

BEYOND MULTIMODAL ALIGNMENT: CERTIFYING PHYSICAL LANGUAGE THROUGH RESPONSE SUBSTITUTION AND ORDERED EXECUTION

Kaizhen Tan^{1,2}, Xin Xu², Siru Tao², Yixiao Li², Hanzhe Hong², Yang Feng³, Heqing Du³

¹New York University, New York, NY, USA

²Carnegie Mellon University, Pittsburgh, PA, USA

³Columbia University, New York, NY, USA

ABSTRACT

World models increasingly treat compact multimodal representations as interfaces between perception and physical interaction, yet existing probes do not establish whether information acquired through different sensors carries the same executable meaning, or whether it survives a new action composition. We introduce an operational capability hierarchy and the Disjoint-Bridge Operator-Substitution Certificate (DBOSC), which asks whether independently trained modality compilers enter a frozen response chart interchangeably on evidence outside their training panels. On Cluster Haptic, audio and acceleration representations of the same unseen surface are $4.5\times$ closer in response space than wrong-surface pairings, and the gap holds for all 19 held-out surfaces; unsealing the withheld responses afterwards confirms that every branch also predicts the physics better than the population chart. We then test ordered execution in a controlled elastoplastic system with complementary modality blind spots, where the certificate behaves as an instrument rather than a verdict. At the pre-registered budget its own prerequisite refuses the stack, because the frozen executor cannot advance even an exact chart coordinate through a held-out program. At a converged budget the same rank-three chart executes those programs (oracle NMSE 0.18), fusion improves on both modalities, and 14 of 16 registered checks pass; the two failures share one cause, a diagonal restriction of the fused information matrix doing as well as the full one. Clearing the gate is moreover a property of the executor, not of the chart: an executor emitting whole programs instead of shared per-step dynamics is $38\times$ worse than an entity-blind predictor on the same chart. A matching non-identifiability result explains why compression and fusion alone cannot determine an unseen composition law. Together these results separate attribute access, response substitution, fusion closure, and ordered execution into distinct, separately testable achievements.

1 INTRODUCTION

World models earn a physical interpretation when their internal state can be used, not merely decoded. Recent work has called discrete transition codes a physical language, trained tactile futures to shape action representations, and probed latents for mass, drag, and stiffness (Shang et al., 2026; Jin et al., 2026; Tan et al., 2026). These systems make the same practical bet: a compact representation can become an interface between perception and intervention. The phrase *physical language* adds a semantic commitment. A word inferred from sound should keep its meaning when touch replaces sound, and a phrase assembled from partial evidence should still execute under a new action order.

The usual diagnostics answer narrower questions. Reconstruction tests one decoder. A probe establishes access to an attribute. Alignment places two embeddings nearby. Cross-modal interchangeability comes closest, but it is normally established by training the modalities into a shared space and then reading out agreement in that space; agreement can then follow from the shared objective rather than from the physics. None of these alone says that independently acquired evidence denotes the same observable response function. Cross-modal interchangeability, function representation, and predictive-state learning each supply part of that picture (Zhang et al., 2024; Gondal et al.,

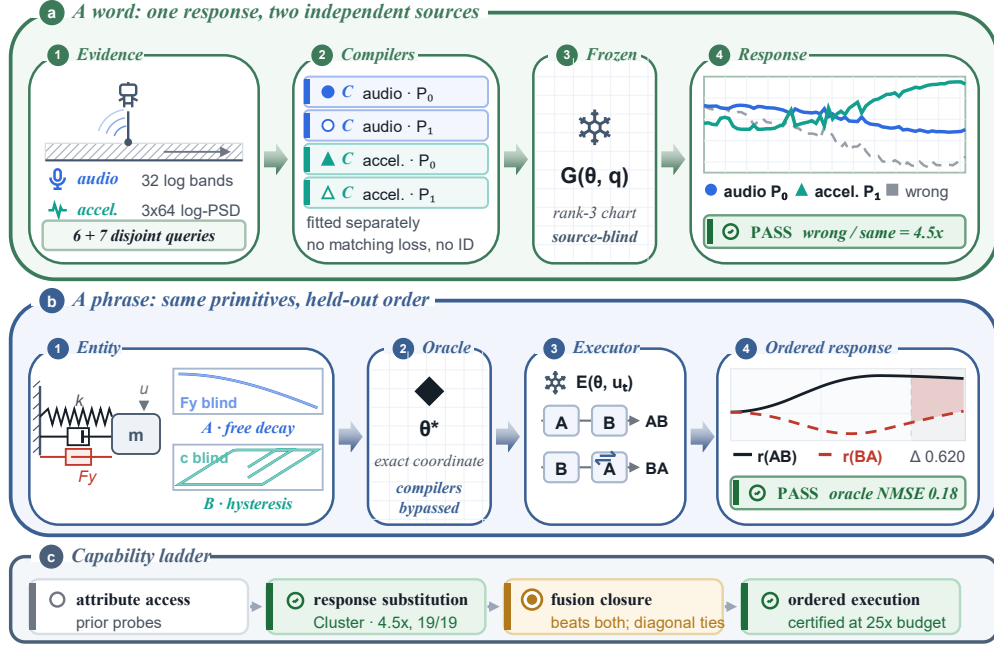


Figure 1: **A word substitutes across sensors; a phrase executes only once its executor does.** (a) One robot scan yields audio and acceleration evidence on disjoint panels P_0, P_1 . Four compilers are fitted independently—no matching loss, no entity identifier—and the frozen decoder $G(\theta, q)$ gives their coordinates a shared denotation. Curves are decoded responses for a held-out surface outside both panels: the two sources agree, a wrong surface does not. (b) A controlled elastoplastic entity with exactly complementary blind spots; each inset draws four hidden levels of the invisible coordinate, and they coincide. The exact θ^* bypasses the compilers and drives a frozen executor on AB and BA ; the reported NMSE is the converged-budget run. (c) The resulting capability profile: substitution certified, ordered execution certified once the executor is adequate, and fusion closure certified except against its own diagonal restriction.

2021; Ingebrand et al., 2024; Littman et al., 2001). We need a test that attaches meaning to what a representation can execute, then asks exactly which capability has been established.

Our answer is operational and query-relative. An entity is represented by its responses to a registered family of interventions. A frozen executor gives modality coordinates a shared denotation. DBOSC then asks whether independent compilers, trained on disjoint evidence panels and without cross-modal matching, substitute for one another on responses outside those panels. Same-entity agreement must beat a wrong-entity bridge and a population response. This design turns cross-modal meaning into a response-space measurement rather than a geometric analogy.

