ACTIONS ARE EXPERIMENTS: IDENTIFYING ACTION SEMANTICS FOR PREDICTION AND DECISION

Anonymous authors
Paper under double-blind review

ABSTRACT

The same numeric action can have different physical effects when a robot's controller, coordinate frame, or body changes. We study the resulting ambiguity in action-conditioned prediction: a model must identify what its commands currently mean. We represent this information as an action-semantic state inferred from short response queries, and instantiate it with local affine operators and temporal response kernels. The contribution is a controlled empirical test of this representation, using fixed predictors, held-out interfaces, and paired wrong-state controls. On a joint holdout of UR5e and action programs, after training on Panda and Sawyer, the response state reduces standardized future MSE from 1.7784 to 0.0203; replacing it with a wrong-body state gives 1.8799. A separate probe-then-decide test has zero normalized regret on 72 supported target cases. Official LIBERO-plus and RoboTwin simulations provide further evidence with distinct limits: static and temporal future prediction on RoboTwin both fall short of the prespecified accuracy threshold, while a separate role-goal decision test shows lower endpoint error. We also report the unmet ranking, support, and contact criteria and account for duplicated task dynamics. These results support response-derived action conditioning for tested local role-space futures and decisions. They do not establish benchmark task success, a visual world action model, or a universally sufficient state.

1 INTRODUCTION

A robot receives a command vector, but the vector alone does not specify its physical intervention. Rotating the control frame changes a command's direction; changing controller gain changes its magnitude and transient response; changing the body can change the number of native action coordinates. A predictor exposed to identical observations and numeric commands can therefore face different futures. Our central claim is that *the current meaning of an action is information that a predictive state may need to retain*.

This issue precedes the choice between pixel, latent, and compact state prediction. World models learn consequences for planning (Ha & Schmidhuber, 2018; Hafner et al., 2019), and world action models connect predicted futures to robot actions (Wang et al., 2026; Shen et al., 2026). An action-conditioned model still needs to ground its command coordinates in the current system. We study this grounding directly: execute a fixed query program, measure its physical responses, and use a compact response-derived state to condition prediction or candidate selection (Figure 1).

Interaction can reveal how a robot currently responds to its commands. We turn those responses into an explicit *action-semantic state*, then test whether it helps a fixed model predict future effects and choose among candidate actions when the action interface changes. Matched and mismatched response states let us identify the contribution of current-system information.

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.

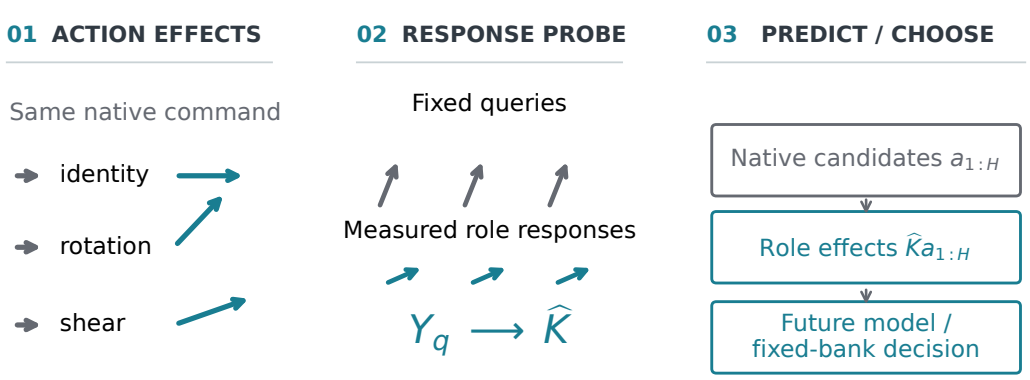

Figure 1: Actions identify their own interface. A fixed response program supplies the current native-to-role operator. The same operator conditions predictions for multiple candidate actions. Arrows are schematic; no image or video enters this state construction. Query budgets differ across experiments.

Our main contributions are:

- **Action semantics as state.** We derive a compact state from a robot's responses to short physical queries, capturing what its commands currently do.
- **Prediction across changing interfaces.** Conditioning on this state improves local future prediction when controllers, action interfaces, or robot bodies change.
- **Decision value beyond prediction error.** We show that the inferred state can improve action selection, evaluating candidate choices by their realized outcomes.

2 Related Work

Identification and adaptation. Classical system identification estimates dynamics through designed excitation (Ljung, 1999). UP-OSI conditions a universal policy on dynamics parameters inferred from recent state–action history (Yu et al., 2017); rapid motor adaptation learns an online adaptation module for control under changing conditions (Kumar et al., 2021). These works already establish the usefulness of interaction-derived conditioning. We study a narrower question: whether an explicit native-to-role response coordinate is useful under matched versus mismatched conditioning, including changes in native action dimension. We do not evaluate superiority over these adaptive-control algorithms or learn an active probe policy.

Predictive and decision-relative state. Predictive state representations describe systems through future tests (Littman et al., 2001; Singh et al., 2012); causal states group histories by predictive distributions (Crutchfield & Young, 1989). Control abstractions preserve distinctions relevant to rewards and transitions (Ferns et al., 2004; Gelada et al., 2019). Our operational definition is relative to a role, a query family, and a decision family. The affine operator is a candidate coordinate for that definition, not a proof that every relevant history has been compressed without loss. The interpretation also connects to sensorimotor accounts (O'Regan & Noë, 2001) and inverse-dynamics-shaped representations (Ivashkov et al., 2026).

World and action models. Fast-WAM studies the need for future imagination at inference (Yuan et al., 2026); LaWAM uses compact latent visual futures (Chen et al., 2026). Video-derived latent actions (Bruce et al., 2024; Garrido et al., 2026) and cross-embodiment conditioning (Wu & Gao, 2026) address related representation problems. Our experiments isolate physical command grounding using measured role responses. They do not constitute a performance comparison with those visual models.

3 Action Semantics as a Testable State Variable

Hidden interfaces. Let x be physical state, $a \in \mathcal{A}_e$ a native command for body e, and η the current action interface. A transition follows

$$x_{t+1} \sim P(\cdot \mid x_t, a_t, \eta_t). \tag{1}$$

If η_t is neither observed nor inferable from the supplied observation, conditioning only on (x_t, a_t) can merge distinct action effects. This is an information limitation of that conditioning choice, not a claim that every observation representation lacks interface information.

