Q-PHYSJEPA: QUERY-CONDITIONED PHYSICAL ABSTRACTION FOR JEPA WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Embodied world models should preserve the physical evidence needed for the task at hand, yet accurate latent prediction and physical-property decoding do not by themselves establish decision relevance. We introduce Q-PhysJEPA, a query-conditioned joint-embedding predictive architecture that introduces task queries before low-dimensional history compression. Visual, proprioceptive, and action histories are aggregated into evidence slots, reweighted by a query-conditioned gate, and compressed into a task-dependent representation. The proposed model predicts candidate execution futures and estimates utilities from their mean and terminal embeddings, combining future prediction, utility regression, and candidate ranking. We evaluate the compression mechanism in a component study with a shared direct candidate-readout architecture, alongside separate physical-content and intervention diagnostics. On ManiSkill tests with physical parameters outside the training ranges, this direct-readout variant achieves 56.29% top-1 candidate-selection accuracy across four objectives, compared with 46.83% for a query-independent representation, a gain of 9.46 percentage points. These measurements concern query-conditioned compression, not the complete future-mediated scoring path. Controlled analyses further show that greater query-dependent physical selectivity need not imply better decisions. The results motivate task-conditioned history compression while distinguishing accessible physical information from its demonstrated use in candidate selection. Codes are available at .

1 INTRODUCTION

World models support embodied decision-making by using interaction histories to predict how the environment may evolve under possible executions (Hafner et al., 2019; 2020). JEPA-based world models perform this prediction in a low-dimensional latent space. The resulting representations are learned through predictive objectives, while the physical information they encode can be examined through property readouts and downstream control (Maes et al., 2026). However, neither low prediction error nor successful property decoding is sufficient to establish that a representation is adequate for embodied decision-making, i.e., given a task objective and a set of candidate executions, a representation need not preserve physical differences that leave the optimal candidate unchanged; it must instead preserve the decision-relevant distinctions that can change that choice (Grimm et al., 2020). Aggregate prediction error does not directly measure this requirement, because a low average may conceal errors on distinctions that determine the preferred candidate (Farahmand et al., 2017). Property decodability is likewise insufficient: it shows that the probed properties are accessible from the representation, but not that the model uses them in making its decision (Ravichander et al., 2021).

We argue that a low-dimensional history representation should be conditioned on the task query, since different queries may prioritize different aspects of possible outcomes, thereby changing which physical evidence from the same interaction history is relevant to candidate selection. As a result, a physical distinction that leaves the preferred candidate unchanged under one query may determine the choice under another. For example, the task queries, such as efficient pushing, safe lifting, and precise placement, require different evidence from the same interaction history: friction and contact dynamics for pushing, object mass and grasp stability for lifting, and geometry and contact alignment for placement. A single query-agnostic representation shared across objectives may therefore retain distinctions useful for some queries while discarding those required by others. We call such

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.

task-dependent compression of interaction history *query-conditioned physical abstraction*, as shown in Figure 1.

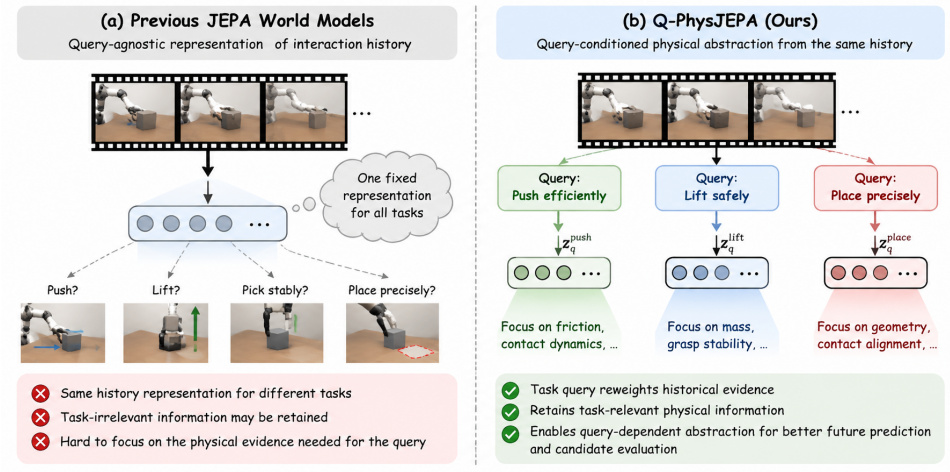

Figure 1: Query-agnostic and query-conditioned compression of interaction history. A query-agnostic bottleneck maps the same history to a representation shared across objectives. Q-PhysJEPA instead uses the task query to reweight historical evidence before the final low-dimensional projection, allowing the same history to yield task-specific abstractions.

However, learning and validating such an abstraction poses two challenges. First, how can a low-dimensional representation preserve the physical distinctions needed for the current query? Second, how can training make these distinctions useful for candidate evaluation, and how can we test whether the model actually uses them? Changing the query may alter a representation or decision even when the model ignores interaction evidence. Query sensitivity alone therefore does not establish physical grounding.

To address the first challenge, Q-PhysJEPA uses the query to gate historical evidence before the final low-dimensional projection, rather than after potentially relevant distinctions have been discarded. To address the second, we couple latent future prediction with query-dependent decision supervision. Predicted candidate futures are trained against embeddings of realized future observations; utility regression and candidate ranking use task-defined utilities from recorded executions. The utility head reads predicted futures rather than raw histories or controller parameters, so decision-loss gradients pass through the predictor into the query-conditioned abstraction. This coupling encourages the representation to retain evidence that predicts physical consequences and distinguishes task-relevant outcomes. It is a learning constraint, not a guarantee of physical identifiability. We therefore distinguish retention from functional use through property readouts and matched query, physical, and temporal interventions, assessing representation, future-prediction, and decision responses separately.

We evaluate query-conditioned compression on ManiSkill using the same direct candidate-readout architecture for both representations. With physical parameters outside the training ranges, this component study achieves 56.29% macro-averaged top-1 accuracy across four objectives, versus 46.83% for a query-independent representation, a gain of 9.46 percentage points. Separate diagnostics examine physical content and decision sensitivity; Appendix A.4 distinguishes these measurements from the complete future-mediated model.

2 Related Work

2.1 JEPAs and Latent World Models

I-JEPA and V-JEPA learn by predicting image and video embeddings (Assran et al., 2023; Bardes et al., 2024). V-JEPA 2-AC and LeWorldModel extend latent prediction to control (Assran et al.,

2025; Maes et al., 2026). DINO-WM plans with predicted pretrained visual features (Zhou et al., 2024), while TD-MPC jointly learns task-oriented dynamics and terminal values (Hansen et al., 2022). LeJEPA addresses embedding collapse (Balestriero & LeCun, 2025); PhyLatent further targets physical-state and counterfactual distinctions (Zeng et al., 2026). Our question is which physical distinctions a compressed history should preserve when the task query changes.

2.2 Task-Relevant Abstraction and Conditioning

The information bottleneck formalizes compression that preserves relevant information (Tishby et al., 2000). In control, DeepMDP and bisimulation-based representations connect abstraction to rewards and transitions (Gelada et al., 2019; Zhang et al., 2021), while value-aware modeling prioritizes decision-relevant prediction errors (Farahmand et al., 2017; Grimm et al., 2020). TIA separates reward-related features from distractors (Fu et al., 2021); TaskSense directs spatial attention before latent encoding (Islam & Huber, 2026). Conditioning itself is established: FiLM modulates features (Perez et al., 2018), and TC-JEPA conditions patch prediction on text (Huang et al., 2026). Thorpe et al. (2026) explicitly advocate query-sufficient physical abstractions. We study a specific learned instantiation: query-gated interaction slots before compression, coupled to candidate-future and utility learning. We claim neither a new principle of conditioning nor information-theoretic sufficiency guarantees.

2.3 Physical Evidence and Functional Use

HPT learns visual–proprioceptive representations (Wang et al., 2024), and DensePhysNet recovers mass and friction through interactions (Xu et al., 2019). Video-model studies examine intuitive physics, motion readouts, and dynamic-property recovery (Garrido et al., 2025; Joseph et al., 2026; Zhan et al., 2025). Work on physical identifiability distinguishes recoverability, retention, and predictive use (Tan et al., 2026; Zhang et al., 2026). In language models, probing controls and amnesic interventions likewise separate encoded information from behavioral use (Ravichander et al., 2021; Elazar et al., 2021). Our diagnostics ask whether query-relative physical information also supports candidate decisions; readout accuracy or query sensitivity alone cannot answer that question.

3 Method

3.1 Problem Formulation

We consider an embodied decision problem in which a learned world model selects one parameter setting from a finite set of candidates for a predefined feedback controller. Formally, at the decision point, the world model receives an interaction history

$$\mathcal{H}_T = (o_0, a_0, o_1, \ldots, a_{T-1}, o_T), \tag{1}$$

where $o_t = (I_t, p_t)$ contains an RGB image I_t and the robot proprioceptive state p_t, and a_t is the low-level action applied between o_t and o_{t+1}.

The world model additionally receives a task query $q \in \mathbb{R}^{D_q}$ and a candidate set $\mathcal{C} = \{c_i\}_{i=1}^{N}$. The query encodes the current task objective and determines how execution outcomes are evaluated, while each candidate c_i specifies a parameter setting for the predefined feedback controller $\pi(\cdot; c_i)$. Instantiating the controller with c_i determines its closed-loop behavior during execution.

Executing the controller $\pi(\cdot; c_i)$ from the environment state reached after $\mathcal{H}_T$ would produce an observation rollout

$$\mathcal{O}_i = \left(o_0^{(i)}, \ldots, o_{L_i-1}^{(i)}\right), \tag{2}$$

where L_i is the realized execution length and $o_0^{(i)}$ is the initial observation at the post-history state. Let $\mathcal{U}_q$ denote the task-defined evaluator that assigns a scalar utility to a realized rollout, with larger values indicating more desirable outcomes. The utility of candidate c_i and the oracle-optimal candidate index are defined as

$$u_i^{(q)} = \mathcal{U}_q(\mathcal{O}_i), \qquad i^\star(q) = \arg\max_{i \in \{1,\ldots,N\}} u_i^{(q)}. \tag{3}$$

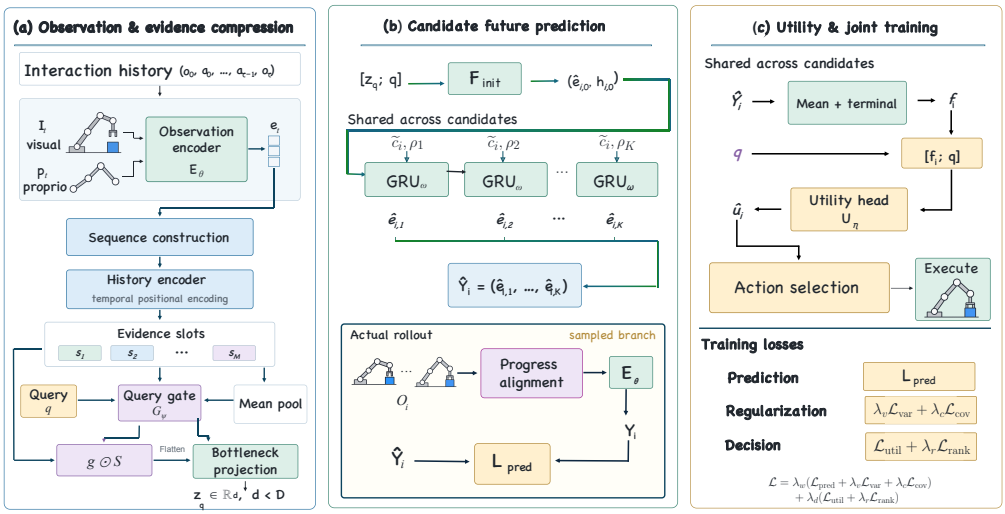

Figure 2: The future-mediated Q-PhysJEPA architecture. Interaction history is encoded into evidence slots S and compressed into a low-dimensional representation z_q via query-conditioned gating. Given $[z_q; q]$, candidate parameters c_i, and execution phases $\boldsymbol{\rho}$, the predictor generates future representations $\widehat{Y}_i$, which a query-conditioned utility head scores for candidate selection. During training, the shared observation encoder maps real future observations to prediction targets.