The two experiments in Figure 1 expose a useful separation. Cluster Haptic provides the positive result: audio and acceleration compile to the same surface-specific response coordinate on unseen surfaces, and unsealing the withheld responses confirms that the shared coordinate is accurate rather than jointly mistaken. An exactly controlled elastoplastic system asks for more. Its two modalities have complementary, provable blind spots; the response profile is compact; and the held-out response depends strongly on action order. There the certificate behaves as an instrument: its own prerequisite refuses a stack whose executor is undertrained, the same stack certifies once that executor converges, and swapping the executor’s factorization stops it again. Together with a non-identifiability construction, these results establish a practical hierarchy: attribute access, response substitution, fusion closure, and ordered execution are distinct achievements. The main contribution is the certificate that tells them apart, a real-apparatus substitution result, and a controlled study of which component supplies ordered execution.

2 OPERATIONAL SEMANTICS

2.1 RESPONSE EQUIVALENCE GIVES A WORD ITS REFERENT

Let Ω be a family of physical entities and $\mathcal{A}^*$ a set of finite intervention programs. An apparatus α fixes the sensor, contact geometry, initial condition, and response preprocessing. The response operator of entity ω is

$$\mathcal{R}_\omega^\alpha(q) := \mathcal{L}(Y_{0:T(q)} \mid \text{do}(q), \omega, \alpha), \quad q \in \mathcal{A}^*, \quad (1)$$

where $\mathcal{L}$ is the response law. Two entities carry the same operational meaning for a registered query family $\mathcal{Q}$ when

$$\omega \equiv_{\alpha, \mathcal{Q}} \omega' \iff \mathcal{R}_\omega^\alpha(q) = \mathcal{R}_{\omega'}^\alpha(q) \quad \forall q \in \mathcal{Q}. \quad (2)$$

The quotient $\Omega / \equiv_{\alpha, \mathcal{Q}}$ is the vocabulary induced by the apparatus and query family. It distinguishes entities through behavior that can be executed and observed. This definition also makes the unit of meaning explicit: not an embedding coordinate, but the response function indexed by the registered interventions.

A modality compiler $C_{m,P}$ maps evidence $e_{m,P}$ from modality m and panel P into a response coordinate θ or a belief over coordinates. A shared executor gives that coordinate a denotation,

$$\llbracket e_{m,P} \rrbracket(q) = G(C_{m,P}(e_{m,P}), q). \quad (3)$$

The executor receives no modality, panel, or entity identifier. Any two compilers that induce the same function $q \mapsto \llbracket e_{m,P} \rrbracket(q)$ are therefore interchangeable for that response family.

2.2 A CAPABILITY LADDER

Figure 1 organizes four increasingly demanding claims. *Attribute access* asks whether a latent carries a physical variable that is recoverable from raw evidence. *Response substitution* asks whether different sensors induce the same entity-specific behavior through one frozen executor. *Fusion closure* asks whether complementary evidence improves the executed predictive law. *Ordered execution* asks whether that law remains valid when familiar primitives appear in a held-out order. Each step changes the empirical object being tested; success at one level supplies no automatic shortcut to the next. Table eftab:capability-logic states what each rung holds fixed.

2.3 WORDS, FUSION, AND PHRASES

We call a representation a physical word when its sensory source can be exchanged without changing what a frozen executor predicts. For a Cluster surface, this word is neither a material name nor an embedding cluster. It denotes the scan-response function induced by the registered apparatus. Audio and acceleration may look unrelated as signals; they share a word when independent compilers recover the same responses on queries outside both evidence panels.

Fusion concerns a different use of the representation. It combines incomplete evidence about one entity inside a shared chart. A successful fusion should improve the response law that the chart executes, not merely narrow a posterior in coordinate space. Our test therefore holds the entity, executor, and intervention fixed while changing only the available evidence. A narrower belief paired with a worse response has achieved statistical concentration, but not fusion closure.

A phrase adds an operation. In the controlled system, A and B are familiar force pulses, while AB and BA are programs assembled from the same primitives. Plastic memory carries the effect of the first pulse into the second, so reversing the order changes the final response. Access to the individual action tokens does not determine this composition rule. The oracle check supplies the exact entity coordinate and asks whether the frozen executor can carry it across the held-out transition.

The holdout follows the unit of meaning. A word-level experiment withholds the entity and asks whether different sensors still select its response function. A phrase-level experiment keeps the primitives familiar but withholds the edge that joins them. Cluster provides real sensory variation and a broad response grid for the first question. The controlled oscillator provides exact blind directions and physical memory for the second. Together they separate two claims that a single latent-space score would conflate.

2.4 WHY THE HIERARCHY IS STRICT

Proposition 1 (Compression and information fusion do not identify composition). *Low-rank response compression on observed queries, strict squared-loss Bayes-risk reduction from each modality, and strict posterior-trace reduction under fusion do not suffice to identify, and hence cannot guarantee accurate execution of, an unseen ordered query.*

A short construction makes the separation concrete. Let the observed primitive responses be $\mathcal{R}(A) = U$ and $\mathcal{R}(B) = V$, with independent Gaussian coordinates U, V , and let the two modalities provide noisy measurements of U and V separately. The observed response chart is exactly two dimensional. Each measurement lowers Bayes risk for an ordered response, and combining them strictly contracts the posterior covariance. Now consider two physical systems that agree on every observed variable but use opposite composition laws,

$$\begin{aligned}\mathcal{R}_\sigma(AB) &= U + V + \sigma\gamma UV, \\ \mathcal{R}_\sigma(BA) &= U + V - \sigma\gamma UV, \quad \sigma \in \{-1, +1\}.\end{aligned}\tag{4}$$

Their evidence and primitive-response distributions are identical, while their commutators have opposite signs. Any learner fitted to the common observed law must therefore return the same ordered prediction in both systems and be wrong in at least one. Compression and fusion have done their advertised jobs; the unseen operation is simply not identified by those jobs. Appendix A supplies the posterior and risk calculation.

