Recoverability After Model–Reality Mismatch

A Same-State Counterfactual Evaluation Protocol

Seonhye Gu*
AI-Based Surgical Robot Innovation Lab

July 2026

Abstract

Physical AI systems must act when execution no longer matches the assumptions used for planning or control. Prior work studies predictive world models, mismatch detection, recovery policies, safety filters, replanning, and human handover. Yet these capabilities are commonly evaluated as separate systems or from different rollout histories, making it difficult to isolate the effect of the response itself.

We introduce a same-state counterfactual evaluation protocol that reconstructs a matched simulator execution state immediately after a task-relevant mismatch and varies only the post-mismatch response. In an Isaac Sim 4.1 and ORBIT-Surgical dual-arm reach proxy, a target is shifted by 6 cm at step 20 and an occlusion proxy is optionally applied. From the reconstructed state, we compare CONTINUE, REPLAN, REOBSERVE, RESHAPE, and HANDOVER. In the preregistered anchor condition with $n = 20$ seeds per branch, CONTINUE succeeded in 0/20 trials, REPLAN in 19/20, REOBSERVE in 17/20, and RESHAPE in 18/20. A visibility-threshold rule selected HANDOVER in all trials; because human completion was not implemented, terminal resolution after handover was not evaluated.

The evidence supports a narrow claim: recoverability at the tested state is intervention-conditioned, and mismatch detection alone does not determine the appropriate response. We do not claim a learned world model, a learned recoverability estimator, or clinical autonomy. The contribution is an evaluation protocol for measuring post-mismatch response profiles under controlled conditions.

Keywords: Physical AI, recoverability, mismatch, counterfactual evaluation, replanning, surgical robotics

1 Introduction

Predictive models allow embodied agents to anticipate future states and plan actions [1, 2]. Deployment nevertheless produces model–reality mismatch: targets move, visibility changes, contacts differ from expectation, dynamics shift, and human actions alter the task. Recent work addresses pieces of this problem through online mismatch estimation [3], safety filtering [4], replanning and specification repair [8, 9], learned recovery policies [5, 6], or human intervention [10, 11].

These methods answer different questions: Is the current state unsafe? Which parameter changed? Can a recovery controller return the system to a viable region? Should a human take over? Their evaluations also differ in task, onset state, and trajectory history. Consequently, when two responses produce different outcomes, it is often unclear how much of the difference is due to the response rather than the state from which it was applied.

We therefore ask:

> At the same post-mismatch execution state, do different intervention classes produce different recoverability outcomes?

We operationalize recoverability as the empirical success proportion of a response branch:

$$\hat{R}_a(S,t) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[y_i(S,a,t) = \text{success}], \tag{1}$$

where S is the matched post-mismatch execution state, a is the response, t is the intervention time, and N is the number of trials.

*This work was independently conducted by the author while affiliated with the AI-Based Surgical Robot Innovation Lab. The affiliation is provided for identification purposes only and does not imply official laboratory output or institutional endorsement.

The paper makes one contribution: a protocol that reconstructs a matched mismatch-onset state, forks multiple responses from it, and compares terminal outcomes under a common success criterion. The anchor experiment demonstrates the protocol in a controlled surgical-reach proxy; it is not presented as a clinical benchmark or a general law of recovery.

2 Related Work

2.1 Predictive Physical AI and Mismatch Estimation

Latent world models support policy learning and planning by predicting future representations rather than acting only reactively [1, 2]. Their usefulness depends on model adequacy. *World Models* explicitly showed that a controller can exploit inaccuracies in a learned simulator [1]. More recent work estimates mismatch online: Rapid Mismatch Estimation infers sudden changes in manipulated-object dynamics from proprioception and adapts within roughly 400 ms on a real manipulator [3]. Uncertainty-aware latent safety filters use world-model epistemic uncertainty to detect out-of-distribution hazards and synthesize safe actions [4].

These studies identify or mitigate model inadequacy. Our protocol begins after a mismatch state has been instantiated and asks how alternative responses compare from that state.