Decision equivalence and probe identifiability. Fix a role observation $r(x)$, physical context x, candidate bank $\mathcal{B}$, and utility family $\mathcal{U}$. Write $U_{x,\eta}(a)$ for expected utility and $A^*_{x,\eta}(U) = \arg\max_{a\in\mathcal{B}} U_{x,\eta}(a)$ for the optimal-action set. Define

$$\eta \sim_{x,\mathcal{B},\mathcal{U}} \eta' \quad \Longleftrightarrow \quad A^*_{x,\eta}(U) = A^*_{x,\eta'}(U) \quad \text{for every } U \in \mathcal{U}. \tag{2}$$

This is an equivalence relation by equality of the optimal-action sets. It defines a decision-semantic state $\sigma = [\eta]$ for the specified context and family. A fixed tie rule chooses among equal optima. Separately, a query program $q \in \mathcal{Q}$ yields responses Y_q. Inferring σ from Y_q requires the probes to distinguish decision-relevant classes. Indistinguishable response distributions with disjoint optimal-action sets preclude guaranteed recovery. We test particular query programs empirically; we do not assume identifiability from the definition.

For a learned coordinate $\hat{\sigma}$, decision sufficiency is measured by

$$\mathrm{Regret}(\hat{a}; x, \eta, U) = \max_{a\in\mathcal{B}} U_{x,\eta}(a) - U_{x,\eta}(\hat{a}), \qquad \hat{a} = \arg\max_{a\in\mathcal{B}} \widehat{U}(x, \hat{\sigma}, a). \tag{3}$$

Low operator error, low prediction error, and low decision regret are different empirical properties. For a two-action bank $\{a, b\}$, if two contexts mapped to the same representation have opposite preferences with margin $\delta > 0$, any deterministic decoder incurs regret at least δ in one context. For larger banks this two-action argument requires ruling out a third action that is good in both. Conversely, a uniform utility error bound $\sup_a |\widehat{U}(a) - U(a)| \le \epsilon$ gives regret at most 2ϵ. Appendix A gives the short proofs. These observations motivate paired controls and direct decision evaluation, rather than treating regression accuracy as a sufficiency certificate.

Local coordinates. A static local response is represented by an affine operator

$$c_e(a) = b_e + K_e a, \qquad K_e \in \mathbb{R}^{d_r \times d_e}. \tag{4}$$

The role dimension d_r can stay fixed while native dimension d_e changes. In the fixed-body controller studies we also fit a one-lag increment model,

$$\Delta r_t = K_0 a_t + K_1 a_{t-1} + b + \epsilon_t, \qquad K = K_0 + K_1. \tag{5}$$

For RoboTwin transients, step-response kernels $S_1, \ldots, S_H$ instead predict cumulative displacement through

$$\widehat{\Delta r}_{t+h} = \sum_{j=1}^{h} S_{h-j+1}(a_{t+j} - a_{t+j-1}), \tag{6}$$

with the pre-program command specified by calibration. These are experiment-specific coordinates. A static K is not assumed to encode contact modes, object state, or arbitrary temporal dynamics.

What makes this a state variable? A fixed action encoder $e(a)$ returns the same code for a numeric command wherever it is used. In contrast, $c_e(a)$ depends on responses measured at the current anchor. The inferred operator persists while the predictor evaluates multiple candidates: a new candidate changes a, whereas a new interface can change K_e even when the candidate is fixed. This distinction specifies the information exposed to the future model. It does not require identifying a unique physical controller, because different controllers may produce equivalent responses for the role and horizon under study. Conversely, an operator identified at one anchor need not remain valid after a large configuration change. The experiments test local reuse within their registered programs, rather than assuming a globally constant map.

4 EXPERIMENTAL DESIGN

Queries, fitting, and costs. Fixed-body interface experiments use a balanced 20-step contact-free response program. A standardized ridge observer with penalty 10^{-4} maps prefixes of length $L \in \{0, \dots, 6\}$ to a full-response operator. The rank-four family uses $L = 3$ after its recovery audit; controller studies use six responses. Single-arm body transfer uses a zero command for bias and signed joint-coordinate responses; LIBERO-plus uses a four-query affine prefix. RoboTwin calibrates temporal response on each body, including the target: 24 signed direction pairs fit the maps and eight validate them. Thus "short queries" does not mean a universal three-action budget. Appendix B specifies the separate protocols.

From offline fitting to a new interface. The procedure has separate identification and use phases. During observer fitting, training-interface response prefixes are paired with the operators estimated from their fuller response programs. A future predictor then fits training-program outcomes using the chosen response representation. At evaluation, the observer and future-predictor parameters remain frozen. The current interface supplies its own query responses, which produce one operator or kernel; this state is reused across the held-out programs or candidates. A new calibration is interaction with the target system, even though it is not a gradient update to the learned predictor. RoboTwin uses directly calibrated response convolution in place of the learned future ridge. Its target-body calibration must therefore be counted when interpreting transfer.

Within this protocol, a candidate selector can be written as

$$\hat{a}(g) = \arg\min_{a \in \mathcal{B}} \left\| \widehat{\Delta r}(a; \hat{\sigma}) - g \right\|_2, \tag{7}$$

for a desired role displacement g, as in the RoboTwin endpoint test. The realized-bank oracle replaces the predicted displacement by the executed one. It is an evaluation reference, not an input to the selector. The single-arm directional tests instead maximize projected progress; their registered utility normalizations are consequently kept separate from endpoint-error ratios.

For four-step future programs, fixed-body predictors use native actions and canonical effects $[a_{1:4}, \widehat{K} a_{1:4}]$; the cross-body predictor uses four shared-role effects without padding the state with a robot identifier. A standardized linear ridge is fit only on training conditions and programs. These small models isolate the conditioning variable. Targets are measured local displacement at registered horizons, standardized using training targets. The full-response reference uses more measurements; it is not a strict error lower bound under finite-sample fitting.

Controls and access restrictions. Native-action predictors omit response state. Training-mean state removes current-system information. Paired wrong state preserves the conditioning interface but supplies another system's response; its predictor checkpoint is reused without refitting. Raw-response bilinear controls test an unstructured alternative and generally have more parameters. Full-response and analytic-prefix controls test measurement and learned-prior effects. These are information controls, not a uniform capacity-matched benchmark. The LIBERO-plus native and semantic predictors do have equal parameter counts.