3 OPERATIONAL CERTIFICATES

3.1 DISJOINT-BRIDGE OPERATOR-SUBSTITUTION CERTIFICATE

DBOSC separates the construction of a response vocabulary from the test of multimodal meaning. A response-rich training source defines the chart and executor; both are then frozen. Each modality compiler is fitted independently from its own evidence, with neither a matching loss nor an entity identifier. Let P_0 and P_1 be disjoint evidence panels, and let $\mathcal{Q}_{\text{bridge}}$ contain response queries outside both panels. For entity i , the decoded response of branch (m, P) is $\hat{\mathcal{R}}_{m,P}^{(i)} = [G(C_{m,P}(e_{i,m,P}), q)]_{q \in \mathcal{Q}_{\text{bridge}}}$.

We normalize response distance by the variation among training entities,

$$D_{\mathcal{Q}_{\text{bridge}}}(R, R') = \frac{\|R - R'\|_F^2}{Z}, \quad Z = \frac{1}{N_{\text{tr}}} \sum_{j=1}^{N_{\text{tr}}} \|G_{\mathcal{Q}_{\text{bridge}}}(\theta_j) - G_{\mathcal{Q}_{\text{bridge}}}(0)\|_F^2.\tag{5}$$

For a test set of N entities, let $\mathcal{O} = \{(a, P_0; b, P_1), (a, P_1; b, P_0)\}$ denote the two cross-panel orientations. The finite estimator is

$$D_{\text{same}} = \frac{1}{2N} \sum_{i=1}^N \sum_{(u,v) \in \mathcal{O}} D_{\mathcal{Q}_{\text{bridge}}}(\hat{\mathcal{R}}_u^{(i)}, \hat{\mathcal{R}}_v^{(i)}).\tag{6}$$

Let $\mathcal{U}$ contain the four modality-panel branches and write $\bar{R} = G_{\mathcal{Q}_{\text{bridge}}}(0)$. The two controls are

$$\begin{aligned}D_{\text{wrong}} &= \frac{1}{2N(N-1)} \sum_{i \neq j} \sum_{(u,v) \in \mathcal{O}} D_{\mathcal{Q}_{\text{bridge}}}(\hat{\mathcal{R}}_u^{(i)}, \hat{\mathcal{R}}_v^{(j)}), \\ D_{\text{pop}} &= \frac{1}{4N} \sum_i \sum_{u \in \mathcal{U}} D_{\mathcal{Q}_{\text{bridge}}}(\hat{\mathcal{R}}_u^{(i)}, \bar{R}).\end{aligned}\tag{7}$$

Every branch must first execute observed development responses better than the population coordinate. On test entities, the substitution certificate is

$$D_{\text{same}} < \min\{D_{\text{wrong}}, D_{\text{pop}}\}.\tag{8}$$

The wrong-entity comparison tests specificity, the population comparison tests collapse, and the disjoint panels remove shared query instances as an alignment key. Appendix B gives the access contract and the corresponding finite sums.

3.2 FUSION AND ORDERED EXECUTION

For a belief μ over response coordinates, the frozen executor induces the response law $(G_Q)_\# \mu$. We score the complete query, time, and channel vector with the energy score (Gneiting & Raftery, 2007),

$$\text{ES}(F, \mathbf{y}) = \frac{1}{\sqrt{D}} (\mathbb{E}\|X - \mathbf{y}\|_2 - \tfrac{1}{2}\mathbb{E}\|X - X'\|_2), \quad X, X' \stackrel{\text{iid}}{\sim} F. \quad (9)$$

The ordered certificate starts with an oracle check: the exact chart coordinate of a held-out entity must execute new programs better than the population coordinate. This isolates the executor’s capability from evidence inference. Complementary fusion is meaningful only after this instrument works; it must improve both the predictive law and its mean response over either modality, the prior, and matched controls. The final target is the response commutator

$$\Delta_\omega^{A,B} = r_\omega(AB) - r_\omega(BA). \quad (10)$$

where $r_\omega(q)$ is the complete normalized readout trajectory. We evaluate the stack in order: oracle execution, partial beliefs, fusion, and finally the commutator. Each stage inherits the response meaning established by the one before it. Appendix C gives the complete protocol and decision rule.

4 EXPERIMENTAL INSTANTIATIONS

4.1 CLUSTER HAPTIC

Cluster Haptic records synchronized acceleration, audio, force, and position while a controlled three-axis machine scans 118 surfaces under varied direction, velocity, normal-force, and repeat conditions (Eguchi et al., 2026a;b). We split complete surface identities. A small axis set contains a reference scan and every single-coordinate change; the remaining combinations form $\mathcal{Q}_{\text{bridge}}$. Parity divides the axis set into two disjoint panels. A test compiler therefore sees a new physical object through one small panel and must denote its responses throughout the composition grid.

The response is the three-axis log spatial acceleration spectrum. A train-only rank-four response basis compresses each scan; a second, rank-three basis organizes complete surface response profiles. Decoding a surface coordinate through these two fixed maps produces the spectrum for any registered scan query. Four linear-kernel compilers map audio or acceleration from either panel into this chart. They receive neither surface identity nor a cross-modal matching target.

Training responses define the chart, and development responses verify that every compiler predicts new compositions. At test time, the two cross-panel orientations swap which modality sees which probes, so no result belongs to a privileged sensor or panel. The final comparison is among decoded response functions and fixed controls. Signal processing, split counts, and the access contract appear in Appendix B.

4.2 CONTROLLED ORDERED RESPONSE

The second system is a unit-mass elastoplastic oscillator with stiffness k , damping c , yield force F_y , and plastic displacement p :

$$f = \text{clip}(k(x - p), -F_y, F_y), \quad \dot{v} = u - cv - f, \quad \dot{x} = v. \quad (11)$$

One modality observes a small-amplitude free decay. Its initial elastic force remains below the smallest yield force in the entity grid, and mechanical energy decreases thereafter; plastic displacement therefore stays at zero and the entire sequence is exactly invariant to F_y . The other modality observes a fully settled quasistatic hysteresis loop. Velocity is zero at every recorded point, so the damping term disappears and the sequence is exactly invariant to c . Both experiments retain stiffness information. Training, development, and test entities occupy disjoint Cartesian grids.

Three force pulses form finite action words. Every word ends with the same 60-step zero-input readout, and the driven segment is excluded from the score. Fitting words contain all three primitives but only self transitions and the $B \leftrightarrow C$ edge. Development introduces the separate $A \leftrightarrow C$ edge. The held-out test contains AB and BA , and every longer word containing either adjacency remains absent throughout fitting and development. Thus the model has seen A , B , and both positions in

a word; what it has not seen is their transition edge. Reversing that edge changes the normalized readout with RMS 0.620, so the held-out operation is physically consequential rather than a symbolic relabeling.