Note that although candidate executions may produce rollouts of different lengths, $\mathcal{U}_q$ maps them to comparable scalar utilities. Consequently, $i^\star(q)$ identifies the candidate preferred under query q, and $c_{i^\star(q)}$ is the corresponding controller parameter setting.

Our goal is therefore to identify the preferred candidate $i^\star(q)$ from $(\mathcal{H}_T, q, \mathcal{C})$ without first executing the candidate controllers by using the world model. Realized candidate rollouts and their utilities provide supervision during training, but are unavailable at decision time. The model must therefore infer the consequences of each candidate from the observed interaction history and its controller setting, and determine which candidate is preferred under the given query.

3.2 Q-PhysJEPA Overview

JEPA provides a natural foundation for the formulated problem by learning compact representations through latent-space prediction. Its predictive objective, however, is not necessarily aligned with the decision problem defined in Eq. 3. A standard JEPA-style model learns a latent state z that is useful for predicting future representations, typically by minimizing prediction error over the future observations encountered during training. Such an objective encourages z to preserve information that is predictive on average, but does not specifically require it to retain the physical distinctions that determine which candidate is preferred under a particular task query q.

A natural extension is to condition future prediction or outcome evaluation on q. This alone is insufficient when the interaction history has already been compressed into a query-agnostic latent representation. Once information has been removed in forming the compressed latent vector z, downstream conditioning cannot recover a physical distinction that was discarded before the query was introduced. Consequently, even a query-conditioned predictor may lack the query-relevant physical evidence required to distinguish candidate outcomes.

Q-PhysJEPA addresses these limitations by making both representation formation and learning supervision decision-aware. First, we introduce a *query-conditioned physical abstraction*. The model organizes the interaction history into a set of evidence summaries S and uses the task query to reweight this evidence before the final low-dimensional bottleneck, producing z_q, which is used for prediction. The query therefore influences which physical distinctions survive compression and the

prediction process, rather than being applied only after a fixed latent state has been formed. Second, we ground this abstraction through candidate-conditioned latent future prediction and query-dependent decision supervision. For each candidate controller setting c_i, the model predicts a latent execution future $\widehat{Y}_i$ that is supervised by the corresponding realized rollout. The predicted future is then evaluated under the task query to obtain $\widehat{u}_i^{(q)}$. Utility regression and candidate-ranking losses further emphasize distinctions in predicted outcomes that change candidate preference. Together, these objectives encourage z_q to preserve physical evidence that is both predictive of execution consequences and relevant to the current decision.

Fig. 2 illustrates the general pipeline of Q-PhysJEPA, and the following sections describe the core components including the query-conditioned history compression, candidate-conditioned latent future prediction, and joint learning objectives.

3.3 Query-Conditioned History Compression

Based on the interaction history, Q-PhysJEPA constructs the query-conditioned representation z_q through two stages: interaction evidence extraction, which organizes the observed interaction into a set of evidence slots, and query-conditioned physical abstraction, which selectively compresses these slots according to the task query.

Interaction Evidence Extraction. For each observation $o_t = (I_t, p_t)$, the observation encoder combines visual and proprioceptive information into a fused representation e_t:

$$\begin{aligned} \widetilde{p}_t &= (p_t - \mu_p) \oslash \sigma_p, \\ e_t &= E_\theta(o_t) = F_\theta([V_\theta(I_t); \widetilde{p}_t]) \in \mathbb{R}^D, \end{aligned} \tag{4}$$

where V_θ is the visual encoder, F_θ is a fusion MLP, μ_p and σ_p are training-set normalization statistics, and $\oslash$ denotes element-wise division.

Physical evidence is reflected in both individual observations and the environment's responses to actions. We therefore pair each observation representation with its preceding low-level action:

$$x_t = [e_t; \widetilde{a}_{t-1}; v_t] \in \mathbb{R}^{D+D_a+1}, \tag{5}$$

where

$$\widetilde{a}_{t-1} = \begin{cases} \mathbf{0}_{D_a}, & t = 0, \\ (a_{t-1} - \mu_a) \oslash \sigma_a, & 1 \le t \le T, \end{cases} \qquad v_t = \begin{cases} 0, & t = 0, \\ 1, & 1 \le t \le T. \end{cases} \tag{6}$$

Here, μ_a and σ_a are training-set normalization statistics and v_t distinguishes the missing initial action from a genuine zero-valued action.

The resulting interaction tokens are contextualized by a temporal Transformer (Vaswani et al., 2017):

$$R = \mathrm{Transformer}_\phi\Big(\{\mathrm{LN}(W_x x_t + b_x) + \pi_t\}_{t=0}^{T}\Big) \in \mathbb{R}^{(T+1)\times D}, \tag{7}$$

where π_t denotes the temporal positional encoding. Rather than immediately compressing R into a single latent vector, we use M learnable aggregation queries to extract a set of interaction evidence slots:

$$S = \mathrm{LN}(Q_{\mathrm{slot}} + \mathrm{MHA}(Q_{\mathrm{slot}}, R, R)) \in \mathbb{R}^{M\times D}. \tag{8}$$

Here, $Q_{\mathrm{slot}} \in \mathbb{R}^{M\times D}$ contains learned aggregation queries shared across examples and is distinct from the task query q. The resulting slots $S = (s_1, \ldots, s_M)$ summarize complementary evidence from the interaction history without imposing predefined physical semantics on individual slots.

Query-Conditioned Physical Abstraction. Given the interaction evidence S, the task query determines how strongly each slot contributes to the final compressed representation. We compute a query-dependent gate as

$$\overline{s} = \frac{1}{M}\sum_{m=1}^{M} s_m, \qquad g_q = \mathrm{sigmoid}(G_\psi([\overline{s}; q])) \in (0,1)^M. \tag{9}$$

The reweighted slots are then concatenated and projected into the low-dimensional abstraction:

$$r_q = [g_{q,1}s_1; \ldots; g_{q,M}s_M] \in \mathbb{R}^{MD},$$
$$z_q = \tanh(\mathrm{LN}(W_b r_q + b_b)) \in \mathbb{R}^d, \qquad d < D. \tag{10}$$

The continuous gates allow the same interaction evidence to be weighted differently across task queries without assigning fixed semantics to the individual slots or making a hard selection. As a result, z_q can preserve different physical distinctions from the same interaction history depending on the current objective. Consistent with the JEPA predictive setup, the downstream future predictor operates on the compact representation z_q and does not directly access R, S, or e_T.

3.4 Candidate-Conditioned Latent Future Prediction and Utility Estimation

Given the query-conditioned abstraction z_q, Q-PhysJEPA compares candidate controller settings through their predicted execution consequences. Rather than scoring c_i directly, the model predicts how the same inferred post-history condition would evolve under each candidate and evaluates the resulting latent future under the task query q.

Candidate-Conditioned Latent Rollout Prediction. For each $c_i \in \mathcal{C}$, the predictor rolls z_q forward into a candidate-conditioned latent future

$$\widehat{Y}_i = P_\omega(z_q, q, c_i, \boldsymbol{\rho}) = (\widehat{e}_{i,1}, \ldots, \widehat{e}_{i,K}), \qquad \widehat{e}_{i,k} \in \mathbb{R}^D, \tag{11}$$

where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_K)$ specifies K normalized execution phases with $0 < \rho_1 < \cdots < \rho_K = 1$. Using normalized phases provides a common prediction coordinate for candidate executions with different realized lengths. Following the JEPA principle of predicting target representations in latent space, training targets are obtained by aligning each recorded rollout $\mathcal{O}_i$ with these phases and encoding the corresponding observations with E_θ. Specifically, phase ρ_k is matched to $o^{(i)}_{\lceil \rho_k (L_i - 1) \rceil}$, whose encoding provides the target $y_{i,k}$. The predictor P_ω is implemented recurrently with a GRU network (Cho et al., 2014). All candidate branches share an initialization derived from $[z_q; q]$ and are subsequently rolled forward with their respective normalized controller parameters $\widetilde{c}_i$. Thus, each candidate is evaluated from the same inferred post-history condition, while its controller setting determines the branch-specific latent evolution.

Utility Estimation from Predicted Futures. We evaluate each candidate with a query-conditioned process–terminal utility head. Its input combines the predicted execution and final outcome:

$$f_i = \left[\frac{1}{K} \sum_{k=1}^{K} \widehat{e}_{i,k}; \widehat{e}_{i,K} \right], \qquad \widehat{u}_i^{(q)} = U_\eta([f_i; q]). \tag{12}$$

where the mean-pooled component summarizes execution, while the terminal component explicitly preserves the predicted final outcome. The shared head U_η has two 128-unit GELU hidden layers and a scalar output. It receives only $[f_i; q]$, not raw history, controller parameters, or realized execution length. Thus, candidate-dependent information reaches the score through predicted consequences. At decision time, we select $\widehat{i}(q) = \arg\max_i \widehat{u}_i^{(q)}$.

3.5 Joint Learning Objective

Q-PhysJEPA is trained with two complementary forms of supervision: the latent future prediction grounds candidate-conditioned predictions in realized execution outcomes, while utility-based supervision identifies which differences between these outcomes matter for candidate selection under the current query. Together, they encourage the query-conditioned abstraction z_q to retain physical evidence that is both predictive and decision-relevant.