Core experiments read simulator role state, not RGB, future video, task reward, success labels, or hidden-interface identifiers. A separate ablation reads one initial image through a fixed encoder. Candidate outcomes are used for evaluation after selection rules are specified. Response queries are contact-free in the core interface studies; the initial decision anchors can retain existing contact, but this does not demonstrate learned contact-state sufficiency. Every reported confirmation uses its original split and evaluation criteria. Subsequent analyses do not change the interpretation of failed tests.

Holdouts and counting. The interface future pilot nominally has three held-out task labels, 16 maps, and four programs. An existing audit shows that its task labels duplicate free-space dynamics: 192 prediction rows represent 64 unique map–program units, and 768 ranking cases represent 256 map–direction units. We report those effective counts without claiming they are independent random samples. Other primary future sets contain 128 controller–map–program combinations, 24 frequency–program combinations, 64 UR5e anchor–program combinations, and 32 LIBERO-plus

Table 1: Two local decision tests with prespecified protocols. Gains and regret use each study's normalization and must not be pooled across studies. Full agreement is action-index agreement with the full-response rule.

Holdout	Cases / support	Gain vs. mean	Gain vs. wrong	Regret	Full agree.
Rotation	96/72	0.1719	0.4977	0.00116	1.000
UR5e body	72/72	0.2323	0.3174	0	1.000

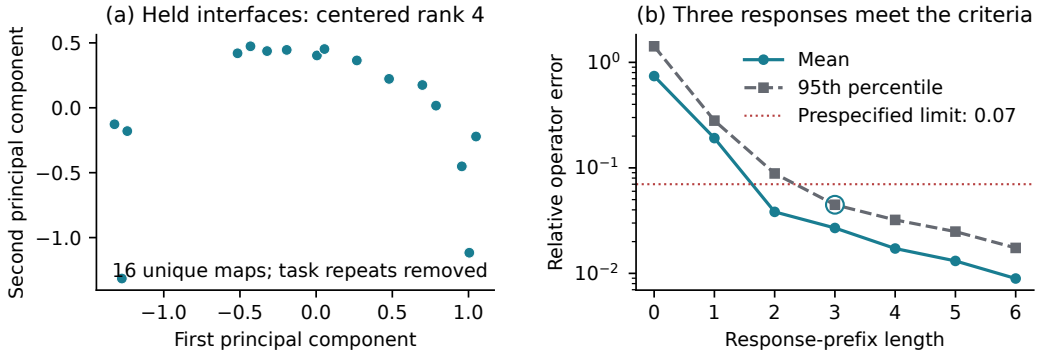

Figure 2: Recovery beyond a yaw shortcut. Left: PCA visualization of 16 unique held-out maps; centered rank is measured in the original four dimensions. Right: observer errors across 48 task–map rows. The three task labels do not supply 48 independent interfaces. Markers show point estimates, not confidence intervals.

initial-condition–program combinations. These are crossed, clustered designs sharing resets, bodies, or programs. The fixed-bank results are descriptive point estimates; no row-wise confidence intervals or population-level significance claims are attached.

5 Results: Grounding Commands, Then Testing Their Use

5.1 Recoverable action state can change decisions

The independent rotation confirmation uses four held-out task/scene combinations, 12 anchors, and eight rotations. It yields 0.00116 normalized regret and agrees with the full-response action in all 96 cases. Table 1 shows gains against no-response and wrong-state controls. These are evaluated-case averages; support is reported separately rather than silently filtering unsupported cases.

The representation also recovers interfaces beyond rotation. Sixteen held-out 2×2 maps span four centered matrix dimensions. Two responses exceed the prespecified tail-error limit (0.0885 versus 0.07); three are the shortest prefix meeting all recovery criteria, with mean relative error 0.0269 and tail error 0.0447 (Figure 2). This is recovery of the tested operator family, not evidence that three probes identify arbitrary interfaces.

A harder rank-four downstream test falls short of its decision criteria: support is 55/96 below the required 60, and worst-task exact agreement is $13/16 < 0.85$. Positive utility gains and 0.9426 gap closure do not satisfy these requirements. A post hoc comparison of the tested prefixes leaves support unchanged (Appendix D).

For body transfer, the common role is three-dimensional EEF motion. Panda and Sawyer have seven native arm coordinates; held-out UR5e has six. After response calibration at each new anchor, a fresh 96-candidate bank and 18 directions at four UR5e anchors produce 72 supported cases. The prefix selects the realized best bank action in every case, yielding zero regret versus 0.2323 for training-mean and 0.3174 for wrong-body state. This is calibrated local candidate selection, not uncalibrated transfer or task completion.

Table 2: Primary joint-holdout future MSE (lower is better). n counts unique evaluated combinations, not independent episodes. "Full" uses the full response. Gain/damping prediction improves, but its separate ranking and temporal-memory criteria are unmet.

Held-out axes	n	Native	State	Wrong state	Full
Interface + program	64	0.9435	0.0345	1.7060	0.0343
Gain/damping + map + program	128	0.7849	0.0382	0.2612	0.0380
Frequency + program	24	0.2524	0.0451	0.3050	0.0456
UR5e + program	64	1.7784	0.0203	1.8799	0.0224
Initial condition + program	32	0.14549	0.00146	0.08156	0.00139

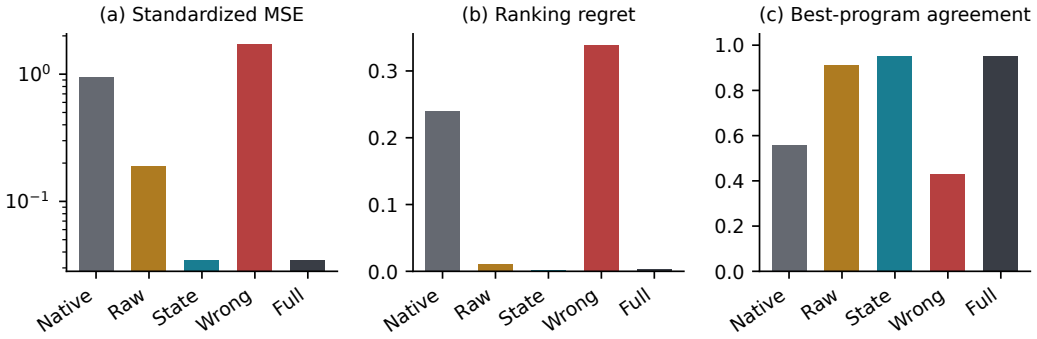

Figure 3: Interface/program pilot. Correct semantic conditioning improves future MSE and horizon-2 ranking. The paired wrong state reuses the same checkpoint. Prediction counts are 192/64 nominal/unique combinations; ranking counts are 768/256. Task labels are dynamically redundant. No error bars are inferred from those repeated labels.