The coordinates hidden from each modality still control the ordered response. Modality A hides yield force; modality B hides damping; both retain stiffness. The held-out readout therefore requires their complementary information rather than a shared stiffness estimate alone.

A rank-three SVD of fitting responses defines the response chart. A shared two-layer micro-step executor predicts observable dynamics from the current response, scalar action, and chart coordinate. Its oracle test uses the exact chart coordinate from the held-out entity’s fitting-response profile. Only after freezing this instrument do two GRU compilers emit rank-two Gaussian information factors. Natural parameters add under the common prior,

$$\Sigma_S = (I + \sum_{m \in S} \Lambda_m)^{-1}, \quad \mu_S = \Sigma_S \sum_{m \in S} \eta_m, \quad (12)$$

which is the standard Gaussian product-of-experts operation (Hinton, 2002; Wu & Goodman, 2018; Sutter et al., 2021). Population, diagonal, point, direct, and observable wrong-entity controls share the same held-out programs. Appendix C specifies the grids, action words, training schedule, and gate.

5 RESULTS

5.1 CLUSTER CERTIFIES ENTITY-SPECIFIC RESPONSE SUBSTITUTION

Every Cluster compiler predicted repeat 1 composition responses better than the population chart center on development surfaces (Figure 2c). The result holds across both panel assignments and both modalities.

The test comparison is decisive. In response space, the same-surface audio–acceleration bridge is $4.5\times$ closer than a wrong-surface bridge and also beats population substitution (Figure 2d). The aggregate is not carried by a few surfaces: the gap holds for all 19 held-out identities, median ratio $5.2\times$.

A target-free comparison cannot by itself exclude two sensors agreeing on the same wrong response. We therefore unsealed the withheld repeat 1 test responses *after* every chart, compiler and threshold was fixed. All four branches beat the population chart center on responses that entered no fit—acceleration 0.687, audio 0.830, tracking their development values (Table 2)—and bridge distance correlates with true response error across surfaces ($r = 0.63$). A short bridge is evidence of an accurate response, not of a shared mistake. The two independently trained sensors therefore recover the same entity-specific response coordinate on surfaces absent from fitting.

5.2 THE PREREQUISITE REFUSES AN UNTRAINED INSTRUMENT

The controlled system clears the prerequisites for an ordered test. Both blind directions are exactly invariant, every hidden coordinate changes the held-out response, and the true AB/BA commutator is substantial. The rank-three chart retains 98.95% of centered fitting-response energy.

The registered run trains the executor for 1,200 updates, and at that budget the oracle check fails: given the exact chart coordinate the executor is worse than the population response on both development and held-out programs (NMSE 1.22 and 3.52). Because that check is the pipeline’s own prerequisite, nothing downstream of it is evidence about the representation—the instrument that would carry it is untrained. Repeating the identical pipeline, seed and thresholds with the single change of a converged budget resolves it: sweeping the budget over $\{1.2, 3, 6, 12, 30\} \times 10^3$ updates crosses the 0.80 gate between 6,000 and 12,000, and at 30,000 the same chart executes the sealed programs with oracle NMSE 0.076 on development and 0.184 on AB/BA (Figure 3a). The count of registered checks passed rises with it, from 8 of 16 at the registered budget to 14 of 16.

The gate keeps its teeth, because clearing it is a property of the executor rather than of the chart. A *program-level* executor that reads the action word as a token sequence and predicts the whole scored readout is the more accurate of the two on the seventeen fitting words—absolute NMSE 0.0017 against an entity-blind predictor, versus 0.0893—so it is not a weak baseline. Yet it is $3.6\times$ worse

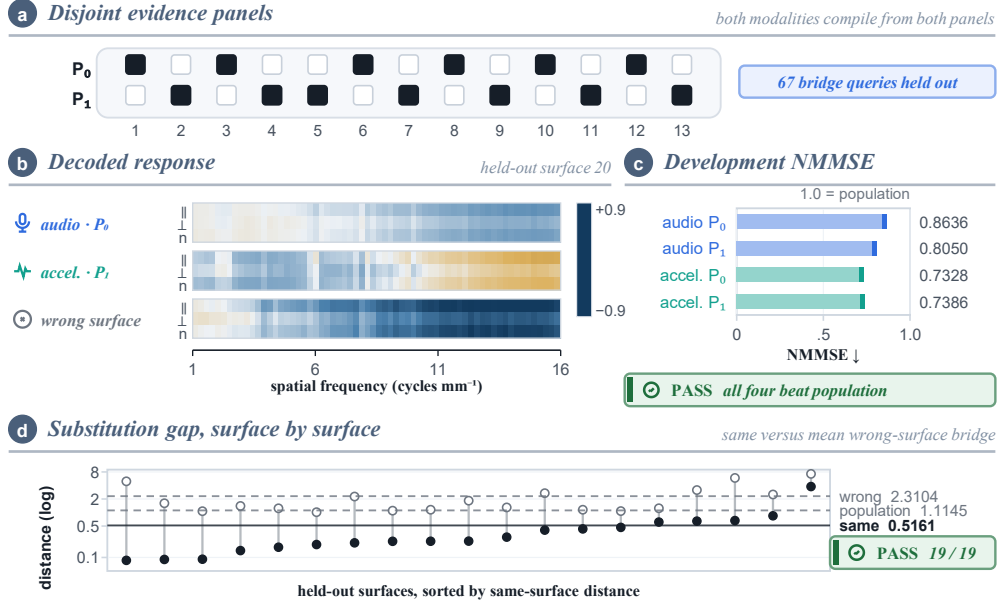


Figure 2: **Two sensors recover the same entity-specific response.** (a) The 13 axis queries split into disjoint panels; the 67 composition queries form $\mathcal{Q}_{\text{bridge}}$. (b) Standardized log spatial-PSD decoded through the frozen chart for one held-out surface: audio from P_0 and acceleration from P_1 agree, the same decoder driven by a wrong surface does not. (c) Development NMMSE per branch, population at 1.0. (d) Each held-out surface sorted by same-surface distance: filled marks are the same-surface bridge, open marks the mean wrong-surface bridge, rules the aggregate estimators of Eqs. 6–7.

