A framework for non-exact minimization in markov decision processes. In *Proceedings of the Fifth International Conference on Knowledge Based Computer Systems*, 2004.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy Lillicrap, and David Silver. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588: 604–609, 2020. doi: 10.1038/s41586-020-03051-4.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, pages 1015–1022, 2010.

Antti Valmari. Stubborn sets for reduced state space generation. In *Advances in Petri Nets 1990*, volume 483 of *Lecture Notes in Computer Science*, pages 491–515. Springer, 1991. doi: 10.1007/3-540-53863-1_36.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems*, 2017.

Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *International Conference on Learning Representations*, 2021.

Hongxin Zhang, Zeyuan Wang, Qiushi Lyu, Zheyuan Zhang, Sunli Chen, Tianmin Shu, Behzad Dariush, Kwonjoon Lee, Yilun Du, and Chuang Gan. COMBO: Compositional world models for embodied multi-agent cooperation. In *International Conference on Learning Representations*, 2025.

Zijie Zhao, Honglei Guo, Shengqian Chen, Kaixuan Xu, Bo Jiang, Yuanheng Zhu, and Dongbin Zhao. Empowering multi-robot cooperation via sequential world models. In *International Conference on Learning Representations*, 2026.

A Exact benchmark dynamics and stochastic coupling

The main experiments use the frozen main stochastic two-robot shared-intersection benchmark. We document the exact transition equations here because the central claim depends on order-sensitive *stochastic* dynamics rather than a hand-labeled commutativity relation.

The physical state is $s = (q_1, q_2)$, where each $q_i \in [-1.2, 1.2]$ is the signed position of robot i on its lane and the shared intersection is at zero. The action set contains advance and retreat for each robot. For an action applied to robot i with direction $d \in \{-1, +1\}$, let $j \neq i$. The shared-zone slowdown factor is

$$h(q_i, q_j) = \exp\left[-\frac{1}{2}\left(\frac{q_j}{0.32}\right)^2\right] \exp\left[-\frac{1}{2}\left(\frac{q_i}{0.55}\right)^2\right], \qquad c_{\mathrm{occ}} = 1 - 0.72\, h. \tag{12}$$

A second centrality variable controls asynchronous safety/communication delay,

$$c_{\mathrm{delay}} = \exp\left[-\frac{1}{2}\left(\frac{q_i}{0.42}\right)^2\right] \exp\left[-\frac{1}{2}\left(\frac{q_j}{0.42}\right)^2\right], \qquad p_{\mathrm{delay}} = 0.05 + 0.38\, c_{\mathrm{delay}}. \tag{13}$$

For independent exogenous draws $\xi_{\mathrm{act}}, \xi_{\mathrm{jit}} \sim \mathcal{N}(0,1)$ and $u \sim U[0,1]$, define

$$\eta = 0.24\big(1 + 0.05\, \xi_{\mathrm{act}}\big) \begin{cases} 0.58, & u < p_{\mathrm{delay}}, \\ 1, & \text{otherwise}, \end{cases} \tag{14}$$

and update

$$q_i' = \mathrm{clip}(q_i + d\, \eta\, c_{\mathrm{occ}} + 0.006\, \xi_{\mathrm{jit}}, -1.2, 1.2)\,, \qquad q_j' = q_j. \tag{15}$$

Event-keyed common random numbers. When comparing ab and ba, each physical action receives one exogenous tuple $(\xi_{\mathrm{act}}, \xi_{\mathrm{jit}}, u)$, and the same tuple is reused for that action in both serializations. Hence any difference between the two endpoints is caused by the serialization-dependent intermediate state, not by independently resampled disturbances. A single paired observation is

$$Z(s; a, b) = \mathrm{clip}\left(\frac{\|s_{ab}^{(2)} - s_{ba}^{(2)}\|_2}{\sqrt{2}\,(2 \cdot 0.24 + 4 \cdot 0.006)}, 0, 1\right), \tag{16}$$

and the main-benchmark swappability target is $\mu(s, a, b) = \mathbb{E}[Z(s; a, b)]$. The threshold is fixed at $\epsilon = 0.04$.

The candidate pool contains 200 initial states and all six unordered action pairs, for 1,200 diamonds. Half of the states are sampled broadly from $[-1, 1]^2$; half are sampled from a clipped Gaussian centered on the shared zone. Two same-robot pairs are globally safe in this benchmark, while the four cross-robot pairs are only approximately 46.5–47.0% safe, so the difficult part of the task is genuinely state dependent.

B ADD implementation details

Each action pair receives its own Gaussian process over the two-dimensional physical state. The frozen main-benchmark settings are: RBF length scale 0.6, GP normalized observation-noise parameter $\alpha = 0.4$, warm start 3 points/pair, pair-coverage floor 8, batch size 5, total discovery budget 150, and per-pair cap 30. The final conservative relation uses $\beta = 2.5$ and $\epsilon = 0.04$. The value $\beta = 2.5$ was chosen only from development seeds 1600–1603 and frozen before confirmatory evaluation.

For the historical exact variance-reduction instantiation, a GP posterior covariance K_t gives the one-step posterior variance after hypothetically observing candidate x,

$$\sigma_{t|x}^2(u) = \sigma_t^2(u) - \frac{K_t(u, x)^2}{\sigma_t^2(x) + \sigma_n^2}. \tag{17}$$

Let $r_t(u \mid x) = \sigma_t(u) - \sigma_{t|x}(u)$ and $w_t(u) = p_t^{\mathrm{safe}}(u) \exp(-(U_t(u) - \epsilon)_+/\tau)$. The frozen exact score used in the original confirmatory discovery experiment was

$$\alpha_t^{\mathrm{VR}}(x) = \frac{1}{\sqrt{n_{ab}}} \sum_{u \in \mathcal{U}_t^{ab}} w_t(u) r_t(u \mid x). \tag{18}$$

The target set uses $p_t^{\text{safe}}(u) > 0.35$, the boundary-proximity scale is $\tau = 0.012$, and pair-level utility is divided by $\sqrt{n_{ab}}$ to impose diminishing returns. The later component ablation and fresh confirmation show that exact global variance reduction is not empirically necessary for the discovery gain. We therefore treat Eq. (18) as a historical confirmatory instantiation, while Eq. (9) is the simpler empirically supported acquisition.