5.2 Matched response state improves held-out future prediction

Table 2 collects five primary future tests. The repeated pattern is specific: correct response conditioning reduces error, and replacing that state while retaining the checkpoint removes much of the gain. MSE is standardized within each experiment; values across rows have different target scales and are not a common leaderboard.

Structured effects versus extra features. Figure 3 resolves the interface pilot into prediction and ranking. The native predictor has 54 fitted coefficients, the structured semantic predictor 102, and the raw-response bilinear control 378. The raw expansion includes response–action interactions, yet has MSE 0.1873 compared with 0.0345 for the structured state. The contrast therefore cannot be summarized as a larger model beating a smaller one. It also does not show that every unstructured architecture would fail: architecture and regularization remain part of this finite comparison. Replacing the state in the semantic checkpoint with a paired wrong interface raises MSE to 1.7060, while all learned weights stay fixed. Together, these controls support both the usefulness of the chosen structure and the importance of matching it to the current system.

The accompanying ranking comparison is a separate measured outcome, not an inference from MSE. It evaluates the same frozen future predictions against realized program consequences. Its apparent task diversity is limited by the duplicated free-space dynamics, so this positive result concerns held-out interfaces and programs. It cannot establish transfer to new object dynamics merely because nominal task names differ.

The UR5e/program result is the clearest body-transfer test: the future ridge fits only Panda/Sawyer and training programs, while current UR5e response queries supply its interface state. The predictor receives no body identifier. State conditioning reduces MSE by 98.86% versus padded native actions, and improves every held-out anchor and horizon. The much larger raw-response bilinear model is worse; this establishes the usefulness of the structured coordinate on this split, not a general sample-complexity theorem.

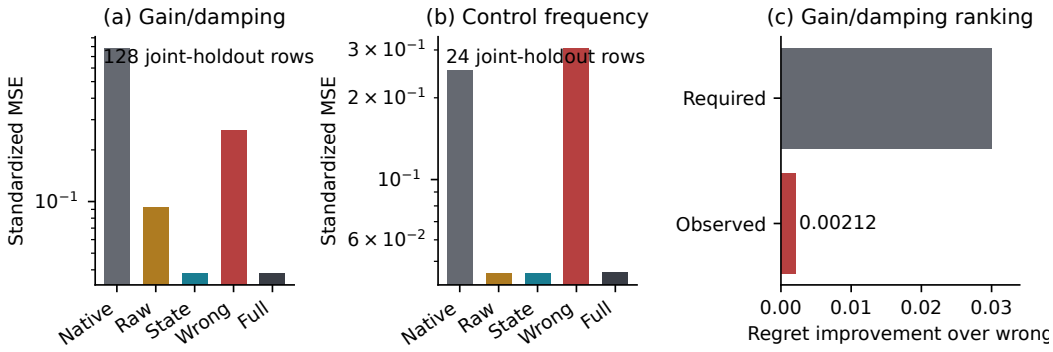

Figure 4: Two distinct controller interventions. Prediction improves under gain/damping and frequency changes. The gain/damping ranking contrast remains below its prespecified threshold. Frequency was evaluated for prediction only.

The controller interventions supply genuinely changing dynamics rather than duplicated task labels. Gain/damping and control frequency both support bounded future prediction. However, gain/damping's ranking improvement over the wrong state is only 0.00212 against a required 0.03, and its dynamic-over-static improvement is 9.71% against a required 10%. The combined prediction-and-ranking criterion is therefore unmet. The frequency replication was prediction-only and supplies no new ranking claim.

Figure 4 separates the two controller families and makes the failed ranking requirement explicit. Gain/damping changes transient response while the physical reset is held fixed; frequency changes the time discretization of control with the other mechanics settings held fixed. These interventions reduce the ambiguity of attributing a gain to task labels alone. Nevertheless, the value of temporal memory depends on which intervention is tested. A better regression score for a one-lag model on one axis cannot establish that memory is universally necessary, and neither prediction replication supplies evidence of closed-loop task success.

On one official LIBERO-plus task, two held Robot Initial States conditions crossed with held programs produce 98.99% lower MSE than equal-capacity native prediction. Pure background, camera, and light selectors leave the no-render response state unchanged; this checks nuisance invariance of a state that never reads images.

Can an initial image substitute for response state? A separate ablation fixes an ImageNet ResNet-18 encoder and uses one initial image per anchor. Image plus native actions gives MSE 1.11813; adding action state reduces it to 0.02633, while state plus actions alone gives 0.02624. Reusing the combined checkpoint with mean or wrong-body state raises MSE to 0.67775 or 1.13399. Thus the observed gain is associated with the current response state, and this frozen image representation does not supply a substitute. The comparison is restricted to the chosen encoder, initial observation, and local EEF target. It does not imply that images cannot reveal geometry, that richer visual histories cannot identify an interface, or that object-state perception is dispensable. Full protocol details are in Appendix C; no future image or visual policy is evaluated.

5.3 Prediction and decision results can diverge

Official RoboTwin 2.0 bodies (Chen et al., 2025) expose native dual-arm dimensions 12, 14, and 12 for ALOHA, Franka, and UR5, mapped into a common six-dimensional TCP role. Temporal response is independently calibrated on each body. The static and fresh temporal future tests obtain $R^2 = 0.467$ and 0.740, both below the prespecified 0.80 threshold. Neither establishes accurate absolute-future prediction under that criterion.

A separate decision test uses a fresh bank of 48 candidates and 16 role goals per body; no predictor is refit on those outcomes. Temporal state beats all five controls on median endpoint error for each body. Held UR5 has median selected-to-oracle error ratio 1.323 versus 1.725 for static and 2.625 for

Table 3: RoboTwin decision test: median per-goal selected/oracle endpoint-error ratio (lower is better; bank floor 1). Each column has 16 goals sharing one 48-candidate bank. Target-body response calibration is required. The separate absolute-future tests do not meet their accuracy criterion.