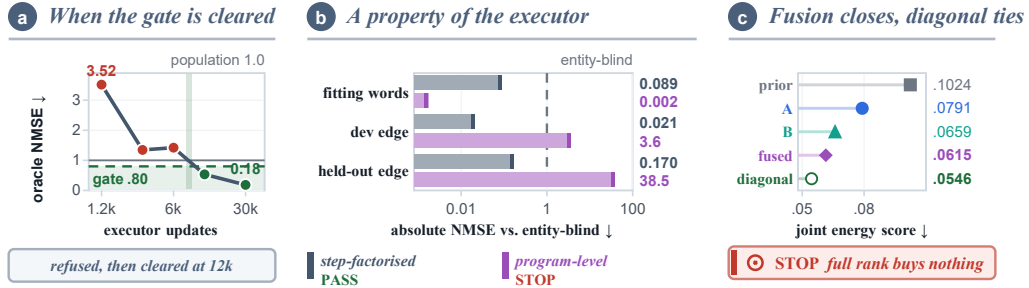


Figure 3: **The prerequisite refuses an untrained instrument; the gate still separates two executors.** (a) Oracle NMSE given the exact chart coordinate, from the registered pipeline at five executor budgets with identical seed and thresholds. At 1,200 updates the prerequisite fails, so nothing downstream is evidence about the representation; the gate is crossed between 6,000 and 12,000 (shaded band). (b) Absolute NMSE against an entity-blind predictor at a matched 30,000-update budget, log scale, three seeds. The program-level executor is the more accurate of the two on the words it was fitted on, yet collapses on any adjacency it was never trained to emit: ordered execution is a property of the executor’s factorization, not of the chart. (c) Held-out joint energy score. Fusion improves on both modalities and on the prior—the closure the certificate asks for—but the diagonal restriction of the same information matrix does slightly better, which is the sole remaining gate failure.

than that predictor on the development $A \leftrightarrow C$ edge and $38.5\times$ worse on the held-out $A \leftrightarrow B$ edge (Figure 3b). Ordered execution requires an executor factorized through machinery the primitives already share; the certificate admits one and stops the other.

5.3 COMPLEMENTARY FUSION CLOSES, WITH ONE EXCEPTION

At the converged budget the belief stack behaves as the semantics predicts. Each unimodal factor improves on the population prior, and their fusion improves on both: joint energy score 0.0615 against 0.0791 and 0.0659, response NMSE 0.319 against 0.549 and 0.407. Posterior trace falls 68% relative to the better single modality and the fused information matrix is full rank ($\lambda_{\min} = 0.587$), while the observable wrong-entity control stays at NMSE 0.840—the gain is entity-specific rather than general smoothing. Table 4 lists every comparison on the same held-out entities and programs.

Fourteen of the sixteen registered checks pass. The two that fail share a single cause: a diagonal restriction of the information matrix does slightly better than the full rank-two matrix, on the joint energy score (0.0546 versus 0.0615) and on the AB/BA commutator (0.174 versus 0.182). Complementary evidence does close inside this chart. What the certificate refuses is the narrower claim that the off-diagonal structure of the fused belief earns its parameters.

5.4 WHAT THE PROFILE LOCALIZES

The fused commutator recovers the held-out order effect far better than every entity-breaking control—0.182 against 0.374 for the population, 0.470 for the direct diagnostic and 1.000 for a zero-order prediction—so the ordered readout is carried by the entity coordinate and not by the action word alone.

Cluster and the controlled system tell one story. The real apparatus certifies that two modalities share an entity-specific response meaning, and unsealing confirms that meaning is accurate rather than jointly mistaken. The controlled system shows what *using* it under a new action order costs: an executor a new phrase can reuse, and an order of magnitude more budget. The certificate localizes which component supplies which capability instead of returning one verdict.

6 RELATED WORK

Recent world models make their physical representations unusually explicit. PhiZero reasons through discrete transition codes, latent-acquisition studies identify which mechanics predictive states expose, and TacWAM uses tactile futures to improve action learning (Shang et al., 2026; Tan et al., 2026; Jin et al., 2026). Predictive state representations and causal states define state through future predictions (Littman et al., 2001; Downey et al., 2017; Shalizi & Crutchfield, 2001), while bisimulation defines behavioral equivalence through matched rewards and transition laws (Ferns et al., 2004). Function Encoders, DeepONet, neural processes, and neural-operator discovery turn finite input-output contexts into reusable function coordinates (Ingebrand et al., 2024; Lu et al., 2021; Garnelo et al., 2018; Chen et al., 2026). We use the same behavioral foundation and make its empirical burden explicit across sensors and intervention families.

Cross-modal interchangeability is the closest methodological thread. Connect–Collapse–Corrupt improves embedding interchangeability for cross-modal tasks trained with uni-modal data; function-contrastive learning aligns disjoint views of the same function; partial-view causal representation learning characterizes which latent content remains identifiable under partial observability (Zhang et al., 2024; Gondal et al., 2021; Yao et al., 2024). Sheaf-based sensor integration formalizes local compatibility and global consistency (Robinson, 2017). DBOSC contributes a complementary measurement protocol, and the differences are the reason it can be read as evidence about physics rather than about optimization. No compiler ever sees the other modality, a matching loss, or an entity identifier, so agreement cannot be inherited from a shared objective. The evidence panels are disjoint, so a shared query instance cannot act as an alignment key. Agreement is measured in the space of executed responses to registered interventions rather than in embedding geometry, and it is scored on queries outside both panels against wrong-entity and population controls that break entity identity and collapse respectively.

Products of experts provide the combination rule (Hinton, 2002); multimodal VAEs use it to combine modality-specific evidence and support missing modalities (Wu & Goodman, 2018; Sutter et al., 2021). Hidden-parameter state-space models instead aggregate variable-sized interaction contexts into uncertain task beliefs (Shaj et al., 2022). These fusion rules are well understood. Our controlled

test targets the semantic step that comes afterward: whether a concentrated fused belief still executes the response family and action order attached to its chart.

7 DISCUSSION

The Cluster result changes what cross-modal alignment means. Audio and acceleration differ in units, sampling, and nuisance structure; their common content becomes visible only after execution. What they share is a coordinate for how one surface responds throughout the scanning program. The frozen chart supplies a concrete referent even when the encoder geometries themselves need not align.