Important distinction between development diagnostics and the final relation. During confirmatory analysis on the main stochastic two-robot benchmark, the 40-item diagnostic shortlist was ranked with $\mu + 2\sigma$ to compare candidate quality at a fixed list size. The final learned relation used downstream is the frozen conservative threshold $\mu + 2.5\sigma \leq \epsilon$. We keep these two objects separate throughout the paper.

C CERTIFICATION STUDY AND WHY THE PAPER USES A BOUNDED EMPIRICAL CLAIM

Before building the full stochastic transition benchmark, the proxy certification study isolated the statistical confirmation problem. Figure 6 reports the resulting individually certified counts. Under an independent Gaussian observation model with $\sigma = 0.01$, a 150-probe discovery phase was followed by 60 independent confirmation probes. ADD certified 24.875 state-dependent diamonds on average versus 20.063 for Random, a 23.99% increase, with zero observed false certificates in either method over 16 fresh seeds. The one-sided paired Wilcoxon test gives $p = 0.003489$.

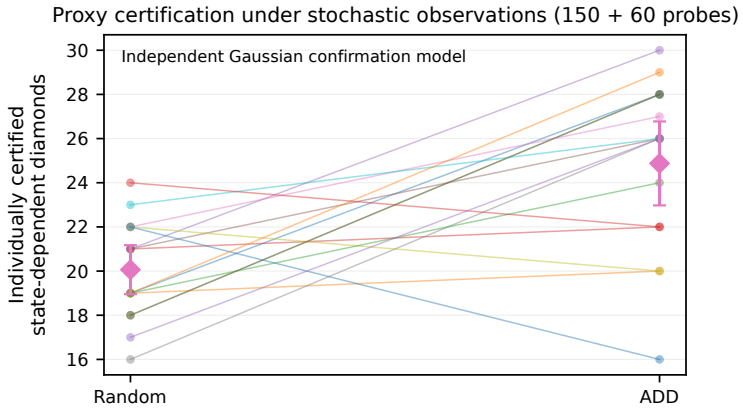


Figure 6: Proxy certification under independent Gaussian observations. This experiment tests the confirmation machinery under a stated Gaussian observation model; it is not presented as a physical deployment guarantee.

Figure 7 summarizes the failed propagation routes. We also evaluated stronger pointwise propagation mechanisms. A single global Lipschitz constant was too conservative in the genuinely state-dependent pairs, a learned local-Lipschitz rule propagated only a few additional cases, and an oracle-norm simultaneous RKHS bound certified zero additional state-dependent diamonds at the same budget. These negative results motivated the paper's more modest language: TRACE-WM uses an empirically inferred relation whose precision is evaluated on held-out states and does not claim a universal safety theorem over continuous state space.

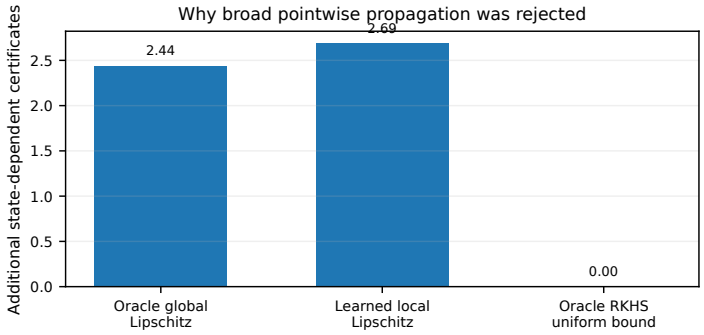


Figure 7: Additional state-dependent certificates produced by broad pointwise propagation schemes in the proxy certification study. The formal routes were rejected because they were numerically too conservative at the available intervention budget.

D METHOD DEVELOPMENT AND NEGATIVE RESULTS

The final method was not obtained by silently tuning until a preferred result appeared. Several versions failed and were frozen as negative evidence.

The proxy-study acquisition rule did not transfer directly to stochastic paired measurements. Figure 8 and Table 2 report the pre-calibration failure. When the proxy-study acquisition rule was applied without correcting the GP noise model, it did not beat Random at budgets 60, 90, 120, or 150. The paired discrepancy in the main stochastic two-robot benchmark has standard deviation roughly 0.02–0.03, whereas the earlier near-deterministic pool used a much smaller GP noise assumption. A development-only calibration sweep therefore preceded the confirmatory ADD experiment; Fig. 9 shows the noise calibration.

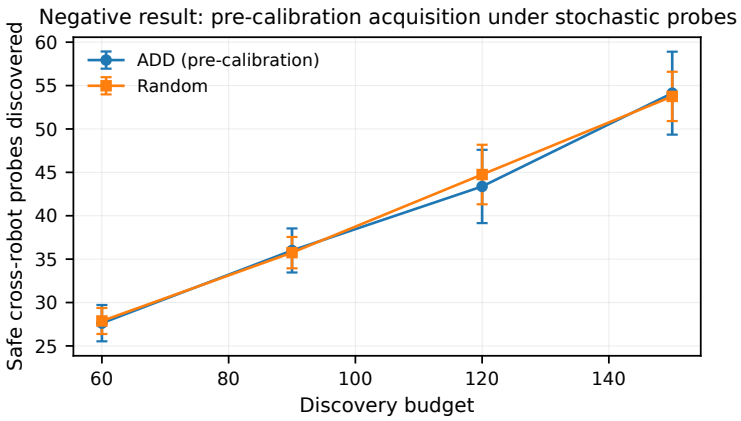


Figure 8: Negative result: the pre-calibration active rule did not beat Random across discovery budgets once observations became genuinely stochastic.

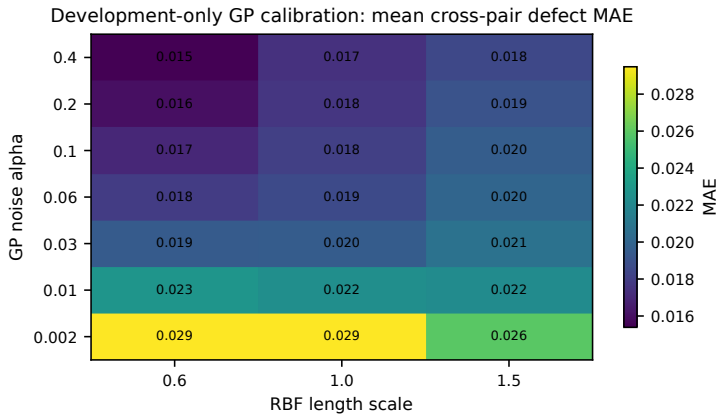


Figure 9: Development-only GP calibration. The stochastic observation scale materially changes defect reconstruction; the final $\alpha = 0.4$, length-scale 0.6 setting was frozen before fresh-seed testing.