Conditioning / control	ALOHA	Franka	UR5
Temporal response	1.000	1.000	1.323
Static response	1.113	1.879	1.725
Native actions	2.387	2.467	2.625
Action PCA	2.716	1.898	2.183
Training mean	1.580	1.217	2.096
Wrong body	2.472	2.974	1.952

Table 4: Limitations under the original evaluation criteria. Improvements in one metric do not establish a result when other prespecified criteria are unmet.

Test	Observed result / unmet requirement	Interpretation
Rank-four decisions	Support $55 < 60$; task agreement $13/16 < 0.85$	Decision evidence insufficient
Gain/damping	Ranking contrast $0.00212 < 0.03$; memory gain $9.71\% < 10\%$	Prediction only
RoboTwin futures	Static/temporal $R^2 = 0.467/0.740 < 0.80$	Absolute future accuracy unestablished
Contact/object routes	Shelf tests fail the K regret comparison; later studies lack robust reversals	Contact-state claim unestablished

native conditioning (Table 3). Its oracle hits are 4/16. These results establish direct decision value for that calibrated role and bank, even though the broader absolute-future requirements failed. The future and decision data are different, so this is not a paired demonstration that a single regression threshold characterizes decision sufficiency.

5.4 Where the evidence remains limited

Table 4 summarizes the unmet criteria alongside the positive results. Contact pilots and engineering work did not establish a robust decision variable beyond free-space K. Some studies ended before a valid decision dataset existed; those runs cannot establish that contact state is unnecessary. Earlier estimator, mechanics, reset, and robustness failures are discussed in Appendix D.

6 Discussion and Limitations

The experiments support one operational conclusion: *when an interface can vary, measuring its responses can provide useful conditioning that a numeric command alone lacks*. The paired wrong-state controls, separate held-out programs, and changes in native action dimension make this claim more specific than a generic benefit from additional features. The results do not establish minimality of the operator representation, global Markov sufficiency, or superiority to modern adaptive-control methods. Query design and safety were engineered for the local simulator regimes, not proved for arbitrary robots.

Our strongest body results still require current-body calibration, and most other studies share one reset or a small number of anchors. Measured role state and exact simulator restoration simplify observation and distribution shift. Crossed programs and repeated goals are not independent episodes; nominal task labels can duplicate the same dynamics. We therefore report finite-design comparisons rather than statistical population claims. The full-response reference can be slightly worse than a prefix under finite fitting and measurement noise.

Connecting prediction error to selection. The relevant error for a candidate selector is how prediction changes comparisons within its bank. For the endpoint rule in Equation 7, let $e(a, g) = \|\Delta r(a) - g\|_2$ and let $\hat{e}$ use the predicted displacement. A uniform endpoint-prediction bound over

the bank gives

$$\max_{a \in \mathcal{B}} \|\widehat{\Delta r}(a) - \Delta r(a)\|_2 \le \epsilon \quad \Longrightarrow \quad e(\hat{a}, g) - \min_{a \in \mathcal{B}} e(a, g) \le 2\epsilon. \tag{8}$$

The implication follows from the reverse triangle inequality and the uniform-utility argument in Section 3. It clarifies the missing step between a regression score and a decision guarantee: average standardized MSE neither bounds the worst candidate error nor expresses that error in the physical units of the goal. Conversely, prediction errors that leave useful candidates correctly ordered can coexist with a good measured decision score. We do not estimate a new uniform bound from the opened tests. The reported prediction metrics and direct bank evaluations remain separate evidence, with each judged by its original contract. Likewise, local success gives no guarantee after the candidate bank, role, or calibrated operating region changes.

Decision margins. For a fixed goal and bank, let a_1 and a_2 be the best and second-best realized candidates, with gap $\gamma = e(a_2, g) - e(a_1, g)$. If every candidate displacement error is at most ϵ and $\gamma > 2\epsilon$, predicted endpoint errors preserve a_1 as the unique best action. Near a tie, the index can change with small regret. Thus UR5e oracle hits, RoboTwin ratios, and average future MSE answer different questions. The realized gap is unavailable to the selector, and these experiments do not estimate a prospective uniform error certificate. The condition clarifies what one would require; it does not upgrade a measured bank result.

The state is role-relative. Free-space EEF or TCP response does not specify object contact, articulation, dexterity, or long-horizon task success. The contact studies illustrate that observing a response change is insufficient to prove an additional decision-relevant coordinate. Similarly, a fixed initial-frame ablation cannot rule out information in richer vision or video representations. No result here demonstrates a complete visual world action model.

Reusing identified action effects. The response state separates identification of the current native-to-role map from evaluation of candidate actions. A fixed query program measures that map once at an anchor; the inferred state is then reused for many unseen action programs, so the predictor does not require a fresh response trial for every candidate. The paired wrong-state comparison is informative for this reason: it keeps the fitted model and candidate program in place while changing only the map supplied to them. On the joint body/program holdout, a shared EEF coordinate lets one fitted predictor evaluate UR5e programs despite a change in native action dimension. RoboTwin uses a temporal response kernel and a separate selector, but preserves the same identify-then-evaluate organization. This structure allows action grounding to condition a fixed prediction or decision system without refitting it at each interface.

The price and lifetime of calibration. The held-out-body results are conditional on responses collected from that body at each evaluated anchor. Freezing predictor weights does not make the test zero-interaction transfer: signed commands, resets, and role measurements provide information about the target before selection. The RoboTwin temporal calibration similarly includes target-body fit and validation queries. These costs would matter on hardware, where a query can consume time, disturb an object, or be unsafe; this study does not quantify them. Nor does one local operator specify when to recalibrate after a change in configuration, contact, or controller. A deployment comparison would have to count calibration actions and physical time alongside decision error, use the same interaction budget for its controls, and state a rule for refreshing the response state. Until then, the supported claim is improvement after the reported local calibration, not a net gain under an unmeasured operating budget.

7 Conclusion

Actions can serve as experiments that identify how commands map to physical consequences. An explicit response-derived state improves local predictions and fixed-bank decisions under the tested interface and body changes. Its sufficiency depends on the intended role and decision family; the present evidence does not establish the broader prediction, ranking, or contact claims discussed above.