The controlled result isolates the executor rather than the chart. A rank-three terminal-response coordinate supports an unfamiliar action composition only when executor parameters are shared across primitives and training extends beyond the budget that certifies substitution. A program-specific executor collapses on unseen adjacencies, including one exposed by development, despite higher accuracy on seen programs. Reconstruction quality and chart rank therefore say little about ordered execution. Compositional phrases require an entity-bearing chart and a factorized executor that reuses machinery fixed by familiar phrases.

The belief experiment makes the same point probabilistic. Gaussian information addition contracts uncertainty by design, but whether that contraction is worth anything is only visible after execution: at the registered budget the contracted belief executed worse than either modality, and at a converged budget it executes better than both. Posterior concentration is therefore not evidence of fusion closure on its own, and a certificate that scores beliefs in coordinate space would have reported the same contraction in both cases.

A response word need not identify an object uniquely: two surfaces share a word whenever the registered scans cannot distinguish their response functions, and one partial observation may support several. The honest representation is a belief over response classes, not a forced point label. The language analogy then has a precise division of labor—the executor defines the dictionary, a modality supplies a belief over entries, fusion updates it, and an action sequence supplies the grammar.

Acquisition and certification remain separate: a compiled coordinate gains physical meaning only by surviving a source swap and supporting new phrases through a shared executor.

The intervention set is therefore part of representation design. Entity holdouts ask whether a sensor recovers a word for a new object; edge holdouts ask whether familiar primitives obey the same grammar in a new order. Those choices are more informative than a larger dataset that exposes no new distinction.

Limitations. The positive certificate rests on one apparatus and one modality pair: audio and acceleration on 19 held-out Cluster surfaces. The protocol needs only a registered intervention family and a frozen decoder, but the evidence is rig-specific, and a second contact geometry would separate a property of the method from a property of this apparatus. The ordered-execution study buys provable blind spots at the cost of realism, and the budget at which the gate is crossed describes this executor rather than a constant. What transfers is treating the oracle check as a prerequisite rather than a result.

8 CONCLUSION

We introduced an operational certificate that gives multimodal physical meaning a response-space test. Cluster Haptic establishes source-blind audio–acceleration substitution on unseen surfaces: two independently trained sensors enter one frozen chart, recover the same entity-specific responses outside their evidence panels, and—once those responses are unsealed—predict them better than the population chart. The controlled system separates that achievement from ordered execution and locates the difference precisely. It is not in the chart, which does support the held-out transition edge once its executor is adequately trained, but in that executor’s factorization and in the budget it requires—and the registered gate refused to certify anything while the instrument was untrained. A physical language therefore knows the response distinctions its words preserve across sources, and the phrases whose execution reuses machinery the primitives already fixed; DBOSC measures the former and the ordered gate the latter.

REFERENCES

- Zituo Chen, Qiaofeng Li, Jiaxin Hu, and Sili Deng. Neural operator discovery from heterogeneous trajectories. *arXiv preprint arXiv:2607.23337*, 2026. doi: 10.48550/arXiv.2607.23337. URL <https://arxiv.org/abs/2607.23337>.
- Carlton Downey, Ahmed Hefny, Boyue Li, Byron Boots, and Geoffrey J. Gordon. Predictive state recurrent neural networks. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6053–6064. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/2bb0502c80b7432eee4c5847a5fd077b-Paper.pdf.
- Michikuni Eguchi, Tomohiro Hayase, Yuichi Hiroi, and Takefumi Hiraki. Cluster haptic texture dataset: Haptic texture dataset with varied velocity-direction sliding contacts. *Scientific Data*, 13(1):756, 2026a. doi: 10.1038/s41597-026-06760-z. URL <https://doi.org/10.1038/s41597-026-06760-z>.
- Michikuni Eguchi, Tomohiro Hayase, Yuichi Hiroi, and Takefumi Hiraki. Cluster haptic texture dataset: Haptic texture dataset with varied velocity-direction sliding contacts. <https://doi.org/10.6084/m9.figshare.29438288.v5>, 2026b. Dataset, version 5.
- Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite markov decision processes. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, UAI '04, pp. 162–169, Arlington, Virginia, USA, 2004. AUAI Press. ISBN 0-9749039-0-6. URL <https://dl.acm.org/doi/10.5555/1036843.1036863>.
- Marta Garnelo, Jonathan Schwarz, Dan Rosenbaum, Fabio Viola, Danilo J. Rezende, S. M. Ali Eslami, and Yee Whye Teh. Neural processes. *arXiv preprint arXiv:1807.01622*, 2018. doi: 10.48550/arXiv.1807.01622. URL <https://arxiv.org/abs/1807.01622>.
- Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437. URL <https://doi.org/10.1198/016214506000001437>.
- Muhammad Waleed Gondal, Shruti Joshi, Nasim Rahaman, Stefan Bauer, Manuel Wüthrich, and Bernhard Schölkopf. Function contrastive learning of transferable meta-representations. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3755–3765. PMLR, 2021. URL <https://proceedings.mlr.press/v139/gondal21a.html>.
- Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002. doi: 10.1162/089976602760128018. URL <https://doi.org/10.1162/089976602760128018>.
- Tyler Ingebrand, Amy Zhang, and Ufuk Topcu. Zero-shot reinforcement learning via function encoders. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 21007–21019. PMLR, 2024. URL <https://proceedings.mlr.press/v235/ingebrand24a.html>.
- Lei Jin, Yiding Ma, Xin Zhang, Chen Gao, Wei Wu, and Yong Li. TacWAM: Anchor-guided world action model with mechanics-aware tactile prediction. *arXiv preprint arXiv:2607.28391*, 2026. doi: 10.48550/arXiv.2607.28391. URL <https://arxiv.org/abs/2607.28391>.
- Michael L. Littman, Richard S. Sutton, and Satinder Singh. Predictive representations of state. In *Advances in Neural Information Processing Systems*, volume 14, pp. 1555–1561. MIT Press, 2001. URL https://proceedings.neurips.cc/paper_files/paper/2001/file/1e4d36177d71bbb3558e43af9577d70e-Paper.pdf.
- Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5. URL <https://doi.org/10.1038/s42256-021-00302-5>.