Table 2: Budget sweep for the pre-calibration active rule. "Safe probes" counts genuinely safe cross-robot interventions found during discovery; shortlist columns report safe items among the final 40 candidates.

Budget	ADD safe probes	Random safe probes	ADD shortlist	Random shortlist
60	27.62	27.88	37.25	38.00
90	36.00	35.75	37.62	36.62
120	43.38	44.75	37.75	39.62
150	54.12	53.75	38.75	39.88

Graph transfer variants. A naive 40-probe transfer mechanism that appended the frozen two-robot prediction as ordinary input features did not outperform learning the multi-agent relation from scratch. A second version combined transferred and residual epistemic uncertainties and became excessively conservative. The final Graph-TRACE construction therefore uses the frozen two-robot predictor as a *mean function* and lets only the residual GP control multi-agent epistemic uncertainty.

E Full stochastic two-robot discovery results

Table 3 reports the complete seed-level discovery results. The confirmatory stochastic discovery experiment uses seeds 1700–1715 and the same 150 paired-probe budget for ADD and Random. The active full-dynamics baseline receives the same intervention budget but learns one-step dynamics and derives swappability post hoc.

Table 3: Seed-level stochastic discovery results. "Probe safe" is the number of genuinely safe cross-robot interventions found during discovery. "Shortlist safe" is the number of safe cases in the final 40-item diagnostic shortlist.

Seed	Probe safe			Shortlist safe / 40		
	ADD	Random	FullDyn	ADD	Random	FullDyn
1700	81	56	72	36	36	40
1701	90	63	69	40	40	36
1702	80	56	72	40	40	35
1703	55	42	67	40	37	35
1704	79	58	72	40	40	38
1705	74	50	74	40	40	33
1706	72	53	65	40	40	29

Seed	Probe safe			Shortlist safe / 40		
	ADD	Random	FullDyn	ADD	Random	FullDyn
1707	78	56	71	40	40	38
1708	87	54	70	40	40	33
1709	72	60	76	40	40	38
1710	72	56	78	40	40	39
1711	73	55	73	40	40	34
1712	60	54	73	40	39	38
1713	73	55	72	40	39	32
1714	73	57	69	40	40	39
1715	62	60	69	36	40	40

Across the 16 seeds, ADD finds 73.8125 safe cross-robot interventions versus 55.3125 for Random (+33.44%), winning all 16 paired seeds (one-sided Wilcoxon asymptotic $p = 2.17 \times 10^{-4}$ because absolute paired differences contain ties). The active full-dynamics learner finds 71.375 safe probes, so the raw ADD–FullDyn difference is not significant and is not claimed as a win. The stronger distinction is relation purity: ADD's final diagnostic shortlist averages 39.50/40 safe versus 36.06/40 for FullDyn (asymptotic $p = 0.00350$ because tied ranks are present).

F MATCHED WORLD-MODEL GENERALIZATION

Tables 4 and 5 summarize the aggregate and seed-level matched results. The matched held-out experiment uses seeds 5000–5015. For each seed, 160 states are available to discovery and world-model training and 40 states are held out from both. Each seed receives a fresh ADD, Random, and FullDyn relation before training a fresh probabilistic endpoint model.

Table 4: Aggregate held-out results across 16 matched seeds. Lower is better for both W2 and dependent order-effect MAE.

Model	Safe-swap W2	Dependent W2	Balanced W2	Dep. order MAE
Oracle	0.04666	0.03597	0.041317	0.02105
FullDyn relation	0.05145	**0.03635**	**0.043900**	0.02591
Global invariance	0.05154	0.04007	0.045807	0.04504
Random relation	0.06730	0.03840	0.052852	**0.01844**
TRACE-WM (ADD relation)	0.06428	0.04058	0.052431	0.01928
Ordered	0.11849	0.04033	0.079410	0.02038

TRACE-WM with the ADD relation reduces held-out balanced W2 from 0.079410 to 0.052431, a 33.97% reduction, with 16/16 desired-direction outcomes and exact one-sided Wilcoxon $p = 1.53 \times 10^{-5}$. Global invariance has dependent order-effect MAE 0.045042 versus 0.019280 for TRACE-WM, a 57.20% reduction, again with 16/16 desired-direction outcomes and $p = 1.53 \times 10^{-5}$. Random-derived TRACE-WM remains statistically tied downstream: 0.052852 versus 0.052431 balanced W2, with exact one-sided $p = 0.264$ for ADD improving over Random.

Figure 10 visualizes the seed-level relation precision. The expanded seed set also reveals relation-calibration variability. Mean per-seed ADD held-out precision is 97.3%, but pooled precision is 95.9%. Seeds 5006 and 5012 have held-out relation precision 76.09% and 80.39%, respectively; pair-level diagnostics attribute nearly all of these false-safe predictions to one cross-robot action pair that received only 8–9 discovery probes and was extrapolated too broadly. Within the original six-seed subset, seed 5001 admits only nine relation elements and has the largest TRACE-WM balanced W2 in that subset (0.0594), indicating that relation coverage is also a source of run-to-run variability.

Table 5: ADD-relation matched held-out seed-level results.