Latent Future Grounding. Let $\mathcal{B}$ denote a minibatch. For each example $b \in \mathcal{B}$, one recorded candidate branch i_b provides latent future targets and the latent prediction objective is:

$$\mathcal{L}_{\mathrm{pred}} = \frac{1}{D|\mathcal{V}|} \sum_{(b,k) \in \mathcal{V}} \|\widehat{e}_{b,k} - y_{b,k}\|_2^2, \tag{13}$$

where D is the observation-embedding dimension in Eq. 4. We write $\widehat{e}_{b,k} := \widehat{e}_{b,i_b,k}$ and $y_{b,k} := y_{b,i_b,k}$, and define $\mathcal{V}$ as the set of valid (b,k) target positions in the minibatch for concise presentation. This objective anchors the predicted candidate future to the latent representation of its realized execution. Both the history and future observations are encoded in the same latent space by E_θ. We additionally apply variance and covariance regularization, denoted by $\mathcal{L}_{\text{var}}$ and $\mathcal{L}_{\text{cov}}$, to discourage representation collapse and excessive feature redundancy (Bardes et al., 2022).

Query-Dependent Utility Learning. Accurately predicting a future does not by itself indicate which aspects of that future are important for the current decision. Two candidate executions may be equally predictable while differing substantially in their utility under query q. We therefore directly supervise the predicted utility $\widehat{u}_{b,i}^{(q_b)}$ using the task-defined utility $u_{b,i}^{(q_b)}$ from Eq. 3.

Let $\mathcal{A}$ denote the set of candidate executions in the minibatch for which ground-truth utility is available. We minimize

$$\mathcal{L}_{\text{util}} = \frac{1}{|\mathcal{A}|} \sum_{(b,i)\in\mathcal{A}} \ell_\beta\Big(\widehat{u}_{b,i}^{(q_b)} - u_{b,i}^{(q_b)}\Big), \tag{14}$$

where ℓ_β denotes the Smooth L1 loss with transition parameter β. Since $\widehat{u}_{b,i}^{(q_b)}$ is estimated from the predicted future and $u_{b,i}^{(q_b)}$ is calculated considering the query, this supervision encourages the latent prediction pathway to preserve outcome differences that affect utility under the current query.

The final decision also depends on candidate utility ordering. We therefore complement utility regression with pairwise ranking supervision (Burges et al., 2005). Let $\mathcal{P}$ contain valid candidate pairs (b,i,j) from the same example whose ground-truth utilities differ by more than τ. The ranking objective is

$$\begin{aligned}\mathcal{L}_{\text{rank}} = \frac{1}{|\mathcal{P}|} \sum_{(b,i,j)\in\mathcal{P}} \text{softplus}\Big(&- \text{sgn}\Big(u_{b,i}^{(q_b)} - u_{b,j}^{(q_b)}\Big) \\ &\cdot \Big(\widehat{u}_{b,i}^{(q_b)} - \widehat{u}_{b,j}^{(q_b)}\Big)\Big),\end{aligned} \tag{15}$$

where $\text{sgn}(\cdot)$ is the sign function and $\text{softplus}(x) = \log(1+e^x)$. The threshold τ excludes near-ties with ambiguous ordering. We set $\mathcal{L}_{\text{rank}} = 0$ if no valid pair exists.

The complete objective combines physical grounding and decision supervision:

$$\begin{aligned}\mathcal{L} = \lambda_w \left(\mathcal{L}_{\text{pred}} + \lambda_v \mathcal{L}_{\text{var}} + \lambda_c \mathcal{L}_{\text{cov}}\right) \\ + \lambda_d \left(\mathcal{L}_{\text{util}} + \lambda_r \mathcal{L}_{\text{rank}}\right),\end{aligned} \tag{16}$$

where λ_w and λ_d balance the two forms of supervision, and λ_v, λ_c, and λ_r weight their corresponding regularization and ranking terms.

The supervision terms serve distinct roles: $\mathcal{L}_{\text{pred}}$ anchors latent predictions to observed execution consequences, while $\mathcal{L}_{\text{util}}$ and $\mathcal{L}_{\text{rank}}$ emphasize outcome differences that affect query-dependent utilities and candidate preferences. Because U_η reads only predicted futures and q, decision gradients reach the predictor and query-conditioned abstraction. Variance and covariance regularization discourage degeneracy; they do not assign physical meanings to slots. These objectives encourage physical grounding, whose retention and functional use are evaluated rather than assumed.

4 Experiments

We evaluate query-conditioned compression with a direct candidate readout (DR) on visual ManiSkill histories. Separate structured and pixel diagnostics examine physical-evidence availability, retention, and decision relevance; their scores are not pooled.

4.1 Experimental Setup

We use ManiSkill3 `PushCube-v1` and `PickCube-v1` (Tao et al., 2025). Push objectives are precise positioning and efficient transport; Pick objectives are stable holding and rapid completion. Mass and table or gripper friction vary but are not model inputs. Each environment has 384/96/192/192 train/validation/ID/OOD episodes. Source histories, candidate branches, and query

views share a split; OOD physical parameters lie outside training ranges. Each candidate parameterizes a predefined feedback controller. The primary metric is top-1 agreement with the highest-utility recorded candidate, averaged equally across four objectives. Structured diagnostics add normalized regret, non-tied pairwise ranking accuracy, and property readouts. Appendix A details protocols and implementation.

4.2 Candidate Selection with a Direct Readout

Table 1: OOD candidate selection on ManiSkill. Top-1 accuracy is the macro-average over four objectives, not closed-loop task success. The difference is in percentage points (pp)

Method	Top-1 accuracy ↑	Difference
No-query (DR)	46.83%	—
Q-PhysJEPA	**56.29%**	+9.46 pp

Q-PhysJEPA improves the reported OOD aggregate by 9.46 percentage points (Table 1). This supports query-conditioned compression for selecting among the tested controller settings, but does not establish an improvement in closed-loop success or on every objective. Task-level counts and seed-level uncertainty are still needed to resolve that scope (Appendix C.1). Frozen readouts provide a separate predictive diagnostic (Appendix B.1): Q-PhysJEPA has the lowest error on five of six reported metrics. Its state heads received additional tuning not matched for the baselines, whereas its action readouts are unchanged. We therefore do not treat the state-readout comparison as a matched-budget test or as an explanation of the decision gain.

4.3 Physical Content and Decision Quality (RQ1, RQ3, RQ5)

These single-seed structured diagnostics describe the relation between query-relative physical content and candidate-selection quality.

Table 2: Physical content and decision quality in the single-seed structured-observation diagnostic. Selectivity is relevant-property R^2 minus nonnegative irrelevant-property R^2 (leakage); ranking is pairwise accuracy, not top-1 accuracy or task success.

	Physical content			Decision quality	
Method	Relevant R^2 ↑	Leakage R^2 ↓	Selectivity ↑	Regret ↓	Ranking ↑
Representation width 8					
Strong late fusion	0.6292	0.4628	0.1664	0.2510	81.84%
Q-PhysJEPA	**0.7001**	**0.2478**	**0.4523**	**0.1680**	**84.70%**
Representation width 16					
Strong late fusion	0.7139	0.6741	0.0398	0.1915	83.11%
Q-PhysJEPA	**0.8368**	**0.6422**	**0.1946**	**0.1500**	**85.42%**

Query-relative content (RQ1). At both latent widths, Q-PhysJEPA increases relevant-property R^2, reduces irrelevant-property leakage, and improves regret and ranking relative to strong late fusion (Table 2). At width 8, selectivity rises from 0.1664 to 0.4523 while regret falls from 0.2510 to 0.1680.

Candidate selection (RQ3). Across plain JEPA, late fusion, Q-Predictor, Q-Gate, and Q-PhysJEPA, Q-PhysJEPA has the lowest reported regret at both widths: 0.1680 and 0.1500 (Figure 3(a)). This supports the tested conditioning combination, but does not establish causal use of individual decoded properties.

Selectivity alone does not determine decision quality (RQ5). Increasing Q-PhysJEPA's width from 8 to 16 lowers selectivity from 0.4523 to 0.1946, yet improves relevant-property R^2 from 0.7001 to 0.8368 and regret to 0.1500. Late fusion shows the same direction. Thus, suppressing

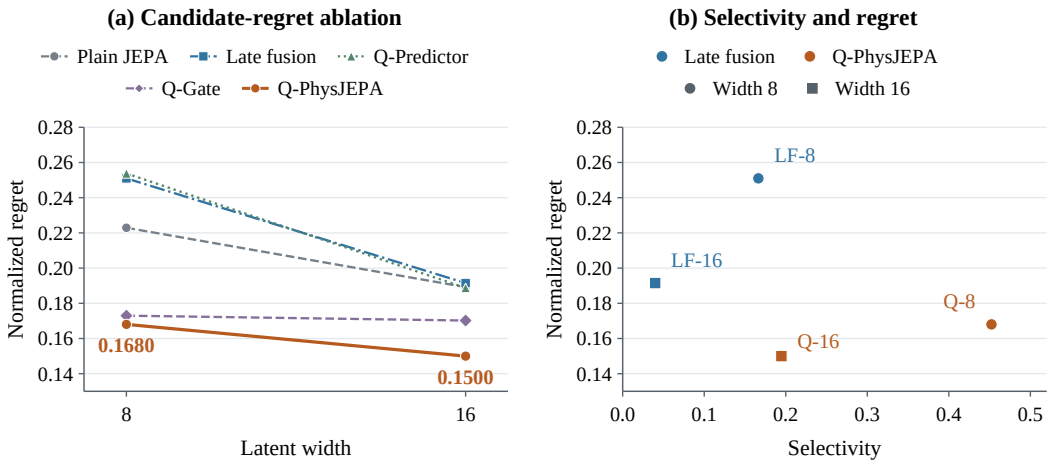

Figure 3: Structured-observation diagnostics (single seed). (a) Candidate regret at latent widths 8 and 16. (b) Selectivity versus regret for late fusion (LF) and Q-PhysJEPA (Q); point labels indicate latent width. Lower regret is better.

irrelevant information alone does not ensure decision quality. Figure 3(b) replots these four conditions; Appendix B.5 discusses their interpretation.

4.4 Available Physical Evidence and Grounding (RQ2, RQ4)

Evidence availability (RQ2). Horizontal pushing identifies friction with accuracy 0.990; microlifting identifies mass with accuracy 1.000. The other property remains at four-class chance (0.250). In a separate pixel diagnostic, friction R^2 is 0.492 with one frame, 0.981 with unordered frames, and 0.994 with the multi-frame estimator (Appendix B.2). These results establish evidence availability, not its retention or use by Q-PhysJEPA.

Grounding controls (RQ4). No-query, query-only, and Q-PhysJEPA remain close on precise pushing and stable picking, below the oracle; balancing substantially lowers accuracy (Appendix B.4). The four-objective gain is not uniform, and query sensitivity does not establish grounding. Matched structured interventions show stronger relative sensitivity to relevant physics; belief interventions affect a separate analytic planner. Temporal masking does not satisfy both the interval criterion and the preregistered 0.02 effect threshold (Appendix C.2). With the current observation fixed, RGB-window masking changes no selections across 36,864 intervention records. These tests do not establish a complete link from physical evidence to predicted futures and decisions.