- Michael Robinson. Sheaves are the canonical data structure for sensor integration. *Information Fusion*, 36:208–224, 2017. doi: 10.1016/j.inffus.2016.12.002. URL <https://doi.org/10.1016/j.inffus.2016.12.002>.
- Vaisakh Shaj, Dieter Büchler, Rohit Sonker, Philipp Becker, and Gerhard Neumann. Hidden parameter recurrent state space models for changing dynamics scenarios. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=ds8yZOUsea>.
- Cosma Rohilla Shalizi and James P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics*, 104(3–4):817–879, 2001. doi: 10.1023/A:1010388907793. URL <https://arxiv.org/abs/cond-mat/9907176>.
- Shuyao Shang, Yuqi Wang, Ruopeng Gao, Xu Chen, Tieniu Tan, Lue Fan, and Zhaoxiang Zhang. PhiZero: A world model built around physical language. *arXiv preprint arXiv:2607.28624*, 2026. doi: 10.48550/arXiv.2607.28624. URL <https://arxiv.org/abs/2607.28624>.
- Thomas M. Sutter, Imant Daunhawer, and Julia E. Vogt. Generalized multimodal ELBO. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=5Y21V0RDBV>.
- 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. doi: 10.48550/arXiv.2607.27017. URL <https://arxiv.org/abs/2607.27017>.
- Mike Wu and Noah D. Goodman. Multimodal generative models for scalable weakly-supervised learning. In *Advances in Neural Information Processing Systems*, volume 31, pp. 5580–5590, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/1102a326d5f7c9e04fc3c89d0ede88c9-Abstract.html.
- Dingling Yao, Danru Xu, Sébastien Lachapelle, Sara Magliacane, Perouz Taslakian, Georg Martius, Julius von Kügelgen, and Francesco Locatello. Multi-view causal representation learning with partial observability. In *International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=OGtnhKQJms>.
- Yuhui Zhang, Elaine Sui, and Serena Yeung-Levy. Connect, collapse, corrupt: Learning cross-modal tasks with uni-modal data. In *International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=ttXg3SKAg5>.

A PROOF OF PROPOSITION 1

Let $U, V \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ and let the two modalities observe

$$E_1 = U + \epsilon_1, \quad E_2 = V + \epsilon_2, \quad \epsilon_1, \epsilon_2 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \tau^2). \quad (13)$$

For observed primitive queries, set $\mathcal{R}(A) = U$ and $\mathcal{R}(B) = V$. The response profile lies exactly in the rank-two chart $Z = (U, V)$. Writing $\alpha = \tau^2/(1 + \tau^2)$, the unimodal and fused posterior covariances are

$$\Sigma_1 = \text{diag}(\alpha, 1), \quad \Sigma_2 = \text{diag}(1, \alpha), \quad \Sigma_{12} = \alpha I. \quad (14)$$

Thus $\text{tr } \Sigma_{12} = 2\alpha < 1 + \alpha = \text{tr } \Sigma_1 = \text{tr } \Sigma_2$.

Now index two systems by an unobserved composition law $\sigma \in \{-1, +1\}$:

$$\mathcal{R}_\sigma(AB) = U + V + \sigma\gamma UV, \quad (15)$$

$$\mathcal{R}_\sigma(BA) = U + V - \sigma\gamma UV, \quad \gamma > 0. \quad (16)$$

The systems induce the same joint distribution over all modality evidence and observed-query responses. Both ordered responses have population mean zero. Under squared loss, each modality strictly reduces Bayes risk for either ordered response because, for example,

$$\mathbb{E}[\mathcal{R}_\sigma(AB) \mid E_1] = \frac{E_1}{1 + \tau^2}, \quad (17)$$

whose variance $1/(1 + \tau^2)$ is the Bayes-risk reduction relative to the zero population predictor.

The commutator is $C_\sigma = 2\sigma\gamma UV$. Any learner using only the common observed law produces the same predictor $\hat{C}$ under both signs. Pointwise,

$$\frac{(\hat{C} - 2\gamma UV)^2 + (\hat{C} + 2\gamma UV)^2}{2} = \hat{C}^2 + 4\gamma^2 U^2 V^2. \quad (18)$$

Taking expectations and using $\mathbb{E}[U^2 V^2] = 1$ shows that at least one indistinguishable system has commutator MSE at least $4\gamma^2$. Exact low-rank compression, useful marginal evidence, and reduced fused uncertainty therefore leave the unseen composition law unidentified. $\square$

Table 1: **A capability ladder, not a single score.** Each rung changes exactly one element of physical-language use while the response vocabulary stays fixed. ◻ PRIORITY marks a claim inherited from prior probing work; ✓ PASS and ✗ OPEN report what the certificates in this paper establish.

Capability	What changes	Executable question	This work
Attribute access	Physical distinction	Can one observation recover it?	◻ PRIORITY
Response substitution	Sensor and evidence panel	Do both views denote the same response function?	✓ PASS
Fusion closure	Available evidence	Does the joint predictive law improve?	✗ OPEN
Ordered execution	Primitive order and transition edge	Does the phrase preserve the intervention law?	✗ OPEN

B CLUSTER PROTOCOL AND ACCESS CONTRACT

We use Cluster Haptic v5 with a category-balanced split of 80 training, 19 development, and 19 test surface identities. All response normalizers, chart bases, and compiler statistics use the training identities.

Response construction. We rotate acceleration into path-parallel, path-perpendicular, and surface-normal axes. The central two-thirds of the measured scan path is resampled at 5 kHz. Welch spectra use 20-mm Hann windows with 10-mm hops in spatial coordinates. Each axis is interpolated to 64 frequencies from 1 to 16 cycles/mm, producing a 3×64 log spatial-PSD response. Audio evidence contains 32 log-band powers from 100 Hz to 18 kHz.

Frozen chart and compilers. A rank-four basis compresses each standardized 192-dimensional acceleration response. Concatenating coefficients across all 80 queries gives a 320-dimensional surface profile; its centered, fixed-sign rank-three SVD defines $\theta \in \mathbb{R}^3$. For branch (m, P) , the compiler uses standardized evidence dimension $d_{m,P}$ and the fixed kernel

$$K_{m,P}(x, x') = 1 + \frac{x^\top x'}{d_{m,P}}, \quad \hat{\theta} = K_{x,T}(K_{T,T} + 10^{-3}I)^{-1}\Theta_T. \quad (19)$$

The panel order is registered, but no surface identifier or material category enters a compiler.