Seed	TRACE W2	TRACE order MAE	$\lvert\widehat{\mathcal{R}}\rvert$	Precision
5000	0.0474	0.0158	63	1.000
5001	0.0594	0.0177	9	1.000
5002	0.0512	0.0171	83	1.000
5003	0.0514	0.0194	82	1.000
5004	0.0498	0.0188	87	1.000
5005	0.0488	0.0211	84	1.000
5006	0.0505	0.0252	92	0.761
5007	0.0459	0.0187	86	1.000
5008	0.0496	0.0178	42	1.000
5009	0.0438	0.0174	56	1.000
5010	0.0555	0.0164	55	1.000
5011	0.0688	0.0207	47	1.000
5012	0.0462	0.0225	102	0.804
5013	0.0539	0.0206	71	1.000
5014	0.0517	0.0194	29	1.000
5015	0.0650	0.0199	47	1.000

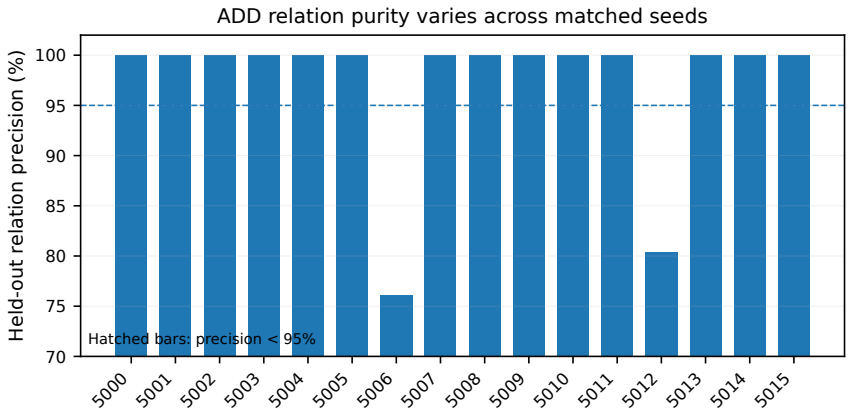


Figure 10: Held-out ADD-relation precision across the 16 matched seeds. Two seeds expose a concentrated cross-robot extrapolation failure rather than uniformly high precision.

G Post-confirmatory stress tests and ablations

Relation-size sweep. Figure 11 reports the relation-size sweep. Nested subsamples of the learned ADD training relation at 10, 20, 40, and approximately 68 swaps yield held-out balanced W2 of 0.063339, 0.060297, 0.056261, and 0.054007, respectively, compared with 0.073152 for zero augmentation and 0.051337 for the full relation. The original hypothesis of saturation near 30 swaps is not supported. The correct conclusion is diminishing returns: additional safe coverage continues to help through the largest tested subsample.

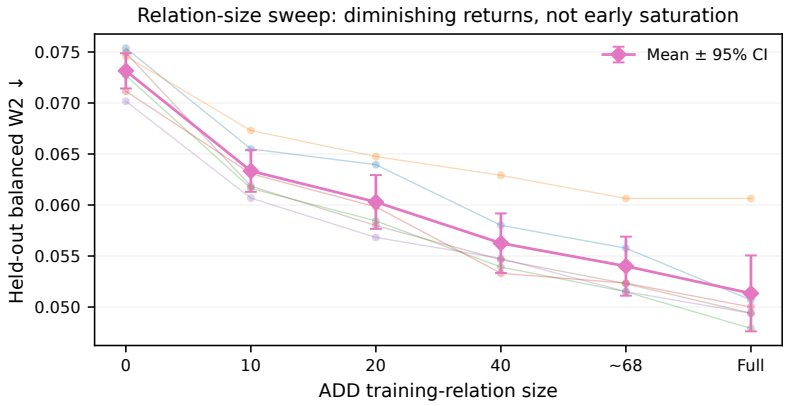


Figure 11: Relation-size sweep. Selective-invariance benefit shows diminishing returns but no early saturation at 30–40 swaps.

ADD acquisition ablation and fresh confirmation. Under the same 150-probe budget, the original exact variance-reduction ADD rule discovers 73.81 genuinely safe cross-robot probes on average versus 55.31 for Random. Removing boundary-proximity weighting reduces discovery to 52.38, removing the pair cap to 61.88, removing the coverage floor to 66.19, and removing safe-probability weighting to 69.63. Removing the diminishing-returns factor gives 74.81. A simpler local uncertainty surrogate discovers 95.13 safe probes in this ablation, so exact global variance reduction is not empirically supported as the source of the gain. Figure 12 summarizes the component ablation.

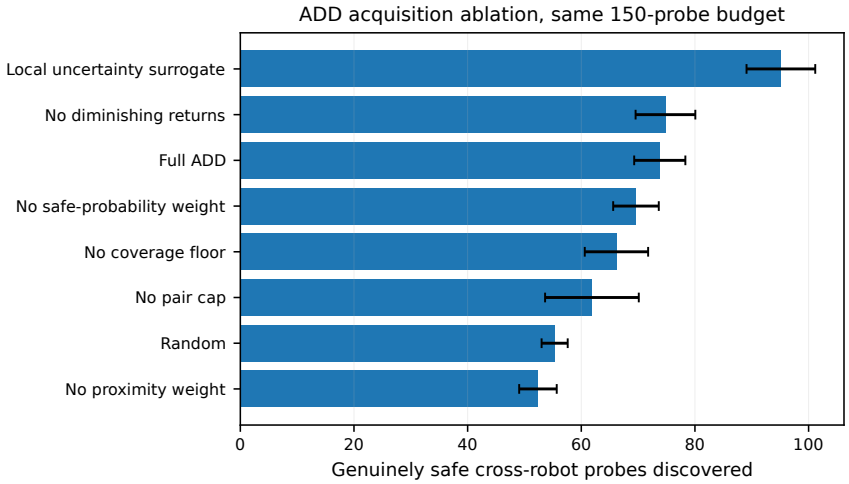
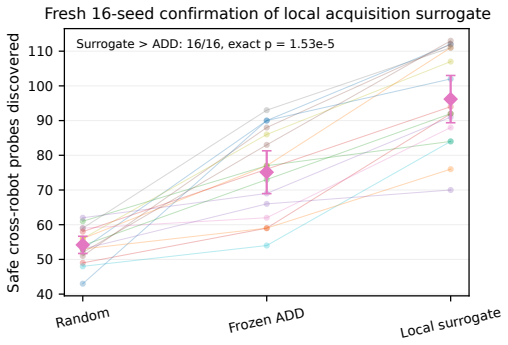
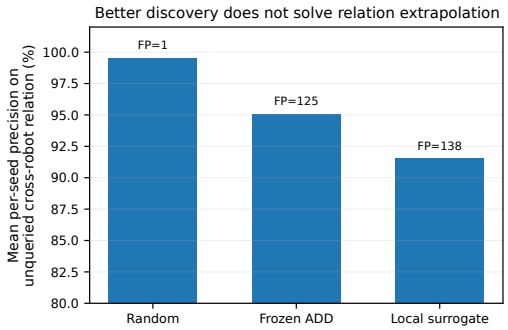


Figure 12: ADD acquisition ablation under the same intervention budget. The exact global variance-reduction term is not empirically necessary for the observed discovery gain.