2.2 Recovery, Replanning, and Runtime Safety

Recovery RL separates a task policy from a recovery policy that intervenes near learned constraint-violation regions [5]. RecoveryChaining learns local recovery policies that return a failed manipulation system to states from which nominal model-based controllers can resume [6]. Reach-avoid safety filters and recovery controllers similarly switch or modify control to avoid unsafe regions and return to a target set [4, 7].

A complementary line repairs plans or specifications after assumption violations. Automated specification repair relaxes violated environment assumptions and identifies new robot skills needed for task completion [8]. VAP-TAMP combines active perception, situation assessment, and task-and-motion replanning for execution-time changes [9].

These works design a particular recovery or replanning mechanism. We instead compare a menu of response classes under a shared onset condition.

2.3 Human Intervention and Shared Autonomy

In surgical robotics, uncertainty has been used to trigger collaborative handoff. Agreement Volatility determines when a robot should cede control and when autonomy can resume during soft-tissue manipulation, reducing reliance on human intervention relative to a variance-only baseline [10]. Adaptive Intervention Mechanism similarly lets a robot request expert help during interactive imitation learning while reducing takeover cost [11]. Human-in-the-loop recovery is therefore not merely a terminal fallback; it is an arbitration problem involving uncertainty, recovery benefit, and human burden.

Our HANDOVER branch is intentionally incomplete: it records the selection of handover but does not model human completion. This prevents an unfair claim that automatic responses outperform human assistance.

2.4 Evaluation and Positioning

Simulation frameworks such as SIMPLER improve reproducibility and study sensitivity to distribution shifts through paired simulation and real-world policy evaluation [12]. Predictive red teaming identifies off-nominal conditions that degrade visuomotor policies and supports targeted data collection [13]. These works evaluate policy robustness across environments or perturbation conditions.

Our focus is narrower and orthogonal: *within one reconstructed post-mismatch state*, vary only the intervention. In the literature reviewed for this paper, we did not identify a standard protocol that jointly (i) reconstructs a matched onset state, (ii) forks a heterogeneous response menu, and (iii) reports response-conditioned terminal outcomes. We therefore frame the contribution as an evaluation protocol, not as a new detector, planner, safety filter, or recovery policy.

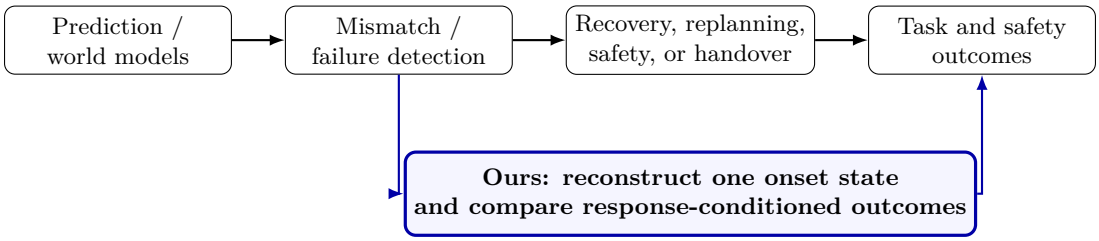


Figure 1: Positioning of the proposed protocol. Prior work typically develops components along the prediction-to-response pipeline. We hold the mismatch-onset state fixed and evaluate alternative responses.

Table 1: Experimental setup.

Component	Setting
Simulator	Isaac Sim 4.1.0
Framework	Isaac Lab 1.0.0, ORBIT-Surgical
Task	Isaac-Reach-Dual-STAR-IK-Rel-Play-v0
Controller	Scripted 6D IK-Rel
Episode horizon	160 steps
Mismatch onset	Step 20
Target shift	+6 cm along Y
Occlusion proxy	L1, visibility 0.35
Sample	Seeds 0–19 per branch, $n = 20$

3 Method

3.1 Environment and Anchor Condition

The controller is scripted inverse kinematics rather than a learned policy. This isolates response evaluation from policy-learning performance.

3.2 Same-State Counterfactual Fork

For each seed, we:

1. execute the nominal rollout through step 19 and record the actions;
2. apply the mismatch at step 20;
3. reset each response branch with the same seed;
4. replay the nominal actions to reconstruct state S;
5. apply only the branch-specific response; and
6. evaluate the terminal outcome.