5 Conclusions

Q-PhysJEPA combines query-conditioned history compression with candidate future prediction and utility learning. The task query guides which information is retained from the interaction history, and the resulting representation is used to predict and evaluate candidate outcomes. A separate direct-readout study supports the value of query-conditioned compression, while structured diagnostics associate query conditioning with more selective physical readouts and lower candidate-selection regret. These studies provide evidence about compression and representation quality but do not measure the complete future-mediated model. The capacity comparisons also show that greater selectivity alone does not ensure better decisions. A useful representation must preserve accurate task-relevant distinctions, even when doing so retains some irrelevant information. The grounding controls leave open whether interaction evidence reliably shapes predicted futures and subsequent choices. Resolving this question requires testing how changes in physical evidence affect representations, predictions, and decisions within the same model.

AI ASSISTANCE STATEMENT

Generative AI tools assisted with manuscript drafting and language editing, literature discovery and organization, code development and refactoring, figure preparation, and the preparation of experimental analyses. These tools were research aids, not components of the evaluated models. AI-generated suggestions are not independent evidence for the reported findings. The authors are responsible for checking AI-assisted text, figures, equations, citations, code, and analyses against the underlying sources and recorded outputs, and take full responsibility for the manuscript and its conclusions.

REFERENCES

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture, 2023. URL https://arxiv.org/abs/2301.08243.

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Mojtaba Komeili, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, Sergio Arnaud, Abha Gejji, Ada Martin, Francois Robert Hogan, Daniel Dugas, Piotr Bojanowski, Vasil Khalidov, Patrick Labatut, Francisco Massa, Marc Szafraniec, Kapil Krishnakumar, Yong Li, Xiaodong Ma, Sarath Chandar, Franziska Meier, Yann LeCun, Michael Rabbat, and Nicolas Ballas. V-JEPA 2: Self-supervised video models enable understanding, prediction and planning, 2025. URL https://arxiv.org/abs/2506.09985.

Randall Balestriero and Yann LeCun. LeJEPA: Provable and Scalable Self-Supervised Learning Without the Heuristics. *arXiv preprint arXiv:2511.08544*, 2025. URL https://arxiv.org/abs/2511.08544.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2105.04906.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mahmoud Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video, 2024. URL https://arxiv.org/abs/2404.08471.

Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning*, pp. 89–96. ACM, 2005. doi: 10.1145/1102351.1102363. URL https://dl.acm.org/doi/10.1145/1102351.1102363.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734. Association for Computational Linguistics, 2014. doi: 10.3115/v1/D14-1179. URL https://aclanthology.org/D14-1179/.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2010.11929.

Yanai Elazar, Shauli Ravfogel, Alon Jacovi, and Yoav Goldberg. Amnesic probing: Behavioral explanation with amnesic counterfactuals. *Transactions of the Association for Computational Linguistics*, 9, 2021. doi: 10.1162/tacl_a_00359. URL https://arxiv.org/abs/2006.00995.

Amir-Massoud Farahmand, Andre Barreto, and Daniel Nikovski. Value-aware loss function for model-based reinforcement learning. In *Proceedings of the 20th International Conference*

on Artificial Intelligence and Statistics, volume 54 of *Proceedings of Machine Learning Research*, pp. 1486–1494. PMLR, 2017. URL https://proceedings.mlr.press/v54/farahmand17a.html.

Xiang Fu, Ge Yang, Pulkit Agrawal, and Tommi Jaakkola. Learning task informed abstractions. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3480–3491. PMLR, 2021. URL https://proceedings.mlr.press/v139/fu21b.html.

Quentin Garrido, Nicolas Ballas, Mahmoud Assran, Adrien Bardes, Laurent Najman, Michael Rabbat, Emmanuel Dupoux, and Yann LeCun. Intuitive Physics Understanding Emerges from Self-Supervised Pretraining on Natural Videos. *arXiv preprint arXiv:2502.11831*, 2025. URL https://arxiv.org/abs/2502.11831.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G. Bellemare. DeepMDP: Learning continuous latent space models for representation learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2170–2179. PMLR, 2019. URL https://proceedings.mlr.press/v97/gelada19a.html.

Christopher Grimm, Andre Barreto, Satinder P. Singh, and David Silver. The value equivalence principle for model-based reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/3bb585ea00014b0e3ebe4c6dd165a358-Abstract.html.

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

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020. URL https://arxiv.org/abs/1912.01603.

Nicklas A. Hansen, Hao Su, and Xiaolong Wang. Temporal difference learning for model predictive control. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8387–8406. PMLR, 2022. URL https://proceedings.mlr.press/v162/hansen22a.html.

Chen Huang, Xianhang Li, Vimal Thilak, Etai Littwin, and Josh Susskind. Text-Conditional JEPA for Learning Semantically Rich Visual Representations. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2605.03245.

SM Mazharul Islam and Manfred Huber. TaskSense: Focusing on What Matters in World Models. *arXiv preprint arXiv:2608.06544*, 2026. URL https://arxiv.org/abs/2608.06544.

Sonia Joseph, Quentin Garrido, Randall Balestriero, Matthew Kowal, Thomas Fel, Shahab Bakhtiari, Blake Richards, and Mike Rabbat. Interpreting Physics in Video World Models. *arXiv preprint arXiv:2602.07050*, 2026. URL https://arxiv.org/abs/2602.07050.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://arxiv.org/abs/1711.05101.

Lucas Maes, Quentin Le Lidec, Damien Scieur, Yann LeCun, and Randall Balestriero. LeWorldModel: Stable End-to-End Joint-Embedding Predictive Architecture from Pixels. *arXiv preprint arXiv:2603.19312*, 2026. URL https://arxiv.org/abs/2603.19312.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. URL https://arxiv.org/abs/1709.07871.

Abhilasha Ravichander, Yonatan Belinkov, and Eduard Hovy. Probing the probing paradigm: Does probing accuracy entail task relevance? In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pp. 3363–3377. Association for Computational Linguistics, 2021. doi: 10.18653/v1/2021.eacl-main.295. URL https://aclanthology.org/2021.eacl-main.295/.

Kaizhen Tan, Xin Xu, Siru Tao, Yixiao Li, Hanzhe Hong, Yang Feng, and Heqing Du. What Can Latent World Models Know? Physical Parameter Identifiability in Multimodal Predictive Representations. *arXiv preprint arXiv:2607.27017*, 2026. URL https://arxiv.org/abs/2607.27017.

Stone Tao, Fanbo Xiang, Arth Shukla, Yuzhe Qin, Xander Hinrichsen, Xiaodi Yuan, Chen Bao, Xinsong Lin, Yulin Liu, Tse kai Chan, Yuan Gao, Xuanlin Li, Tongzhou Mu, Nan Xiao, Arnav Gurha, Viswesh Nagaswamy Rajesh, Yong Woo Choi, Yen-Ru Chen, Zhiao Huang, Roberto Calandra, Rui Chen, Shan Luo, and Hao Su. ManiSkill3: GPU parallelized robotics simulation and rendering for generalizable embodied ai, 2025. URL https://arxiv.org/abs/2410.00425.

Adam J. Thorpe, Stepan Tretiakov, Cheng-Hsi Hsiao, Su Ann Low, Xingjian Li, Hassan Iqbal, Neel P. Bhatt, Ufuk Topcu, and Krishna Kumar. Physically Viable World Models: A Case for Query-Conditioned Embodied AI. *arXiv preprint arXiv:2605.30542*, 2026. URL https://arxiv.org/abs/2605.30542.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000. URL https://arxiv.org/abs/physics/0004057.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1706.03762.

Aleksandar Vujinović and Aleksandar Kovačević. ACT-JEPA: Novel joint-embedding predictive architecture for efficient policy representation learning. *IEEE Access*, 14:78895–78906, 2026. ISSN 2169-3536. doi: 10.1109/access.2026.3696039. URL http://dx.doi.org/10.1109/ACCESS.2026.3696039.

Lirui Wang, Xinlei Chen, Jialiang Zhao, and Kaiming He. Scaling Proprioceptive-Visual Learning with Heterogeneous Pre-trained Transformers. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-3952. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/e0f393e7980a24fd12fa6f15adfa25fb-Abstract-Conference.html.

Zhenjia Xu, Jiajun Wu, Andy Zeng, Joshua Tenenbaum, and Shuran Song. DensePhysNet: Learning Dense Physical Object Representations via Multi-Step Dynamic Interactions. In *Proceedings of Robotics: Science and Systems*, 2019. doi: 10.15607/RSS.2019.XV.046. URL https://www.roboticsproceedings.org/rss15/p46.html.

Xi Zeng, Haojie Ren, and Ziying Song. PhyLatent: Learning Dynamics-Relevant Representations for JEPA World Models. *arXiv preprint arXiv:2608.05720*, 2026. URL https://arxiv.org/abs/2608.05720.

Guanqi Zhan, Xianzheng Ma, Weidi Xie, and Andrew Zisserman. Inferring Dynamic Physical Properties from Video Foundation Models. *arXiv preprint arXiv:2510.02311*, 2025. URL https://arxiv.org/abs/2510.02311.

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. URL https://arxiv.org/abs/2006.10742.

Xiangteng Zhang, Yang Guan, Bo Zhang, Hongyang Li, Ya-Qin Zhang, and Shengbo Eben Li. On the Identifiability of Controlled World Models. *arXiv preprint arXiv:2607.22430*, 2026. URL https://arxiv.org/abs/2607.22430.

Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. In *Proceedings of Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.016. URL https://www.roboticsproceedings.org/rss19/p016.html.

Gaoyue Zhou, Hengkai Pan, Yann LeCun, and Lerrel Pinto. DINO-WM: World models on pre-trained visual features enable zero-shot planning. *arXiv preprint arXiv:2411.04983*, 2024. URL https://arxiv.org/abs/2411.04983.

APPENDIX CONTENTS

A Protocols and Implementation **15**
A.1 Evaluation Tracks and Data Splits . . . 15
A.2 Candidates, Utilities, and Decision Metrics . . . 15
A.3 Frozen-Readout Protocol . . . 16
A.4 Architecture and Training Configurations . . . 17
A.5 Baselines and Structured Ablations . . . 18

B Supplementary Diagnostics **19**
B.1 Frozen-Representation Readouts . . . 19
B.2 Physical-Evidence Diagnostics (RQ2) . . . 19
B.3 Candidate-Regret Ablations (RQ3) . . . 20
B.4 Grounding Controls (RQ4) . . . 21
B.5 Capacity, Selectivity, and Regret (RQ5) . . . 22

C Detailed Results and Evaluation Scope **23**
C.1 Per-Objective Main Results and Uncertainty . . . 23
C.2 Physical and Belief Interventions . . . 24
C.3 Runtime, Failure Cases, and Generalization Scope . . . 25
Execution Outcomes . . . 26

The appendix separates the visual candidate-selection study, structured physical diagnostics, and frozen readouts. Their inputs, objectives, and checkpoints differ; results are not pooled across these tracks. Appendix A.4 presents the future-mediated architecture from the main text, then specifies the separate direct-readout (DR) configuration used for the reported compression study.