REPRODUCIBILITY STATEMENT

The supplement specifies queries, splits, measurement units, controls, and unmet evaluation criteria. Figure and table generators read saved reports without training or simulator execution. The accompanying audit maps numerical claims to JSON fields and hashes, distinguishes current-file digests from historical bindings, and lists missing or changed dependencies explicitly. No threshold or experimental outcome was changed during manuscript preparation.

AI USE STATEMENT

During the preparation of this work, the authors used GPT6 to improve the clarity and grammar of the manuscript. No AI tools were used to generate scientific content, analyze data, or produce research ideas. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

Jake Bruce, Michael D Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, Yusuf Aytar, Sarah Maria Elisabeth Bechtle, Feryal Behbahani, Stephanie C.Y. 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 Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 4603–4623. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/bruce24a.html.

Jialei Chen, Kai Wang, Kang Chen, Shuaihang Chen, Feng Gao, Wenhao Tang, Zhiyuan Li, Weilin Liu, Zhuyu Yao, Boxun Li, Yuanbo Xu, and Chao Yu. LaWAM: Latent world action models for efficient dynamics-aware robot policies, 2026. URL https://arxiv.org/abs/2606.15768.

Tianxing Chen, Zanxin Chen, Baijun Chen, Zijian Cai, Yibin Liu, Zixuan Li, Qiwei Liang, Xianliang Lin, Yiheng Ge, Zhenyu Gu, Weiliang Deng, Yubin Guo, Tian Nian, Xuanbing Xie, Qiangyu Chen, Kailun Su, Tianling Xu, Guodong Liu, Mengkang Hu, Huan-ang Gao, Kaixuan Wang, Zhixuan Liang, Yusen Qin, Xiaokang Yang, Ping Luo, and Yao Mu. RoboTwin 2.0: A scalable data generator and benchmark with strong domain randomization for robust bimanual robotic manipulation, 2025. URL https://arxiv.org/abs/2506.18088.

James P. Crutchfield and Karl Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108, Jul 1989. doi: 10.1103/PhysRevLett.63.105. URL https://link.aps.org/doi/10.1103/PhysRevLett.63.105.

Senyu Fei, Siyin Wang, Junhao Shi, Zihao Dai, Jikun Cai, Pengfang Qian, Li Ji, Xinzhe He, Shiduo Zhang, Zhaoye Fei, Jinlan Fu, Jingjing Gong, and Xipeng Qiu. LIBERO-Plus: A progressive robustness benchmark for visual-language-action models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 38574–38583, June 2026.

Norman Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite markov decision processes. In *Conference on Uncertainty in Artificial Intelligence*, 2004. URL https://arxiv.org/abs/1207.4114.

Quentin Garrido, Tushar Nagarajan, Basile Terver, Nicolas Ballas, Yann LeCun, and Michael Rabbat. Learning latent action world models in the wild, 2026. URL https://arxiv.org/abs/2601.05230.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G. Bellemare. DeepMDP: Learning continuous latent space models for representation learning, 2019. URL https://arxiv.org/abs/1906.02736.

David Ha and Jürgen Schmidhuber. World models, 2018. URL https://arxiv.org/abs/1803.10122.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels, 2019. URL https://arxiv.org/abs/1811.04551.

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

Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. RMA: Rapid Motor Adaptation for Legged Robots. In *Proceedings of Robotics: Science and Systems*, Virtual, July 2021. doi: 10.15607/RSS.2021.XVII.011.

Michael L. Littman, Richard S. Sutton, and Satinder Singh. Predictive representations of state. In *Proceedings of the 15th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, pp. 1555–1561, Cambridge, MA, USA, 2001. MIT Press.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. LIBERO: Benchmarking knowledge transfer for lifelong robot learning. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 44776–44791. Curran Associates, Inc., 2023. doi: 10.52202/075280-1939. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/8c3c666820ea055a77726d66fc7d447f-Paper-Datasets_and_Benchmarks.pdf.

Lennart Ljung. *System Identification: Theory for the User*. Prentice Hall, 2 edition, 1999. URL https://rt.isy.liu.se/en/books/sysid/.

J. Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–973, 2001. URL https://static.cambridge.org/content/id/urn%3Acambridge.org%3Aid%3Aarticle%3AS0140525X01570119/resource/name/S0140525X01000115a.pdf.

Qiuhong Shen, Shihua Zhang, Yue Liao, Qi Li, Zhenxiong Tan, Shizun Wang, Shuicheng Yan, and Xinchao Wang. World action models: A survey, 2026. URL https://arxiv.org/abs/2606.20781.

Satinder Singh, Michael James, and Matthew Rudary. Predictive state representations: A new theory for modeling dynamical systems, 2012. URL https://arxiv.org/abs/1207.4167.

Siyin Wang, Junhao Shi, Zhaoyang Fu, Xinzhe He, Feihong Liu, Chenchen Yang, Yikang Zhou, Zhaoye Fei, Jingjing Gong, Jinlan Fu, Mike Zheng Shou, Xuanjing Huang, Xipeng Qiu, and Yu-Gang Jiang. World action models: The next frontier in embodied AI, 2026. URL https://arxiv.org/abs/2605.12090.

Zhuoyuan Wu and Jun Gao. OSCAR: Omni-embodiment action-conditioned world model for robotics, 2026. URL https://arxiv.org/abs/2606.04463.

Wenhao Yu, Jie Tan, C. Karen Liu, and Greg Turk. Preparing for the unknown: Learning a universal policy with online system identification. In *Proceedings of Robotics: Science and Systems*, Cambridge, Massachusetts, July 2017. doi: 10.15607/RSS.2017.XIII.048.

Tianyuan Yuan, Zibin Dong, Yicheng Liu, and Hang Zhao. Fast-WAM: Do world action models need test-time future imagination?, 2026. URL https://arxiv.org/abs/2603.16666.

Yuke Zhu, Josiah Wong, Ajay Mandlekar, Roberto Martín-Martín, Abhishek Joshi, Kevin Lin, Soroush Nasiriany, and Yifeng Zhu. robosuite: A modular simulation framework and benchmark for robot learning. In *arXiv preprint arXiv:2009.12293*, 2020.