We therefore froze a fresh comparison before observing seeds 8000–8015. The local-uncertainty score discovers 96.19 safe cross-robot probes, versus 75.13 for the original exact variance-reduction instantiation and 54.19 for Random. It wins 16/16 against both baselines (exact one-sided $p = 1.53 \times 10^{-5}$). However, mean per-seed precision on unqueried cross-robot relation elements is 91.5% for the local score, 95.1% for the original ADD rule, and 99.5% for Random. Figure 13 therefore separates discovery efficiency from relation extrapolation rather than treating the stronger acquisition as a certification fix.



(a) Safe-probe discovery on fresh seeds.



(b) Precision on unqueried cross-robot relation elements.

Figure 13: Fresh 16-seed confirmation of the local-uncertainty ADD acquisition. The simpler score improves discovery but does not solve relation extrapolation.

Relation-composition control. The original matched table does not show downstream ADD superiority over Random. To test whether this is caused by relation composition rather than early saturation, we rerun the comparison on seeds 5000–5015 with common world-model initialization and five independently sampled ADD subsets per seed. Each ADD subset is matched to the Random training-relation size when enough ADD elements are available. Size-matched ADD obtains held-out balanced W2 0.052013 versus 0.052595 for Random (13/16 seeds, exact one-sided $p = 0.0327$) and dependent order-effect MAE 0.018385 versus 0.018918 (12/16, $p = 0.0222$). Full ADD obtains balanced W2 0.052318 and does not beat the size-matched ADD subset. Figure 14 shows that which swaps are admitted matters in addition to how many are admitted.

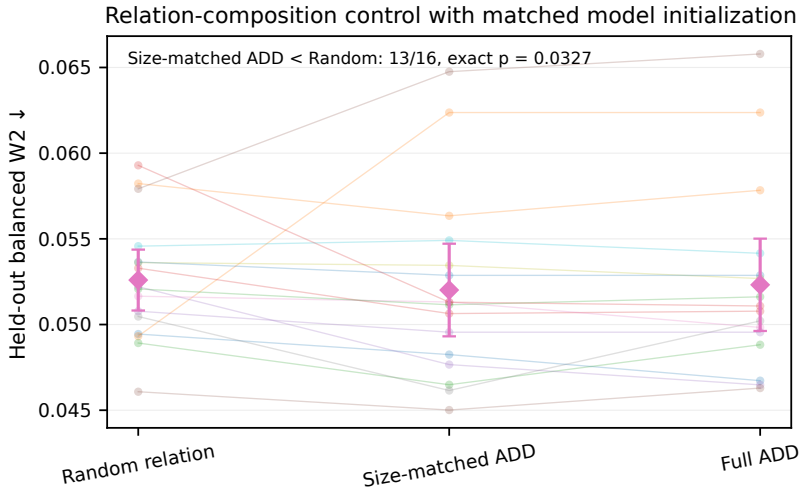
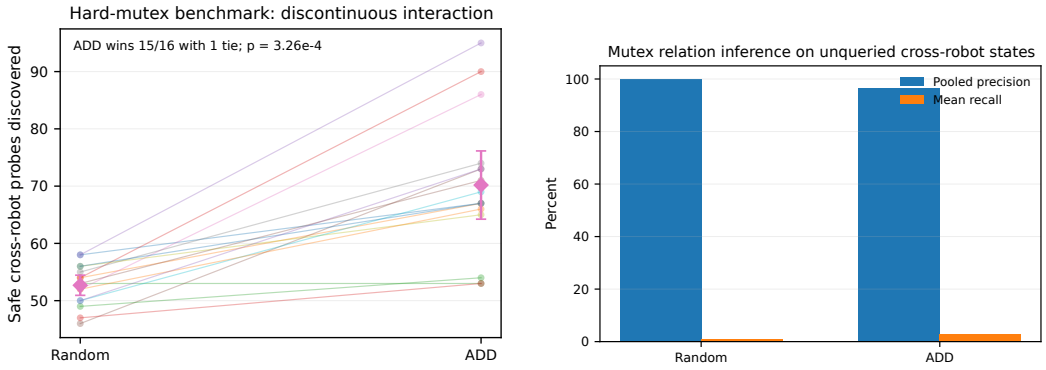


Figure 14: Size-matched relation-composition control with common world-model initialization. ADD-selected elements are modestly more useful than Random-selected elements at comparable relation size, while additional imperfect coverage does not improve over the size-matched subset.

Discontinuous hard-mutex benchmark. To test whether the discovery result depends on a smooth discrepancy surface, we add a hard shared-zone mutex to the original stochastic dynamics. An advance action is blocked discontinuously when it would enter a shared zone of half-width 0.18 while the other robot already occupies that zone. All original stochastic slowdown and delay terms remain, $\epsilon = 0.04$ is unchanged, and no discovery hyperparameter is retuned. Across fresh seeds 9000–9015, the original ADD instantiation discovers 70.19 safe cross-robot probes versus 52.69 for Random, a

33.2% gain with 15 wins, one tie, and no losses ($p = 3.26 \times 10^{-4}$ using the zero-aware signed-rank convention). On unqueried cross-robot states, pooled precision is 96.4% for ADD versus 100% for Random, while mean recall is 2.88% versus 1.04%. Figure 15 therefore supports robustness of active discovery to a discontinuous interaction without resolving the relation precision/coverage tradeoff.

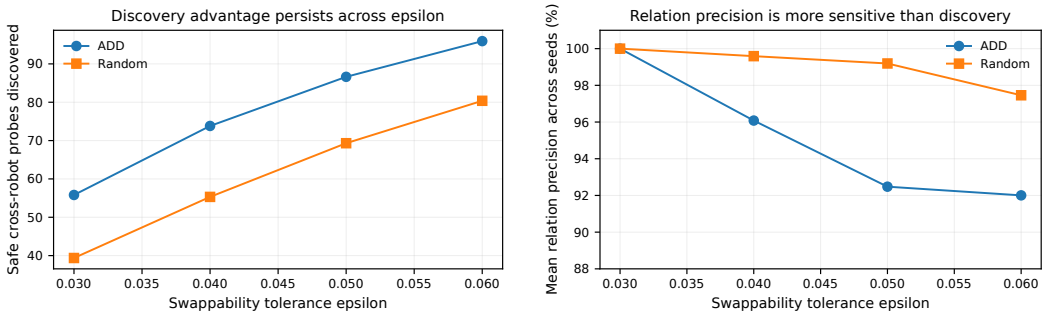


(a) Safe-probe discovery under hard blocking.