A PROTOCOLS AND IMPLEMENTATION

Throughout the appendix, M denotes the number of evidence slots, D the observation-embedding width, d the bottleneck width, N the number of candidates, and K the number of predicted phases, as in Section 3. The task evaluator $\mathcal{U}_q$ maps a realized rollout to utility, whereas U_η predicts utility from a latent future and the query. Readout sample counts and output dimensions use separate symbols defined below.

A.1 EVALUATION TRACKS AND DATA SPLITS

Visual candidate-selection data. The main run uses instrumented versions of ManiSkill3 `PushCube-v1` and `PickCube-v1`, a Panda robot, and `pd_ee_delta_pose` control at 20 Hz. RGB images are resized to 96×96. Each source episode supplies 32 frames sampled at uniformly spaced indices over a fixed probe history, with aligned proprioception and recorded actions. This is a resampled history, not necessarily 32 consecutive control steps. The 25-dimensional proprioceptive vector contains nine joint positions, nine velocities, and the end-effector position and quaternion. Queries are four-dimensional one-hot objective encodings, not free-form language. Physical parameters are not network inputs.

Each environment has 384/96/192/192 source episodes for training, validation, ID testing, and OOD testing. Both query views use the same history and nine candidate executions; all views and branches remain in the source episode's split. Push and Pick data seeds are 81011 and 91021. The saved post-probe simulator state is restored before each candidate, so comparisons share the same initial condition. Failed candidate executions are retained rather than filtered by success.

Physical distributions. Mass is expressed as a multiplier of the environment's default object mass, not in kilograms. Training, validation, and ID testing sample mass scale uniformly from $[0.60, 1.40]$, table friction from $[0.15, 0.65]$, and gripper friction from $[0.70, 1.80]$. Push varies mass and table friction; Pick varies mass and gripper friction. OOD mass ranges are $[0.30, 0.50]$ and $[1.50, 2.00]$; table-friction ranges are $[0.05, 0.12]$ and $[0.70, 0.90]$; gripper-friction ranges are $[0.30, 0.60]$ and $[1.90, 2.30]$. Static and dynamic friction are set to the same coefficient. Each OOD split contains 64 mass-only, 64 friction-only, and 64 joint-shift histories. Low and high ranges alternate by episode index; within a range sampling is uniform. Unshifted factors retain their ID distribution. In joint shifts, both changed factors use the same low/high branch.

Separate diagnostic tracks. The structured experiment has 320/80/160 training/validation/test histories, 48 matched counterfactual groups, data seed 90217, and model seed 24117. Its 14 channels comprise object position and velocity in the horizontal and vertical directions; horizontal-force, grip-force, and lift-acceleration commands; measured normal and tangential forces; two contact indicators; pusher position, gripper aperture, and end-effector height. The physical identification and pixel diagnostics in Appendix B.2 use separate synthetic datasets. Frozen readouts use seeds 42, 43, and 44, whereas the visual candidate-selection run uses 101, 211, and 307. These seed sets and observation protocols are not interchangeable.

A.2 CANDIDATES, UTILITIES, AND DECISION METRICS

Fixed candidate controllers. Push uses the Cartesian product of normalized end-effector command limits $\{0.04, 0.20, 0.38\}$ and goal offsets $\{-0.060, 0, 0.060\}$ m along the initial object-to-goal direction. Each controller applies planar feedback for at most 38 steps, stops when its control-point error is below 0.004 m, and holds for five steps. Pick crosses gripper commands $\{-0.40, -0.70, -1.00\}$ with three lift-command-limit/hold-length pairs: $(0.10, 20)$, $(0.20, 12)$, and $(0.38, 5)$. It first closes for ten steps, then lifts for at most 38 steps with a 0.006 m control-point tolerance, and holds for the specified duration. The command limits are normalized controller inputs, not physical displacements in metres.

The network's candidate vector has five components. Push encodes $[0, a_{\max}, \delta/0.02, 0, 0]$, where δ is the offset in metres; Pick encodes $[1, n_{\text{hold}}/20, 0, g, a_{\max}]$, where g is the gripper command. The fixed feedback controllers use simulator object and goal poses. This access belongs to the execution harness; it is not an extra input to the learned history representation. Candidate selection therefore

evaluates a learned selector coupled to these predefined controllers, not a standalone visual control policy.

Task utilities. Let $b_-(x;s) = 1 - \mathrm{clip}(x/s, 0, 1)$ and $b_+(x;s) = \mathrm{clip}(x/s, 0, 1)$. The task evaluator uses

$$u_{\text{precise}} = 0.35 s_{\text{env}} + 0.35 b_-(e; 0.10) + 0.20 b_-(o; 0.06) + 0.10(1 - F), \tag{17}$$

$$u_{\text{efficient}} = 0.55 s_{\text{env}} + 0.25 b_-(n; 40) + 0.15 b_-(E; 25) + 0.05 b_-(e; 0.10), \tag{18}$$

$$u_{\text{stable}} = 0.20 s_{\text{env}} + 0.35 s_{\text{hold}} + 0.20 b_-(\delta_{\text{slip}}; 0.05) + 0.15(1 - d_{\text{drop}}) + 0.10 b_+(n_g; 25), \tag{19}$$

$$u_{\text{rapid}} = 0.55 s_{\text{env}} + 0.25 b_-(n_*; 65) + 0.15 b_-(E; 45) + 0.05(1 - d_{\text{drop}}). \tag{20}$$

Here s_{env} is terminal environment success; e is terminal planar Push error; o is maximum positive overshoot along the initial goal direction; F indicates a final off-table object; n is the action count; and $E = \sum_t \|a_t\|_2^2$ is an action-effort proxy, not measured energy. For Pick, δ_{slip} is the maximum change in object–gripper relative position over the last 15 frames, and n_g counts grasped frames. The drop flag requires a rise exceeding 0.04 m followed by a loss exceeding 0.04 m. s_{hold} requires success throughout the final ten frames, slip at most 0.025 m, and no drop. The completion index n_* is the zero-based first successful frame, or the trace length if none succeeds. Distances are in metres; time-related scales are frame or action counts. All coefficients are fixed by the task evaluator.

Aggregation and ties. For a fixed history–query pair, write $u_i = u_i^{(q)}$, with $u^* = \max_i u_i$ and $u_* = \min_i u_i$. The selected index $\hat{i} = \hat{i}(q)$ is defined in Section 3.4. Top-1 correctness is $\mathbb{1}[u_{\hat{i}} \geq u^* - 10^{-8}]$; score ties select the first candidate in the fixed ordering. Accuracies are first computed per objective and training seed, then macro-averaged equally across the four objectives and three seeds. The candidate selector's normalized regret is

$$r_{\text{sel}} = \frac{u^* - u_{\hat{i}}}{u^* - u_*}, \qquad u^* - u_* > 10^{-8}. \tag{21}$$

The visual evaluator excludes degenerate ranges from the regret mean and returns an undefined value if none remain. Its ranking accuracy excludes pairs with utility difference at most 10^{-8}, pools valid pairs, and counts a predicted tie as incorrect. The structured evaluator instead floors the regret denominator at 10^{-8}, assigns zero regret to an all-tied set, and averages per-example ranking accuracies; an example with no valid pair receives ranking accuracy one. These conventions distinguish the diagnostic scores from the visual metrics. The training ranking threshold 0.01 is separate from the evaluation tolerance.

A.3 Frozen-Readout Protocol

Following the frozen-encoder design of ACT-JEPA (Vujinović & Kovačević, 2026), separate heads predict 30 consecutive future proprioceptive states starting at $t+1$, or four future seven-dimensional actions. Anchors are sampled at stride four from each query's highest-utility recorded candidate branch, without success filtering, using up to 64 preceding frames. For Q-PhysJEPA, the frozen features concatenate the query-conditioned history representation and current-observation embedding; the readout is not restricted to z_q.

The shared readout design is a two-layer temporal cross-attention decoder with width 128, four heads, feed-forward width 256, and dropout 0.1. Per-coordinate means and standard deviations are fitted only on valid training targets, with a 10^{-4} standard-deviation floor. The recorded coordinate and quaternion ordering is retained without rotation alignment. The original head schedule uses 500 AdamW (Loshchilov & Hutter, 2019) updates, batch size 256, learning rate 6×10^{-4}, weight decay 0.1, cosine decay after 40 warmup updates, and gradient clipping at 1. Validation standardized MSE is checked every 100 updates and selects the checkpoint.

Let N_v be the number of valid readout prediction–target pairs and D_y their output dimension: $D_y = 25$ for state prediction and $D_y = 7$ for action prediction. These differ from the evidence-slot count M and latent embedding width D. The metrics are

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_v D_y} \sum_{j=1}^{N_v} \|\widehat{y}_j - y_j\|_2^2}, \qquad \mathrm{ATE} = \frac{1}{N_v} \sum_{j=1}^{N_v} \|\widehat{y}_j - y_j\|_2. \tag{22}$$

RMSE is evaluated for state and action readouts; ATE is reported for state readouts. Predictions are mapped back to the recorded target scale, padding is excluded, and valid-timestep errors are aggregated within each seed before averaging over 42, 43, and 44. ATE is the mean Euclidean error of the full 25-dimensional state, not a position-only error in metres. Its joint positions, velocities, and pose components have heterogeneous scales.

The tuning records include 2,000-update state-head searches over learning rates 3×10^{-4} and 6×10^{-4} and weight decays 0.01 and 0.1, followed by Q-PhysJEPA-only schedules extending to 4,000 updates. The reported state table combines the selected Q-PhysJEPA readouts with the original baseline settings, rather than a common final tuning budget. Action readouts retain their original settings. The state comparison therefore does not isolate representation quality from readout optimization.

A.4 Architecture and Training Configurations

Forward interface of Q-PhysJEPA. Q-PhysJEPA denotes the future-mediated architecture in Sections 3.3–3.5. Interaction observations and their preceding actions are encoded into slots S, and query-conditioned gating produces the bottleneck z_q. The candidate predictor receives $(z_q, q, c_i, \boldsymbol{\rho})$ and is initialized only from $[z_q; q]$. The uncompressed history features R, evidence slots S, and current-observation embedding e_T have no direct route to the predictor. The reference dimensions are $D = 64$, $d = 16$, $M = 8$, and $K = 4$, with phases $(0.25, 0.5, 0.75, 1)$.

All candidates share the initializer and GRU parameters. Their normalized controller settings determine their respective latent rollouts. For candidate i, the utility feature is exactly the process–terminal summary in Eq. 12:

$$f_i = \left[K^{-1} \sum_{k=1}^{K} \widehat{e}_{i,k}; \widehat{e}_{i,K} \right], \qquad \widehat{u}_i^{(q)} = U_\eta([f_i; q]).$$

With the four-dimensional query, U_η is a $132 \to 128 \to 128 \to 1$ MLP with GELU hidden activations and an unclipped scalar output. It has no direct input from z_q, e_T, the candidate parameters, or the realized execution length. Thus, all history- and candidate-dependent information used for scoring passes through the predicted future. Mean pooling alone is not the process–terminal feature used by this specification.