A Decision Bounds and Scope of the Definition

Two-action necessity. Let $\mathcal{B} = \{a, b\}$. Suppose a deterministic representation maps two contexts h, h' to the same code and their utilities satisfy $U_h(a) - U_h(b) \geq \delta > 0$ and $U_{h'}(b) - U_{h'}(a) \geq \delta$. A deterministic decoder returns the same action in both contexts. Choosing a incurs at least δ regret in h'; choosing b incurs at least δ in h. Thus the representation cannot guarantee regret below δ in both. A randomized decoder choosing a with probability p has expected regrets at least $(1-p)\delta$ and $p\delta$, so their maximum is at least $\delta/2$. For a larger bank, a third action can avoid this contradiction; the preference reversal alone does not suffice unless the candidate restriction or a corresponding regret lower bound is established.

Uniform-error sufficiency. For $a^* \in \arg\max_a U(a)$ and $\hat{a} \in \arg\max_a \widehat{U}(a)$,

$$U(a^*) - U(\hat{a}) = U(a^*) - \widehat{U}(a^*) + \widehat{U}(a^*) - \widehat{U}(\hat{a}) + \widehat{U}(\hat{a}) - U(\hat{a}) \tag{9}$$

$$\leq \epsilon + 0 + \epsilon = 2\epsilon. \tag{10}$$

This requires a uniform bound over the candidate bank. An average future MSE is not such a bound, and utility sensitivity to future errors also matters.

Equivalence does not imply identifiability. Equation 2 is reflexive, symmetric, and transitive because it compares identical optimal-action sets for each utility. Its definition does not imply that a finite response prefix distinguishes different equivalence classes. If two interfaces give the same response distribution under every permitted probe, their response-derived codes have the same distribution. A decoder cannot be guaranteed to choose different, disjoint optimal actions from identical observations. Our tests therefore distinguish operator recovery, causal matching, and direct decision performance. The elementary bounds are explanatory observations, not a new theory of system identification.

B Experimental Protocols and Evaluation Units

The fixed-body studies use LIBERO (Liu et al., 2023); single-arm body tests use robosuite Lift (Zhu et al., 2020); the initial-condition replication uses official LIBERO-plus (Fei et al., 2026); dual-arm experiments use official RoboTwin 2.0 (Chen et al., 2025). These environments supply physics and robot assets. The reported metrics are local role-space predictions and candidate decisions, not their benchmark task-success scores.

Continuous interfaces. The independent rotation confirmation has 12 anchors from four unseen task/scene combinations and eight unseen rotations between -75° and 75°. The frozen observer is used without training on confirmation outcomes. The harder family uses $M = R(\theta)\,\mathrm{diag}(s_1, s_2)R(\phi)$, positive determinant, and anisotropic singular values between 0.72 and 1. There are 24 fit and 16 held-out maps crossed with three task labels per split. The observer's training and evaluation rows number 72 and 48. A sufficient prefix must meet mean relative error at most 0.03, tail error at most 0.07, mean polar error at most 1.5°, tail polar error at most 4°, and the other frozen improvement and consistency checks. Prefix length is audited over the fixed grid, not tuned on downstream outcomes.

The interface future model has 576 nominal fit rows and 192 primary test rows. Its posthoc duplicate audit reduces these to 192 and 64 unique map–program combinations. Future programs have four steps, with eight fit and four held-out programs per map. The targets are cumulative EEF-local displacement at horizons 1, 2, and 4. The corresponding primary ranking count is 768 nominal and 256 unique map–direction combinations. These counts describe the finite design; no claim of random-sample independence is made.

Gain/damping and frequency. At one physical reset with fixed task and frame, fit gain scales are $\{0.55, 1.00, 1.45\}$ crossed with damping ratios $\{0.75, 1.25\}$. Test gains $\{0.75, 1.25\}$ cross damping $\{0.90, 1.10\}$. Twelve fit and eight test maps cross eight fit and four test programs. The future model fits 576 rows and evaluates 128 primary controller–map–program combinations. Hidden gain and damping values are not model inputs. The ranking contrast and the optional dynamic-over-static comparison both fall short of their prespecified thresholds.

The prediction-only frequency replication holds gain, damping, task, map, and reset fixed. Fit frequencies 10, 20, and 40 Hz and test frequencies 15, 25, and 35 Hz cross 32 fit and eight held-out programs, giving 96 fit and 24 primary rows. Frequency labels are not model inputs. The six-response prefix conditions the dynamic model. This result does not resolve the separate gain/damping ranking failure.

Single-arm body transfer. Panda and Sawyer are fit bodies and UR5e is held out. A zero command estimates bias; one-sided native-coordinate queries form the prefix affine operator and signed responses form the fuller reference. Four new anchors per body cross 48 four-step programs. The predictor fits 256 Panda/Sawyer rows using programs 0–31. The primary set is four UR5e anchors crossed with programs 32–47, giving 64 rows. The role-effect input has dimension 12. Native actions are padded to dimension 28. No robot identifier is supplied. Both coordinate count and realized mechanics change, but each target anchor receives response calibration.

A separate decision replication uses four UR5e anchors, 96 candidates, and 18 target directions. All 72 cases meet the prespecified support minimum. Reported utility gains and regret are averages under the original per-case normalization. Near-zero oracle gaps motivate a separate support count and aggregate gap-weighted closure rather than an average of unstable per-case ratios. Earlier zero-drift mechanics and state-binding tests failed; the later positive result does not change their interpretation.

Official initial conditions. One task is fixed. Base and Robot Initial States difficulty 1/2 cross 32 fit programs, yielding 96 fit rows. Difficulty 4/5 cross 16 held-out programs for 32 primary rows. Programs have six steps. The four-query affine prefix is collected before future rollouts, and condition identifiers are excluded. Camera observations and both renderers are disabled. Exactly unchanged responses under appearance selectors concern this no-render state, not a visual model.

Dual-arm body tests. Native dimensions for ALOHA, Franka, and UR5 are 12, 14, and 12; the shared left/right TCP role has dimension six. Complete retained collision geometry is required. Temporal maps are fit from 24 signed direction pairs per body/arm and independently validated on eight pairs. Calibration includes the held body. The future tests each have 48 primary held-UR5/program rows (16 programs at three horizons); these are 16 program trajectories, not 48 independent episodes. Static and temporal future banks and the later decision bank are hash-disjoint. The decision test has 48 candidates and 16 goals per body. Goals share candidate outcomes, so they cannot be treated as independent fresh rollouts. There are no task objects, task rewards, success labels, camera observations, or demonstrations in this test.