All D0 and D1–D3 branches passed the predefined replay quality check. Here, "same-state" means a matched execution state reconstructed by the same seed and nominal action replay and verified by those checks. It does not claim byte-identical equality of all hidden simulator variables.

3.3 Response Classes

CONTINUE. Maintain the pre-mismatch target and continue execution.

REPLAN$_{d20}$. Hold for 20 steps, then update the scripted IK target using the shifted goal. This is a goal update in the scripted controller, not a general task-and-motion planner.

REOBSERVE. Hold, restore visibility, and then update the target.

RESHAPE. Apply an environment-reshaping proxy and then update the target.

Table 2: D0 response profile at the anchor state.

Response	Successes	Empirical rate
CONTINUE	0/20	0%
REPLAN_{d20}	19/20	95%
REOBSERVE	17/20	85%
RESHAPE	18/20	90%
HANDOVER	–	Not evaluable

Table 3: Rule comparison.

Comparator	Selected response	Automatic resolution
Best observed menu response	REPLAN_{d20}	19/20
UQ-inspired rule	HANDOVER in 20/20	Not evaluable
Situation rule	REOBSERVE	17/20

HANDOVER. Enter a human-assistance stub. Human completion is not implemented, so post-handover terminal resolution is not evaluated.

3.4 Baselines and Success Criterion

The UQ-inspired binary rule selects HANDOVER when the visibility or distance threshold fires. The situation rule selects REOBSERVE for an occluded shift and REPLAN for a shift without occlusion.

A branch is labeled a successful automatic resolution if:

1. the end-effector is within 0.02 m of the shifted target;
2. no forbidden axis-aligned bounding box is entered; and
3. the branch does not terminate in HANDOVER.

Occlusion is implemented as a controller-gain proxy rather than geometric or RGB occlusion.

4 Results

4.1 Anchor Response Profile

The response classes separate strongly at the tested state. CONTINUE never resolves the shifted target, while REPLAN_{d20}, REOBSERVE, and RESHAPE each provide successful paths in most trials.

4.2 Controls and Rule Comparisons

Without the occlusion proxy, CONTINUE succeeds in 1/20 trials and REPLAN_{d20} in 19/20. Within this narrow setup, the target mismatch appears more consequential than the gain-scale occlusion proxy. This does not support a general claim that occlusion is unimportant.

The UQ-inspired rule selects HANDOVER in 20/20 trials. Because human completion is absent, its terminal resolution cannot be compared directly with the automatic branches. The situation rule selects REOBSERVE and succeeds in 17/20 trials.

The best observed menu response is an oracle-style post hoc comparison, not a learned online selector.

5 Discussion

5.1 Detection Does Not Determine the Response

The binary rule detects the visibility condition but collapses all trials into HANDOVER. By contrast, multiple non-handover responses resolve the same matched state. This motivates separating mismatch detection,

$$m_t = D(\hat{s}_{t+1}, s_{t+1}, c_t), \tag{2}$$

from response selection,

$$a_t^R = \pi_R(m_t, s_t, g). \tag{3}$$

This study trains neither component. It only demonstrates why they should be evaluated separately.

5.2 Recoverability Is Intervention-Conditioned

The tested state is unrecoverable under CONTINUE but usually recoverable under REPLAN, REOBSERVE, or RESHAPE. A single binary label $R(S)$ therefore hides relevant structure. A more useful object is the response profile

$$\{R_{\text{continue}}(S), R_{\text{replan}}(S), R_{\text{reobserve}}(S), R_{\text{reshape}}(S), R_{\text{handover}}(S)\}. \tag{4}$$

This is the paper's main interpretation. It does not imply that REPLAN is generally optimal, nor that the ordering will persist across tasks, severities, or realistic surgical conditions.

6 Limitations

The evidence is limited to one reach proxy, one primary anchor condition, scripted IK control, and $n = 20$ trials per branch. Occlusion is a gain-scale proxy, HANDOVER lacks human completion, and best-of-menu is an oracle comparison. The simulator does not include tissue deformation, bleeding, smoke, anatomical variation, or clinical workflow. The results should therefore be read as a controlled protocol demonstration rather than evidence for clinical autonomy or general recoverability laws.