(b) Unqueried relation precision and recall.

Figure 15: Hard-mutex stress test. The active discovery advantage persists under a discontinuous shared-resource interaction, while relation extrapolation remains conservative and imperfect.

Sensitivity to ϵ. Figure 16 summarizes the tolerance sweep. At $\epsilon \in \{0.03, 0.04, 0.05, 0.06\}$, cross-robot safe prevalence is approximately 30.5%, 46.75%, 64.0%, and 78.1%. ADD exceeds Random in safe-probe discovery by 41.7%, 33.4%, 25.0%, and 19.4%, respectively. Relation generalization is less stable: ADD mean relation precision across the same tolerances is 100.0%, 96.1%, 92.5%, and 92.0%. This separates robust active discovery from the harder problem of generalizing a safe relation to unqueried states.

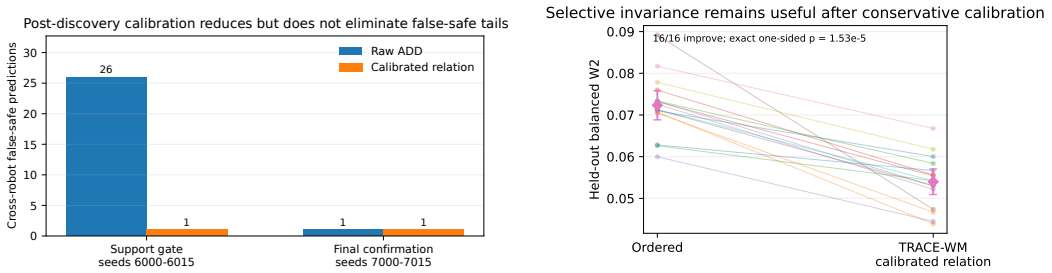


(a) Safe-probe discovery.

(b) Held-out relation precision.

Figure 16: Sensitivity to the swappability tolerance. The discovery advantage persists across tolerances, while relation precision is more sensitive to the safe-set boundary.

Fresh-seed relation-calibration audit. After the 16-seed precision failures were identified, we tested two conservative post-discovery gates on entirely new seed families without changing ADD's discovery policy. On seeds 6000–6015, requiring at least 12 cross-robot probes per pair and interpolation inside the queried-state convex hull reduces cross-robot false-safe predictions from 26 to 1 and raises pooled precision from 77.59% to 98.61%, but the minimum nonempty per-seed precision remains 92.31%. A second pre-specified confirmation on seeds 7000–7015 additionally raises the cross-robot UCB multiplier to $\beta = 3.0$. It yields 46 true positives and one false positive among 47 accepted cross-robot rewrites, pooled precision 97.87%, minimum nonempty per-seed precision 85.71%, and retains only 53.49% of raw ADD cross-robot true positives. It therefore fails the pre-specified 99% pooled precision, 95% minimum precision, and 60% coverage-retention gates.



(a) Cross-robot false-safe audit.

(b) Downstream W2 after calibration.

Figure 17: Fresh-seed relation calibration. Simple support gating removes catastrophic extrapolation but does not satisfy the pre-specified precision and coverage criteria. Selective invariance remains useful downstream despite this imperfect relation inference.

Figure 17 reports the fresh-seed calibration audit. The failed calibration gate is a negative result, not a tuned replacement for ADD. We therefore stop threshold tuning on these seed families and treat reliable out-of-sample certification of relation elements as an open problem.

H Local rewrite composition and trace error

Figure 18 accompanies the composition audit. A locally accepted rewrite at state s does not imply a global equivalence relation. We therefore evaluate what happens when an accepted local swap is followed by unseen suffix actions.

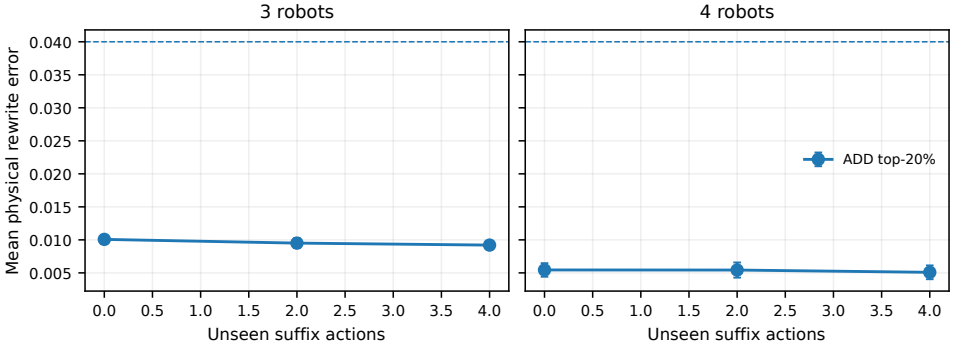


Figure 18: Physical error after appending unseen suffix actions to transferred multi-agent rewrites. The tested selective rewrites remain well below the local $\epsilon = 0.04$ reference on average.

A useful deterministic bound can be stated under a Wasserstein-Lipschitz assumption. If transition kernel P_a satisfies

$$W_1(P_a\mu, P_a\nu) \le L_a W_1(\mu, \nu), \tag{19}$$

then a local swap with error at most ϵ_j, followed by a suffix $a_{j+1}, \dots, a_T$, contributes at most

$$\epsilon_j \prod_{k=j+1}^{T} L_{a_k} \tag{20}$$

to the terminal trace error. By the triangle inequality, a sequence of local rewrites obeys

$$W_1(\tau, \tau') \le \sum_j \epsilon_j \prod_{k \in \mathrm{suffix}(j)} L_k. \tag{21}$$

When all $L_k \le 1$, this simplifies to the sum of local rewrite errors. We use Eq. (21) only as an interpretive bound; the benchmark does not assume known global contraction constants.