Prediction targets and joint supervision. For the recorded branch i_b, phase ρ_k selects observation $o^{(i_b)}_{\lceil \rho_k (L_{i_b} - 1) \rceil}$, which the observation encoder E_θ maps to the latent target. These are embeddings of actual observations, not projections of utility labels or physical-parameter annotations. Realized lengths are used only to align training targets; realized future observations and utilities are unavailable to the forward path at decision time. The losses and weights follow Eq. 16: prediction aligns latent futures with observed executions, while utility regression and ranking backpropagate through U_η, the predictor, and the query-conditioned abstraction. Variance and covariance regularization constrain embedding statistics. This interface defines a learning constraint, not physical identifiability.

Reference configuration. The joint-objective reference schedule is AdamW, 30 epochs, batch size 8, learning rate 3×10^{-4}, weight decay 10^{-4}, and gradient clipping at 1. Its loss weights are $(\lambda_w, \lambda_d, \lambda_v, \lambda_c, \lambda_r) = (1, 1, 16, 8, 0.1)$, with Smooth L_1 transition $\beta = 0.1$ and ranking threshold $\tau = 0.01$. These settings specify the complete model; they are not the configuration that produced the direct-readout accuracy or runtime measurements below.

Direct-readout compression study (DR). Table 1, Appendix C.1, and the candidate-scorer runtime table evaluate the compression component with a separate direct readout. Denoting its parameters by ξ, its score is

$$\widehat{u}^{(q)}_{i,\mathrm{DR}} = D_\xi([z_q; q; c_i]),$$

with z_q replaced by a query-independent representation for the no-query control. This is distinct from $U_\eta([f_i; q])$. Its auxiliary target is a fixed projection of normalized utilities, not a future-observation embedding. We label the query-conditioned variant ***Q-PhysJEPA (DR)*** in the main

comparison; it retains the source identifier `q_physjepa` in the archived exports. The 56.29% versus 46.83% comparison therefore measures the compression/readout combination, not the complete future-mediated architecture.

The archived run is `exp_20260901_233757_cd53e9`. Its 96×96 image encoder follows a patch-Transformer design (Dosovitskiy et al., 2021), using 12×12 patches, two blocks, four heads, and width 64. A two-layer history encoder produces eight slots and a width-16 bottleneck. Push and Pick are trained separately with seeds 101, 211, and 307, AdamW, batch size 32, learning rate 3×10^{-4}, weight decay 10^{-4}, at most 80 epochs, patience 12, and clipping at 5. Validation mean query utility selects the checkpoint. The auxiliary utility-projection loss and utility regression each have weight one; ranking has weight 0.2, Smooth L_1 transition 0.1, and training tie threshold 0.01. This auxiliary loss is not $\mathcal{L}_{\text{pred}}$ from Eq. 13.

Relation to the supplementary tracks. The structured comparisons use frozen representations and a separately trained candidate head (Appendix A.5); the state/action readouts use the features in Appendix A.3. These measurements remain assigned to their respective interfaces and checkpoints rather than being combined as evaluations of the full model in Figure 2.

A.5 Baselines and Structured Ablations

Visual controls. The DR no-query baseline removes query dependence from the history representation, not from the candidate scorer: it still receives the task query. Q-PhysJEPA (DR) gates evidence before compression. Query-only removes the history representation while retaining the query and candidate input. The no-query and Q-PhysJEPA DR implementations each contain 317,401 parameters and share data, candidates, training schedule, and model seeds. The oracle chooses the best realized utility among the same nine executed candidates; it is not a learned model given additional history features.

Structured variants. Plain JEPA uses a query-independent representation and predictor. Strong late fusion introduces the query after compression, using a detached base representation for its query-specific route. Q-Predictor conditions the predictor but not compression; Q-Gate conditions compression but retains the query-independent predictor. Q-PhysJEPA conditions both. Frozen representations are subsequently evaluated by a separately trained, query-aware candidate decision head. These variants do not use the future-only utility architecture as a shared scoring implementation.

The structured models use 64-dimensional tokens, eight slots, four heads, two encoder layers, and bottleneck widths 8 or 16. All five instantiate 137,024 parameters at width 8 and 160,184 at width 16, including inactive routes; equal counts do not imply identical active computation. Training uses seed 24117, batch size 64, AdamW learning rate 7×10^{-4}, weight decay 10^{-4}, at most 80 epochs, and patience 14. The downstream head is trained for at most 60 epochs with patience 10. Auxiliary targets project action value, task success, and constraint violation into 64 dimensions with fixed projection seed 90243. The representation and scoring procedures are therefore separate from the visual experiment. Only the reported seed is used for the comparisons; reserved confirmation seeds are not additional measurements.

B Supplementary Diagnostics

B.1 Frozen-Representation Readouts

We compare frozen readouts of ACT (Zhao et al., 2023), ACT-JEPA (Vujinović & Kovačević, 2026), and Q-PhysJEPA under the protocol in Appendix A.3. Q-PhysJEPA achieves the lowest reported

Table A1: Frozen-representation readout errors on Push and Pick. Lower is better for all metrics; best results are in bold.

	Push			Pick		
Method	State RMSE ↓	ATE ↓	Action RMSE ↓	State RMSE ↓	ATE ↓	Action RMSE ↓
Behavior-cloning baseline						
ACT	0.0824	0.2449	0.0422	0.0804	0.3101	0.0649
Joint-embedding predictive models						
ACT-JEPA Joint	0.0832	0.2187	0.0402	0.0775	0.2902	0.0634
ACT-JEPA Two-stage	0.0806	**0.2083**	0.0372	0.0788	0.2987	0.0665
Q-PhysJEPA†	**0.0733**	0.2104	**0.0260**	**0.0701**	**0.2378**	**0.0535**

† Q-PhysJEPA uses additionally tuned state readout heads; its action readout results are unchanged. ACT baselines retain their original configurations. The tuning budgets are therefore not matched. All values are rounded to four decimal places; gray highlights our method.

error on five of six readouts. The exception is Push ATE, where ACT-JEPA Two-stage obtains 0.2083 versus 0.2104. Q-PhysJEPA's action RMSE is 0.0260 on Push and 0.0535 on Pick, versus 0.0372 and 0.0665 for ACT-JEPA Two-stage. Because only Q-PhysJEPA receives additional state-head tuning, the state-readout results provide a diagnostic comparison rather than a benchmark with matched tuning budgets.

B.2 Physical-Evidence Diagnostics (RQ2)

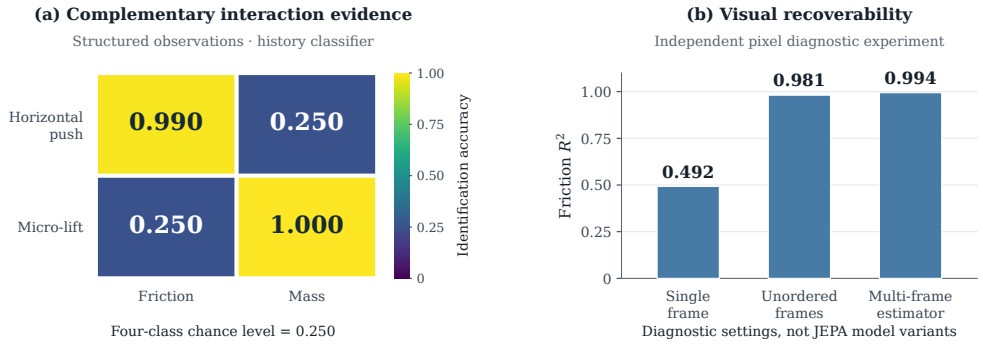

Figure A1: Complementary interaction evidence and independent visual recoverability. The left panel reports four-class identification using structured observations, and the right panel reports friction readouts from pixel inputs. These panels use distinct diagnostic settings. They compare neither Q-PhysJEPA model variants nor temporal ordering within an otherwise identical estimator.

Horizontal pushing supports friction identification with accuracy 0.990 while mass remains at chance (0.250). Conversely, micro-lifting yields mass accuracy 1.000 and friction accuracy 0.250. These results show that the two interactions provide complementary evidence about the physical properties. In the independent pixel diagnostic, friction R^2 is 0.492 with one frame, 0.981 with unordered frames, and 0.994 with the multi-frame estimator. The strong unordered-frame result indicates that multiple frames can provide useful information without preserving their order, potentially through static or state-difference cues. Because the estimators also differ, this comparison does not isolate the effect of temporal order.

Structured identification. The left panel uses analytic probe trajectories with dataset seed 12345, four friction coefficients $\{0.1, 0.3, 0.5, 0.7\}$, and four masses $\{0.25, 0.35, 0.45, 0.55\}$ kg. Each probe has 576/96/256 training/validation/test trajectories with disjoint appearance identities. Sequences contain 16 samples at 0.005 s intervals and structured motion, force, contact, appearance, and probe-phase features. The learned estimator is a width-64 GRU with separate four-class friction and mass heads, trained with AdamW for 100 epochs, batch size 128, learning rate 10^{-3}, and weight decay 10^{-4}; validation loss selects the checkpoint. Results average seeds 701, 809, and 907. This is a structured belief-identification test, not an end-to-end visual JEPA comparison.

Pixel recoverability. The independent horizontal-glide dataset uses seed 77411 and 24/8/16 training/validation/test context groups. Each group contains eight friction levels spanning 0.08–0.65, yielding 192/64/128 trajectories; appearance identities are disjoint across splits. Images have 128×128 pixels, and motion is sampled every 0.02 s after a two-step pulse. Mass, grip friction, and stiffness are fixed at 0.50 kg, 0.55, and 700 N/m. The multi-frame estimator fits a quadratic to tracked pixel positions in the free-glide interval and infers friction from its acceleration. Single-frame and unordered controls instead fit ridge regressions ($\alpha = 10^{-4}$) on image moments with recorded action/contact features. The single frame is selected on validation data. The unordered control uses means, standard deviations, extrema, and quartiles over frame features, rather than a new random frame permutation at test time. Different estimator classes and input summaries therefore contribute to this comparison; the 0.013 R^2 gap is not an isolated temporal-order effect. No confidence interval for that gap is reported.

B.3 Candidate-Regret Ablations (RQ3)

Figure 3 compares plain JEPA, late fusion, Q-Predictor, Q-Gate, and Q-PhysJEPA. The corresponding strong late-fusion regrets reported in Table 2 are 0.2510 and 0.1915. For variants whose exact values are not tabulated, the comparison relies on the plotted results. The ablations support the tested conditioning design, while identifying a causal connection between property readouts and decisions requires matched interventions on the evidence or belief with the remaining decision problem held fixed. The separate interventions in Appendix C.2 provide complementary evidence, but do not jointly track property readouts, predicted latent futures, and decisions on the same physical counterfactual pairs.

The five conditioning routes and their matched instantiated parameter counts are specified in Appendix A.5. The "late fusion" curve denotes the strong late-fusion condition in Table 2. The reported regrets of 0.1680 and 0.1500 improve on that control but remain above the structured protocol's absolute-regret target of 0.10. The ranking gains, 2.86 and 2.31 percentage points, also fall below its three-point target. Consequently, the lowest regret among the tested variants does not imply that every diagnostic criterion was satisfied. This single-seed comparison concerns query conditioning routes, evaluated through the frozen-representation readout protocol in Appendix A.5.