The absolute-future tests require $R^2 \geq 0.80$. The temporal future test additionally requires at least 40% improvement over static response. The separate decision test requires held-UR5 median and 75th-percentile oracle ratios at most 1.35 and 2.0, at least 20% oracle hits, median excess-regret reductions of at least 25% over static and 50% over each other control, and lower median endpoint error for every body. Its positive result does not satisfy the unmet absolute-future requirements.

C Additional Controls and Generated Figures

Table 5: Raw-response controls omitted from the compressed main table. Standardized MSE is experiment-specific. The cross-body raw expansion is substantially larger; these rows are not capacity-matched comparisons.

Primary holdout	Raw response	Response state	Full response
Interface/program	0.1873	0.0345	0.0343
Gain/damping/map/program	0.0923	0.0382	0.0380
Frequency/program	0.0452	0.0451	0.0456
UR5e/program	64.4073	0.0203	0.0224
Initial condition/program	0.10408	0.00146	0.00139

The interface pilot fits 54 native, 378 raw-bilinear, and 102 semantic coefficients. The gain/damping pilot fits 54 native, 702 raw-bilinear, and 102 dynamic coefficients. The body predictor fits 261 padded-native, 6,417 raw-bilinear, and 117 semantic coefficients. The LIBERO-plus native and

semantic predictors both fit 171 coefficients; its raw-bilinear predictor has 2,763. A full-response reference can have higher test error than a shorter prefix because it is a finite noisy estimate passed through a fitted predictor.

Fixed-image ablation. One initial 64×64 image per anchor is encoded by frozen ImageNet ResNet-18 weights; a fixed Gaussian projection supplies 64 features. The ridge fits only 256 Panda/Sawyer rows and is evaluated on 64 UR5e/program rows. The image/action MSE is 1.11813; adding response state gives 0.02633; state/actions alone gives 0.02624. Mean-state and wrong-body substitutions in the combined checkpoint give 0.67775 and 1.13399. These are separate ablation data from the main cross-body future study. No future image is read. The result supports added information from action responses for this target, without showing that visual state is generally unnecessary.

D Limits of the Decision and Contact Evidence

Recovery is not a downstream certificate. The rank-four decision study evaluates 96 cases but only 55 meet the support criterion. The $L = 3$ state has 0.01195 regret, gains 0.1960 and 0.2895 against blind and wrong-state controls, and 0.9426 gap closure. The study nevertheless falls short of its combined requirements. A post hoc comparison over prefixes 1–6 leaves support at 55; longer prefixes cannot repair this evaluated split (Figure 5). Exact index agreement and near-equal utility remain different metrics.

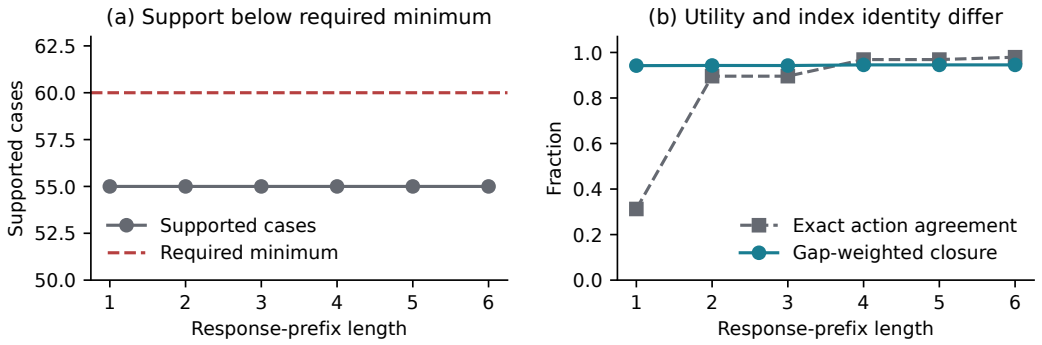

Figure 5: The rank-four decision test does not meet its support requirement. Comparing prefix lengths helps locate the limitation but does not constitute a new confirmatory test.

Earlier negative and incomplete tests. The archived reports include observer discovery and confirmation failures, failures of exact action stability under synthetic noise, earlier cross-body mechanics and state-binding failures, a wrist-load future failure, and incomplete visual, LIBERO-plus, and RoboTwin setups. Later exploratory diagnostics do not establish the claims those tests were intended to assess. An engineering failure prevents a scientific test and should not be interpreted as evidence against the response coordinate.

Object/contact routes. The first object-response studies change the measured response but do not improve held-out regret over the free-space K baseline. Subsequent laptop, switch, dump-bin, can, and bottle studies do not reach a valid decision comparison because of engineering, anchor, or native-budget constraints. In the seal study, both prespecified robot-only validation seeds exceed the axis-orthogonality limit ($0.19331 > 0.15$), so contact physics is not evaluated. Later mirrored-hinge and microwave setups likewise do not reach a decision test.

The book/caddy study reaches a matched K but its first version has reset drift. The equilibrium-aligned version reverses in only one of two seeds; in the other, a book leaves the registered workspace. This condition cannot be excluded after outcomes are known. The hanging-mug/rack study reverses in only one of three release perturbations; in the other two, both model-5 candidates stay hanging and fixed-action regret is zero. These failures do not establish a robust contact-state

quotient. A later topology study fails setup admission before a complete candidate table exists and therefore supplies no positive evidence for a topology-dependent decision state.

Report-level preservation. The underlying reports retain their original decisions, including unsuccessful scientific tests and studies that ended before evaluation. These records are not independent experiments or a denominator for a success rate. The machine-readable audit preserves their identifiers, original decision strings, report paths, and hashes; the summary above describes their relevance to the present claim. A current digest alone does not demonstrate that every historical dependency remains reproducible.

E AUDIT INTERPRETATION

Numerical macros and table rows are generated from report fields and accompanied by full source hashes. Figure provenance records the same sources, and all plots are presentation-only transformations. Building the paper does not run simulations, fit models, or reevaluate experimental outcomes. The audit distinguishes nominal observations, unique combinations, clusters, and horizon repeats; none is automatically an independent sample count.

Historical bindings are verified against their recorded digest, not silently refreshed to the current file. The 16 directly used figure/result reports have no mismatched direct file bindings in the inspected workspace. The wider development chain contains historical binding discrepancies, documented individually in the audit; these are not repaired by replacing the expected hash. Formal scientific decisions remain frozen irrespective of the provenance audit. The local audit contains repository paths for author review and must be distinguished from an anonymized supplementary release.