I GRAPH-TRACE DETAILS AND SENSITIVITY

For acting robots i, j, Graph-TRACE uses six contextual features:

$$[q_i, q_j,\ u_{0.32}, u_{0.42}, \max_k c_k, \sum_k c_k],$$

where $u_r = 1 - \prod_{k \neq i,j}(1 - c_k(r))$ is a permutation-invariant union-occupancy statistic for surrounding robots at radius r. Figure 19 reports the development-only multiplier sweep. The residual GP uses length scales $[0.7, 0.7, 0.8, 0.8, 0.8, 0.8]$, normalized noise $\alpha = 0.4$, and a 40-probe calibration budget stratified evenly across the four cross-robot direction pairs. Thirty percent of multi-agent states are held out. The final relation uses $\beta = 2.5$.

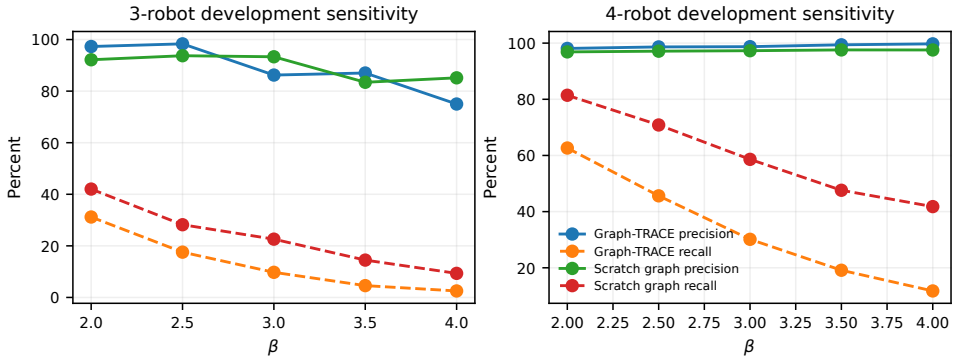


Figure 19: Development-only sensitivity to the conservative multiplier β. Solid curves show precision and dashed curves recall. The final $\beta = 2.5$ operating point was frozen before the 16-seed confirmation.

Table 6 reports the 3- and 4-robot relation-transfer results. Oracle local is the oracle restricted to the locally certifiable rewrite set, defined by the transferred local oracle score, rather than an oracle over the complete ground-truth safe relation. Its precision is therefore 100% on this restricted set, while recall is measured against the full ground-truth safe set. Oracle local can consequently have lower recall than a broader learned predictor without contradiction.

Table 6: Fresh-seed Graph-TRACE relation transfer, averaged over 16 seeds. Predicted and False safe are per-seed mean counts, while Precision and Recall are means of per-seed ratios. This explains why the ratio of the displayed mean counts does not equal the displayed mean precision for the pair-only rows. Recall uses the full ground-truth safe set as its denominator.

Robots	Method	Predicted	Precision (%)	Recall (%)	False safe
3	Pair-only	26.8	99.41	4.44	1.06
3	Graph-TRACE	144.0	97.53	24.09	4.62
3	Scratch graph	255.9	91.88	40.24	23.44
3	Oracle local	313.2	100.00	54.21	0.00
4	Pair-only	56.5	99.83	4.45	0.62
4	Graph-TRACE	547.0	98.31	42.94	9.56
4	Scratch graph	981.9	96.77	75.90	31.81
4	Oracle local	585.7	100.00	46.81	0.00

Graph-TRACE improves recall over pair-only zero-shot transfer by approximately 5.43× at three robots and 9.65× at four robots. Scratch contextual learning obtains more coverage but at lower purity, especially in the three-robot setting. We therefore treat Graph-TRACE as a scaling extension and not as a solved zero-shot generalization result.

J TRACE QUOTIENT AND EXACT PLANNING

Table 7 reports the trace-quotient results. The trace-quotient experiments generate legal asynchronous interleavings while preserving each robot's internal action order. Selective methods canonicalize only adjacent cross-robot swaps accepted by their learned relation. Global invariance canonicalizes all such swaps.

Table 7: Trace quotient averaged over eight fresh seeds.

Robots	Method	Compression (%)	Mean rewrite error	> 0.04 (%)
3	Pair-only	17.60	0.00361	0.98
3	Graph-TRACE	17.45	0.00379	1.17
3	Scratch graph	23.07	0.00517	2.73
3	Global	96.67	0.03302	26.56
4	Pair-only	3.36	0.00237	1.37
4	Graph-TRACE	24.11	0.00774	5.08
4	Scratch graph	65.99	0.01902	14.84
4	Global	98.33	0.02294	20.51

The exact planning proof of concept assigns two forward moves to each robot and optimizes only the cross-robot interleaving. The legal search spaces contain 90 interleavings for three robots and 2,520 for four robots.

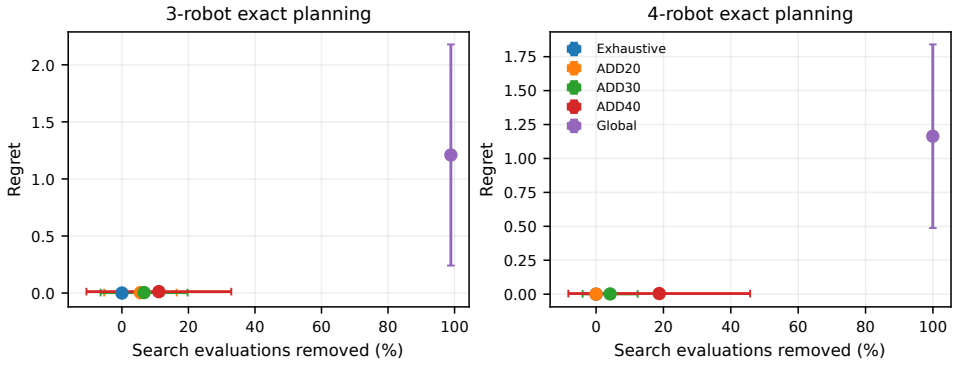


Figure 20: Exact planning compute–quality tradeoff. Selective quotienting removes a modest fraction of evaluations with small regret; global invariance removes almost all evaluations but loses important priority structure.

Table 8: Exact planning summary over four tasks per agent count.