B.4 GROUNDING CONTROLS (RQ4)

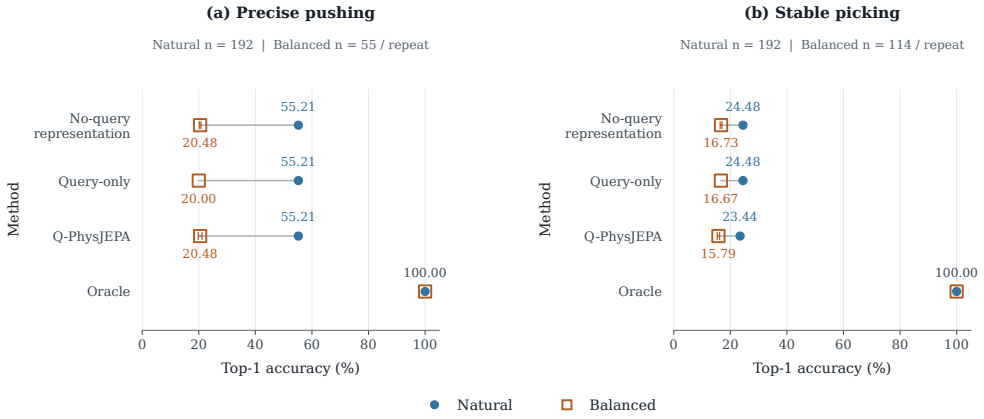

Figure A2: Candidate-selection accuracy on precise pushing and stable picking. Natural evaluation uses 192 examples per task; the balanced subsets contain 55 and 114 examples per repeat, respectively. No-query, query-only, and Q-PhysJEPA results remain close. The oracle selects the highest realized utility among the same nine candidates. These are ID controls from the archived candidate-scoring run.

Table A2 agrees with the ID entries in the archived `balanced_accuracy.csv` export and the rounded figure annotations. The no-query and Q-PhysJEPA natural-distribution entries are the same objective-specific ID results reported in Table A5.

Table A2: RQ4 ID top-1 accuracy (%), verified against the archived metric export. These two-objective controls are distinct from the four-objective OOD macro-average in Table 1.

	Precise pushing		Stable picking	
Method	Natural	Balanced	Natural	Balanced
No-query representation	55.21	20.48	24.48	16.73
Query-only	55.21	20.00	24.48	16.67
Q-PhysJEPA	55.21	20.48	23.44	15.79
Oracle	100.00	100.00	100.00	100.00

These controls show similar performance across the three evaluated methods, with no Q-PhysJEPA advantage on the two displayed objectives. The drop under balancing is consistent with sensitivity to candidate-label frequencies, although it does not identify a shortcut or its mechanism. The controls reuse the same archived models and candidate records, but change the evaluated objective subset and sampling distribution. Their ID scores should therefore not be substituted for the four-objective OOD macro-average.

Balancing and oracle access. Episodes are grouped by the first maximum-utility candidate in the fixed ordering. Precise-pushing label counts for candidates 0–8 are $(1, 0, 0, 9, 36, 16, 14, 105, 11)$; stable-picking counts are $(37, 47, 0, 27, 19, 0, 22, 40, 0)$. Classes with fewer than ten episodes are excluded, and every remaining class is sampled without replacement to match the smallest retained class. Thus, precise pushing retains five classes with 11 episodes each, and stable picking retains six classes with 19 each. Five repeats use sampling seeds 0–4, with identical selected episodes for every method and training seed. Each repeat first averages the three training seeds; Table A2 then averages repeats. The sampling standard deviations are 0.27 pp for both no-query and Q-PhysJEPA on precise pushing, and 0.32/0.41 pp on stable picking, respectively. These are subset-sampling deviations, not confidence intervals over training runs. The oracle uses realized rollout utilities, including simulator-derived outcome metrics; it does not predict unseen outcomes and is not a learnable-control baseline.

Four-objective OOD controls. The same archived exports also contain query-only results for all four OOD objectives. Table A3 completes the comparison on the split used for Table 1, without introducing new checkpoints or pooling it with a future-mediated experiment.

Table A3: Existing OOD direct-readout controls, in percent. Natural scores average the three training seeds; balanced scores additionally average the five fixed subsamples. All values come from the archived metric exports. N/A denotes a single retained oracle-label class, for which the balanced comparison is not defined.

Method	Precise	Efficient	Stable	Rapid
Natural distribution				
No-query (DR)	33.51	55.90	23.44	74.48
Query-only	33.33	78.65	22.92	100.00
Q-PhysJEPA (DR)	33.16	76.22	22.57	93.23
Balanced retained classes				
No-query (DR)	20.38	31.70	17.11	N/A
Query-only	20.00	33.33	16.67	N/A
Q-PhysJEPA (DR)	19.90	33.48	16.83	N/A

Query-only reaches 78.65% on efficient transport and 100.00% on rapid completion, compared with 76.22% and 93.23% for Q-PhysJEPA (DR). For rapid completion, all 192 OOD histories have the same first maximum-utility candidate, index 2; its natural-distribution score therefore does not require distinguishing histories. For efficient transport, balancing retains three classes and gives 33.33% for query-only and 33.48% for Q-PhysJEPA (DR). Thus, the gain over the no-query representation does not establish an advantage over the query-and-candidate prior on these two objectives. These controls concern the archived DR implementation. A current-observation-only control is not included in these exports.

B.5 Capacity, Selectivity, and Regret (RQ5)

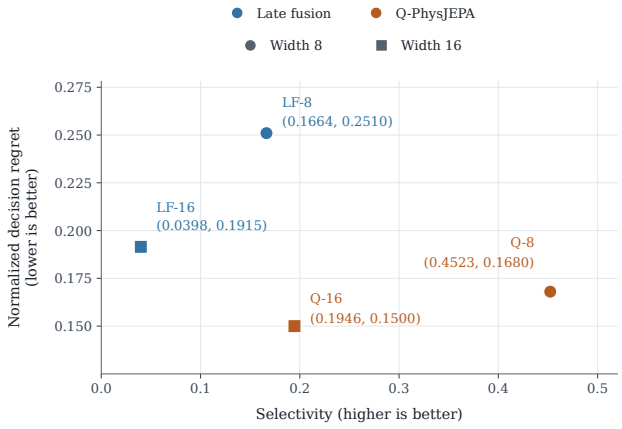

Figure A3: Selectivity and regret for the four conditions in Table 2, replotted from the same experiment. Selectivity is relevant R^2 minus leakage; lower regret indicates better candidate selection.

For Q-PhysJEPA, width 8 yields selectivity 0.4523 and regret 0.1680; width 16 yields 0.1946 and 0.1500. Strong late fusion changes from (0.1664, 0.2510) to (0.0398, 0.1915). Within both model families, the wider representation is less selective but achieves lower candidate regret. Because selectivity is the difference between relevant- and irrelevant-property recovery, either term can change as capacity increases. These four single-seed conditions show a divergence between selectivity and candidate-selection performance over the tested widths, without establishing a general scaling law or a causal effect of width.

Probe and relevance definitions. For each of the five structured objectives, a ridge probe predicts mass, table friction, grip friction, and stiffness from the frozen representation. Probes use the 320 training histories and are evaluated on the 160 test histories. Features are standardized using training statistics; the ridge penalty is 10^{-4} and the intercept is unpenalized. Query–property relevance is defined by the average absolute change in the analytic planner's best attainable utility under a half-range physical intervention, normalized by the largest property effect for that query. It is a value-sensitivity definition, not a claim that the maximizing candidate must change.

Let $\mathcal{R}$ and $\mathcal{I}$ be query–property pairs with normalized relevance at least 0.15 and at most 0.08, respectively. Intermediate pairs are excluded from both groups. Relevant recovery is the mean R^2 over $\mathcal{R}$; leakage is the mean $\max(R^2, 0)$ over $\mathcal{I}$; selectivity is their difference. This clipping prevents negative probe R^2 from artificially improving the leakage score. These pairwise means need not weight every query equally. Within a width, instantiated parameter counts and training budgets are matched; across widths, parameter counts change. The observed selectivity–regret relationship therefore does not separate bottleneck width from total model capacity, and its four points are not independent seed replications.

Value sensitivity versus choice relevance. The preceding relevance groups describe sensitivity of the best attainable *value*, whereas the decision criterion in the Introduction concerns which candidate should be selected. Write $V_q(\vartheta) = \max_i u_i^{(q)}(\vartheta)$ for physical condition ϑ. A change in V_q need not change the maximizing candidate; conversely, the maximizing candidate can change while V_q stays fixed. For example, candidate utilities $(0.9, 0.6)$ and $(0.6, 0.9)$ have the same maximum but opposite preferred candidates. The property groups in Table 2 and the structured sensitivity analysis therefore use value sensitivity as an operational grouping rule, not as a certificate of decision sufficiency.

A complementary choice-level quantity is the cost of retaining the previous condition's optimal candidate:

$$d_q(\vartheta \to \vartheta') = V_q(\vartheta') - u^{(q)}_{i_q^\star(\vartheta)}(\vartheta'), \qquad i_q^\star(\vartheta) = \arg\max_i u_i^{(q)}(\vartheta).$$

Ties follow the fixed-order convention in Appendix A.2. This definition distinguishes loss from an outdated choice from a common shift in all candidate values. It is provided to clarify the distinction; it was not used to assign the reported relevance groups or to recompute their scores. Candidate regret and pairwise ranking are evaluated separately. Accordingly, selectivity describes recoverability of value-sensitive properties, while decision quality is assessed by the candidate metrics.

C Detailed Results and Evaluation Scope

This section reports how direct-readout candidate selection varies across objectives, training seeds, and physical shifts. Separate intervention tests examine representation sensitivity, planning with physical beliefs, and temporal masking.

C.1 Per-Objective Main Results and Uncertainty

The direct-readout (DR) compression experiment `exp_20260901_233757_cd53e9` uses training seeds 101, 211, and 307. Each ID and OOD split contains 192 Push and 192 Pick histories, each paired with two objectives and nine candidates. Top-1 accuracy counts a selection as correct when its utility is within 10^{-8} of the maximum. No-query uses a query-independent representation and a query-aware candidate scorer.

Table A4: Per-objective OOD Top-1 accuracy. Differences are Q-PhysJEPA (DR) minus no-query, in percentage points.

Objective	No-query (%)	Q-PhysJEPA (DR) (%)	Δ (pp)	95% CI (pp)
Precise positioning	33.51	33.16	-0.35	[-1.04, 0.35]
Efficient transport	55.90	76.22	+20.31	[15.97, 24.65]
Stable holding	23.44	22.57	-0.87	[-2.08, 0.35]
Rapid completion	74.48	93.23	+18.75	[14.93, 22.40]
Equal-weight macro	46.83	56.29	+9.46	[8.03, 10.94]

Table A5: Corresponding ID Top-1 accuracy under the same evaluation protocol.