7 Conclusion

We introduced a same-state counterfactual protocol for comparing post-mismatch responses from a matched simulator execution state. In the anchor condition, CONTINUE succeeded in 0/20 trials, REPLAN_{d20} in 19/20, REOBSERVE in 17/20, and RESHAPE in 18/20.

The narrow conclusion is that recoverability at the tested state depends on the response, and that mismatch detection alone does not identify the appropriate intervention. Future work should test whether these response profiles persist across broader perturbation cells, replace proxy occlusion with geometric or visual occlusion, and evaluate learned response selection and completed human handover.

References

[1] Ha, D. and Schmidhuber, J. (2018). World Models. *arXiv preprint arXiv:1803.10122*.

[2] Assran, M., Bardes, A., Fan, D., Garrido, Q., Howes, R., et al. (2025). V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning. *arXiv preprint arXiv:2506.09985*.

[3] Jaszczuk, M. and Figueroa, N. (2025). Rapid Mismatch Estimation via Neural Network Informed Variational Inference. In *Proceedings of The 9th Conference on Robot Learning*, 5256–5277.

[4] Seo, J., Nakamura, K., and Bajcsy, A. (2025). Uncertainty-aware Latent Safety Filters for Avoiding Out-of-Distribution Failures. In *Proceedings of The 9th Conference on Robot Learning*, 4442–4472.

[5] Thananjeyan, B., Balakrishna, A., Nair, S., Luo, M., Srinivasan, K., Hwang, M., Gonzalez, J. E., Ibarz, J., Finn, C., and Goldberg, K. (2020). Recovery RL: Safe Reinforcement Learning with Learned Recovery Zones. *arXiv preprint arXiv:2010.15920*.

[6] Vats, S., Jha, D. K., Likhachev, M., Kroemer, O., and Romeres, D. (2024). RecoveryChaining: Learning Local Recovery Policies for Robust Manipulation. *arXiv preprint arXiv:2410.13979*.

[7] Begzadic, A., Shinde, N., Tonkens, S., Hirsch, D., Ugalde, K., Yip, M., Cortes, J., and Herbert, S. (2025). Back to Base: Towards Hands-Off Learning via Safe Resets with Reach-Avoid Safety Filters. In *Proceedings of the 7th Annual Learning for Dynamics and Control Conference*, 1154–1166.

[8] Meng, Q. and Kress-Gazit, H. (2024). Automated Robot Recovery from Assumption Violations of High-Level Specifications. *arXiv preprint arXiv:2407.00562*.

[9] Oloo, A., et al. (2026). Robot Planning and Situation Handling with Active Perception. *arXiv preprint arXiv:2604.26988*.

[10] Thompson, J., Jordan, B., Brown, D. S., and Kuntz, A. (2025). Agreement Volatility: A Second-Order Metric for Uncertainty Quantification in Surgical Robot Learning. In *Proceedings of The 9th Conference on Robot Learning*, 3685–3703.

[11] Cai, H., Peng, Z., and Zhou, B. (2025). Robot-Gated Interactive Imitation Learning with Adaptive Intervention Mechanism. In *Proceedings of the 42nd International Conference on Machine Learning*, 6243–6256.

[12] Li, X., Hsu, K., Gu, J., Mees, O., Pertsch, K., Walke, H. R., et al. (2025). Evaluating Real-World Robot Manipulation Policies in Simulation. In *Proceedings of The 8th Conference on Robot Learning*, 3705–3728.

[13] Majumdar, A., Sharma, M., Kalashnikov, D., Singh, S., Sermanet, P., and Sindhwani, V. (2025). Predictive Red Teaming: Breaking Policies Without Breaking Robots. In *Proceedings of The 9th Conference on Robot Learning*, 1615–1635.

[14] Liang, X., Wang, C.-P., Shinde, N. U., Liu, F., Richter, F., and Yip, M. (2025). MEDiC: Autonomous Surgical Robotic Assistance to Maximizing Exposure for Dissection and Cautery. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 6889–6895.