Robots	Method	Evaluations	Reduction (%)	Regret
3	Exhaustive	90.0	0.00	0.0000
3	ADD20	85.0	5.56	0.0039
3	ADD30	84.0	6.67	0.0039
3	ADD40	80.0	11.11	0.0123
3	Global	1.0	98.89	1.2099
4	Exhaustive	2520.0	0.00	0.0000
4	ADD20	2520.0	0.00	0.0000
4	ADD30	2415.0	4.17	0.0017
4	ADD40	2047.5	18.75	0.0045
4	Global	1.0	99.96	1.1639

Figure 20 and Table 8 summarize the planning study. It is intentionally presented as supporting evidence rather than a headline result: the learned quotient yields real but moderate search reduction, particularly for four robots, and is conservative in difficult shared-zone states.

K STATISTICAL TESTING PROTOCOL

All primary comparisons are paired by random seed. One-sided Wilcoxon signed-rank tests are reported only for hypotheses whose direction was fixed before the corresponding confirmatory seed batch. We use the exact signed-rank distribution when there are no zero differences or tied absolute ranks. When ties or zero differences invalidate the simple exact distribution, we report the asymptotic Wilcoxon value and label it explicitly. We also report desired-direction paired wins because small samples can make rank-test values coarse.

The primary confirmatory tests are:

- ADD safe discovery $>$ Random on seeds 1700–1715: 16/16 wins, asymptotic $p = 2.17 \times 10^{-4}$ because tied absolute ranks are present.
- ADD shortlist purity $>$ FullDyn: asymptotic $p = 0.00350$ because tied ranks are present.
- TRACE-WM held-out balanced W2 $<$ Ordered on seeds 5000–5015: 16/16 wins, exact $p = 1.53 \times 10^{-5}$.
- TRACE-WM dependent order-effect MAE $<$ Global on seeds 5000–5015: 16/16 wins, exact $p = 1.53 \times 10^{-5}$.
- ADD-derived TRACE-WM balanced W2 $<$ Random-derived TRACE-WM: 12/16 desired-direction outcomes, exact $p = 0.264$; no downstream superiority is claimed.
- Fresh local-uncertainty ADD safe discovery $>$ original exact variance-reduction ADD on seeds 8000–8015: 16/16 wins, exact $p = 1.53 \times 10^{-5}$; local ADD $>$ Random on the same seeds: 16/16, exact $p = 1.53 \times 10^{-5}$.
- Size-matched ADD relation W2 $<$ Random relation under common world-model initialization on seeds 5000–5015: 13/16 desired-direction outcomes, exact $p = 0.0327$; dependent order-effect MAE: 12/16, exact $p = 0.0222$.
- Hard-mutex ADD safe discovery $>$ Random on seeds 9000–9015: 15 wins and one zero difference; the zero-aware one-sided Wilcoxon value is $p = 3.26 \times 10^{-4}$.
- Graph-TRACE precision $>$ ScratchGraph40 at three robots: asymptotic $p = 4.91 \times 10^{-4}$ because zero differences are present; at four robots: exact $p = 3.05 \times 10^{-5}$.

We do not claim significance where the data do not support it. ADD's raw safe-probe count is not significantly larger than the active full-dynamics baseline, and the fresh relation-calibration gate does not satisfy its pre-specified precision/coverage success criteria.

L REPRODUCIBILITY CHECKLIST

Table 9: Frozen core hyperparameters.

Component	Frozen setting
Physical position limit	$Q_{\max} = 1.2$
Nominal action step	0.24
Shared-zone coupling	0.72
Occupancy radii	0.32 (other), 0.55 (self)
Actuation noise scale	0.05
Position jitter scale	0.006
Delay probability	$0.05 + 0.38c_{\text{delay}}$
Delayed step multiplier	0.58
Delay radius	0.42
Swappability threshold	$\epsilon = 0.04$
ADD GP kernel	fixed RBF, length scale 0.6
ADD GP noise	$\alpha = 0.4$
ADD conservative multiplier	$\beta = 2.5$, selected on development seeds only
Warm start / coverage floor	3 / 8 per pair
ADD budget / batch / cap	150 / 5 / 30
TRACE-WM network	2 tanh layers, width 64
TRACE-WM likelihood	diagonal Gaussian NLL
World-model training epochs	160
World-model optimizer	Adam, 3×10^{-3}, weight decay 2×10^{-5}
Graph-TRACE calibration	40 probes, evenly stratified
Graph-TRACE GP noise	$\alpha = 0.4$
Graph-TRACE length scales	$[0.7, 0.7, 0.8, 0.8, 0.8, 0.8]$

World-model inputs and outputs. Table 9 lists the frozen hyperparameters. The controlled probabilistic model receives the two physical state coordinates and one-hot encodings of the ordered two-action sequence. It predicts a two-dimensional endpoint Gaussian mean and diagonal log variance. Training uses Gaussian negative log likelihood. In the main stochastic two-robot evaluation, target endpoint moments are estimated with 250 Monte Carlo CRN rollouts per candidate.

Anonymous supplementary artifacts. The supplementary package contains the benchmark construction, paired-probe simulator, ADD discovery code, probabilistic TRACE-WM implementation, Graph-TRACE calibration code, trace-quotient source, frozen CSV outputs, and the analysis artifacts included with this release. The final anonymous release will include only components that have been re-executed or explicitly documented as archived outputs.

M CLAIM BOUNDARY AND REVIEWER-ORIENTED FAILURE CASES

The evidence supports a selective-invariance claim, not a universal commutativity theorem. The following boundaries are explicit:

1. **No deployment-safety guarantee.** Finite-pool precision and the proxy certification study do not establish safety on arbitrary continuous states or real hardware.
2. **No claim that ADD dominates full system identification on every metric.** FullDyn has slightly better raw prediction fit in some matched experiments and comparable raw safe-probe discovery.
3. **No claim that Graph-TRACE maximizes recall.** Scratch graph learning obtains larger raw coverage, but Graph-TRACE occupies a cleaner precision–coverage operating point with the same small calibration budget.
4. **No claim of free long-horizon equivalence.** Local rewrites can accumulate error and the reached state matters; Eq. (21) requires explicit regularity assumptions.
5. **No claim of high-fidelity robotics validation or latent-world-model scaling.** The current environments are controlled stochastic engineering benchmarks, including a discontinuous hard-

mutex stress test, and the predictive model is a two-step Gaussian endpoint model rather than a visual latent sequence model.

These limitations are not hidden failure modes; they define the scope in which the empirical claims are supported.