Objective	No-query (%)	Q-PhysJEPA (DR) (%)	Δ (pp)	95% CI (pp)
Precise positioning	55.21	55.21	+0.00	[-0.87, 0.87]
Efficient transport	60.24	85.07	+24.83	[20.66, 28.99]
Stable holding	24.48	23.44	-1.04	[-2.09, 0.00]
Rapid completion	75.17	94.97	+19.79	[15.97, 23.61]
Equal-weight macro	53.78	64.67	+10.89	[9.46, 12.37]

We compute 95% confidence intervals from 10,000 paired bootstrap samples of source episodes, stratified by task and OOD shift type. Query views from the same episode are sampled together, and effects are averaged across training seeds within each episode before resampling. These post-hoc intervals describe variation across test scenes with the three trained models held fixed. They do not measure variation from retraining.

Across training seeds, OOD macro accuracy is $46.83 \pm 11.81\%$ for no-query and $56.29 \pm 1.39\%$ for Q-PhysJEPA (DR), where $\pm$ is the sample standard deviation. The three paired gains are $+22.53$, $+7.03$, and -1.17 percentage points. The average gain comes mainly from efficient transport and rapid completion, but it is not positive for every training seed.

Table A6: OOD macro accuracy by physical shift. Each row includes 64 histories per task. Friction refers to table friction for Push and gripper friction for Pick.

Shift	No-query (%)	Q-PhysJEPA (DR) (%)	Δ (pp)	95% CI (pp)
Mass only	56.25	65.76	+9.51	[7.03, 11.98]
Friction only	44.01	52.21	+8.20	[5.60, 10.81]
Joint shift	40.23	50.91	+10.68	[8.07, 13.28]

The archived per-objective exports record the source methods and sample counts. All results in this subsection come from the direct-readout implementation described in Appendix A.4.

C.2 Physical and Belief Interventions

Structured physical interventions. The single-seed structured diagnostic uses 48 matched groups, each containing a base history and four single-factor changes. The diagnostic program and episode context are fixed within each group. Each intervention shifts one factor by half its normalized range, with wrap-around, and leaves the other factors unchanged. Sensitivity is the distance between the resulting representations divided by the range-normalized change in that factor.

Table A7: Structured representation sensitivity, averaged over 48 groups. Measurements are classified as relevant when normalized relevance is at least 0.15 and irrelevant when it is at most 0.08. The ratio is relevant sensitivity divided by irrelevant sensitivity. These measurements are separate from the ManiSkill results.

Width	Method	Relevant	Irrelevant	Ratio
8	Late fusion	2.6506	1.1313	2.3429
8	Q-PhysJEPA	3.9793	0.8295	4.7971
16	Late fusion	3.3564	1.6550	2.0281
16	Q-PhysJEPA	5.6238	1.8119	3.1038

At both widths, Q-PhysJEPA responds more strongly to query-relevant changes relative to irrelevant changes than late fusion.

Controlled belief-to-planning test. We hold the analytic planner fixed and vary its input belief over four friction and four mass classes. The inputs come from the task-appropriate probe, the

alternative probe, another evaluation condition, or a uniform prior. The true-physics reference uses the actual physical classes.

Table A8: Belief-intervention success. Counts are in seed order 701, 809, 907, each with denominator 256; the same evaluation conditions are reused across seeds. Counts are recovered exactly from saved full-precision binary success means and denominators. This is a separate structured planning test, not visual Q-PhysJEPA execution.

Condition	Slide counts	Slide (%)	Lift counts	Lift (%)
Correct probe	256, 252, 252	98.96	256, 256, 256	100.00
Wrong probe	64, 64, 64	25.00	64, 64, 64	25.00
Shuffled belief	69, 61, 67	25.65	64, 67, 69	26.04
Uniform belief	64, 64, 64	25.00	64, 64, 64	25.00
True physics	256, 256, 256	100.00	256, 256, 256	100.00

The planner succeeds nearly as often with the task-appropriate probe as with the true physical classes. Alternative, shuffled, and uniform beliefs yield substantially lower success rates.

Temporal masking. The temporal-masking test divides each 32-frame history into four equal windows, keeping candidate actions and recorded outcomes fixed. Let $D(q, w)$ denote the increase in normalized regret after masking window w. The reported interaction is $I = [D(q_0, 3) - D(q_0, 0)] - [D(q_1, 3) - D(q_1, 0)]$.

Table A9: OOD all-input masking interactions. Intervals have 98.75% coverage for four planned comparisons. None meets both the interval criterion and the preregistered absolute-effect threshold of 0.02.

Comparison	Interaction	Interval
Q-PhysJEPA, push	0.00663	$[-0.00474, 0.01835]$
Q-PhysJEPA, pick	0.01342	[0.00159, 0.02718]
Push, difference from no-query	-0.00676	$[-0.01940, 0.00573]$
Pick, difference from no-query	0.01408	$[-0.00025, 0.02971]$

The positive Pick interaction falls below the preregistered effect threshold, and the comparison with no-query does not establish an advantage. In a separate test that holds the current observation fixed, masking RGB windows changes no selected candidates in 36,864 intervention records. All-input masking changes ten selections. Both counts pool the two methods and both test splits, and the records are not independent scenes. The 36,864 denominator applies only to RGB masking. The task relevance of the masked windows has not been independently verified.

Establishing how Q-PhysJEPA uses physical information in candidate selection still requires property readouts, predicted latent futures, and decisions to be measured jointly on the same matched physical counterfactual pairs.

C.3 Runtime, Failure Cases, and Generalization Scope

Runtime and reproducibility. The direct-readout candidate scorer runs on NVIDIA L20 hardware using Python 3.10.20, PyTorch 2.10.0+cu128, ManiSkill 3.0.1, and SAPIEN 3.0.3. Checkpoints are selected by validation mean query utility without using test data.

Table A10: Original candidate-scorer compute cost. Training time is mean ± seed standard deviation; memory is the maximum recorded across seeds. Forward time is batch-amortized per query, including all nine candidates, and excludes loading, rendering, and simulator execution.

Method	Parameters	Train (s)	Peak (MiB)	Forward (ms)
No-query	317401	50.78 ± 5.70	807.65	0.2122
Q-PhysJEPA	317401	61.24 ± 9.42	808.94	0.2117

These batch-amortized timings cover the archived direct candidate scorer. Separate encoding and per-candidate costs were not logged, and the later GRU future-prediction implementation was not timed here.

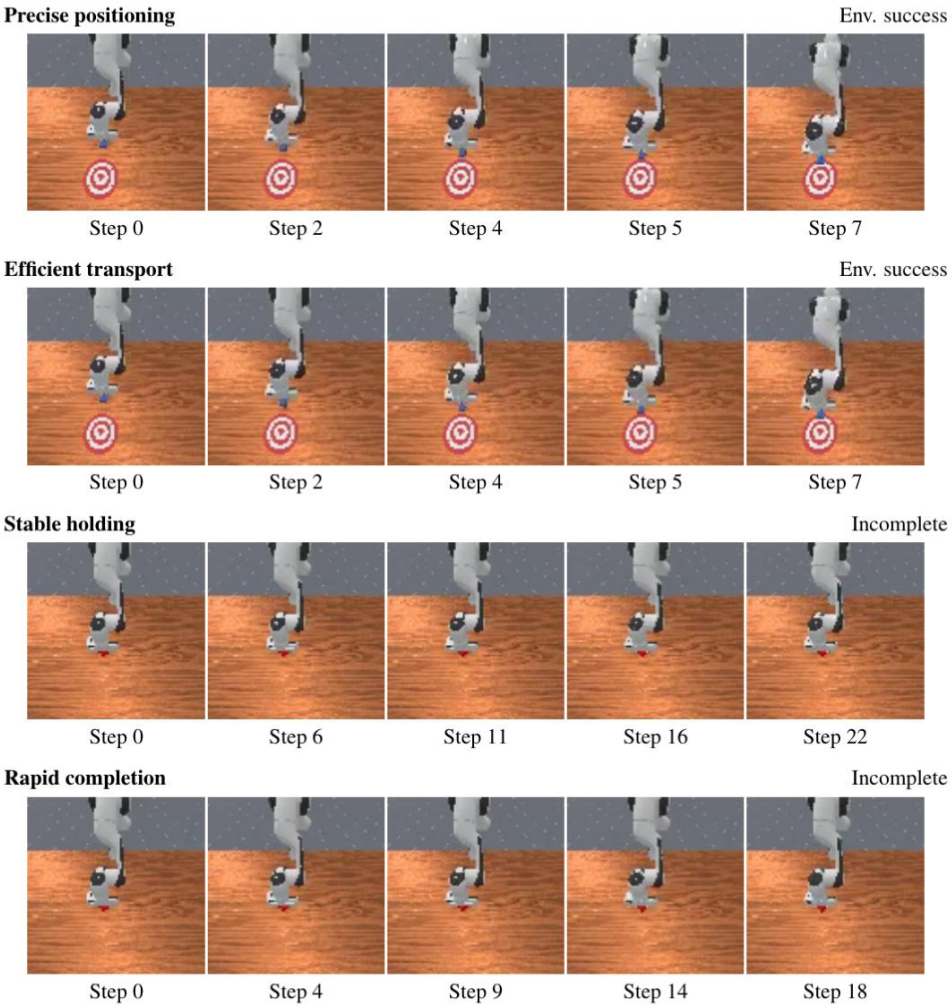

Figure A4: Separate continuous-action model: all four objectives at model seed 42, scene 10000, without success filtering. Five recorded frames per row follow the diagnostic prefix (20 Hz; native 96×96 pixels). Push meets the 10 cm tolerance; Pick is incomplete. These are not DR-selected trajectories.

Execution outcomes. Figure A4 shows an independently trained continuous-action model, not executions selected by the DR scorer. Three models share ten test scenes per objective, giving 30 executions per objective. Both Push objectives satisfy the environment criterion in 30/30 runs, but mean

endpoint errors remain 9.06 and 9.02 cm under a 10 cm tolerance. Both Pick objectives score 0/30 success, including zero sustained holding. The four displayed examples use the same fixed selection rule, without outcome filtering; this subsection gives the protocol and full results. Frozen readouts favor Q-PhysJEPA on five of six metrics, with unmatched state-head tuning (Appendix B.1).

Paired failure examples. For seed 101, the best candidate in OOD rapid-completion episode 00084 is 2, with utility 0.89168. No-query selects 2, whereas Q-PhysJEPA selects 1 with utility 0.85655. In episode 00002, Q-PhysJEPA selects best candidate 2 with utility 0.87919, whereas no-query selects 1 with utility 0.83052. All four selections meet the recorded environment-success criterion, even though their utilities differ. Candidate indices are zero-based. We select the lowest-ID episode for each direction of disagreement, rather than choosing examples by effect size. These cases illustrate selection errors but do not estimate how often particular failure mechanisms occur.

Scope. The candidate-selection and readout studies do not establish online exploration, arbitrary-language query generalization, or physical-robot transfer.