Access contract. The safe cache contains repeat 0 axis evidence for all surfaces, train repeat 1 responses for all queries, and development repeat 1 responses for the 67 composition queries. Test repeat 1 responses are absent. P_0 and P_1 are disjoint and cover the 13 axis queries. Same-entity terms pair audio P_0 with acceleration P_1 and audio P_1 with acceleration P_0 . The full estimator is

$$D_{\text{same}} = \frac{1}{2N} \sum_{i=1}^N [D_{\mathcal{Q}_{\text{bridge}}}(\hat{\mathcal{R}}_{a,P_0}^{(i)}, \hat{\mathcal{R}}_{b,P_1}^{(i)}) + D_{\mathcal{Q}_{\text{bridge}}}(\hat{\mathcal{R}}_{a,P_1}^{(i)}, \hat{\mathcal{R}}_{b,P_0}^{(i)})]. \quad (20)$$

With 19 test surfaces, the certificate averages 38 same-entity terms, all 684 ordered wrong-entity terms, and 76 symmetric population terms.

Artifact integrity. The artifact records the split, query panels, protocol, normalizers, model states, and cache schema, and checks their hashes before evaluation. The safe cache excludes Cluster test responses, and ordered responses are materialized only after every component and threshold is fixed. The held-out responses in Table 2 were extracted from the raw archive by the same feature pipeline afterwards, purely to validate the target-free measure; no chart, compiler, threshold or panel assignment was revisited once they were read, and every branch is reported rather than a selected one.

Table 2: **The substitution certificate in full.** Development NMMSE evaluates repeat 1 composition responses against the frozen chart. The middle block scores the same four branches against the withheld test responses, materialized only after every component and threshold was fixed; all four beat the population chart center on data that entered no fit. The last block is the target-free distance the certificate itself uses. Bars are scaled within each block. The certificate of Eq. 8 holds: **same surface** is closest, and it also wins on 19 of 19 individual surfaces (Figure 2d).

Branch or control		Value ↓
DEVELOPMENT NMMSE		
Acceleration P_0		0.7328
Acceleration P_1		0.7386
Audio P_0		0.8636
Audio P_1		0.8050
Population		1.0000
HELD-OUT TEST NMMSE (unsealed after freezing)		
Acceleration P_0		0.6873
Acceleration P_1		0.6870
Audio P_0		0.8602
Audio P_1		0.8298
Population		1.0000
HELD-OUT TEST DISTANCE		
Same surface		0.5161
Population substitution		1.1145
Wrong surface		2.3104

C ORDERED RESPONSE QUOTIENT DETAILS

C.1 ENTITY AND QUERY SPLITS

The simulator uses $\Delta t = 0.005$. Training entities form a 5^3 grid over $k \in [1.2, 2.4]$, $c \in [0.12, 0.52]$, and $F_y \in [0.18, 0.42]$. Development and test entities occupy disjoint interleaved 4^3 grids. Modality A records 200 free-decay steps from $x_0 = 0.035$; its maximum elastic force remains below the minimum yield force. Modality B records a 90-point fully settled quasistatic hysteresis path.

Primitive pulses have amplitudes $u_A = 4.0$, $u_B = -2.8$, and $u_C = 1.4$, each with 70 driven and 30 coast steps. The word split is

$$\begin{aligned}
 \mathcal{Q}_{\text{fit}} &= \{A, B, C, AA, BB, BC, CB, CC, AAA, BBB, BBC, BCB, \\
 &\quad BCC, CBB, CBC, CCB, CCC\}, \\
 \mathcal{Q}_{\text{dev}} &= \{AC, CA, AAC, ACA, ACB, ACC, BCA, CAA, CAC, CCA\}, \\
 \mathcal{Q}_{\text{test}} &= \{AB, BA\}.
 \end{aligned}$$

Length-three words containing an AB or BA adjacency remain held out throughout fitting and development.

C.2 EXACT BLINDNESS AND RESPONSE RELEVANCE

For modality A, the initial displacement satisfies $\max_k k|x_0| < \min F_y$. Free decay can only reduce mechanical energy, so the trajectory remains elastic and $p = 0$ throughout. Its complete evidence

sequence is therefore invariant to yield force. For modality B, every hysteresis point is evaluated after the system has fully settled. Velocity is exactly zero, the damping term vanishes, and the sequence is invariant to c .

The sealed response remains sensitive to every hidden coordinate. Adjacent-level changes in stiffness, damping, and yield force produce normalized response RMS values 0.09509, 0.03925, and 0.09041. The AB/BA commutator RMS is 0.62016. Wrong-entity fusion is matched within stiffness and replaces evidence only when at least one complementary coordinate changes. Thus neither the blindness checks nor the wrong-entity control can be passed by a stiffness-only representation.

C.3 RESPONSE CHART, EXECUTOR, AND COMPILERS

Each fitting entity contributes a flattened normalized profile of size $17 \times 60 \times 3 = 3060$. The chart is centered before SVD; each component sign is fixed by requiring the largest-magnitude loading to be positive; scores are divided by their train standard deviations. The relative Frobenius residual is

$$\sqrt{\frac{\|R - \hat{R}\|_F^2}{\|R - \bar{R}\|_F^2}} = 0.102265, \quad (21)$$

equivalent to unexplained energy 0.0104582.

The executor receives three normalized observables, one normalized action, and three chart coordinates. Its MLP widths are $7 \rightarrow 128 \rightarrow 128 \rightarrow 3$, with SiLU activations. AdamW uses learning rate 10^{-3} , weight decay 10^{-5} , and batch size 128. The development oracle selects between the fixed 600- and 1,200-update checkpoints before sealed responses are materialized.

Each modality compiler is a 32-dimensional GRU with a rank-two Gaussian information head. The frozen executor supplies its only training target. A point compiler uses the same response objective with a singleton law; the diagonal compiler restricts the information matrix; the direct diagnostic predicts the readout without the chart.

C.4 JOINT FINITE-DISTRIBUTION SCORE

For posterior $\mathcal{N}(\mu, \Sigma)$, a Cholesky factor maps the eight lexicographically ordered nodes in $\{-1, +1\}^3$ to $z_s = \mu + L_\Sigma s$, each with weight $1/8$. Executing each node on AB and BA produces a 360-dimensional response vector. The implementation uses the finite-distribution energy score

$$\sum_s w_s \frac{\|\hat{r}_s - r\|_2}{\sqrt{360}} - \frac{1}{2} \sum_{s,t} w_s w_t \frac{\|\hat{r}_s - \hat{r}_t\|_2}{\sqrt{360}}. \quad (22)$$

This is a proper score for the registered finite predictive distribution.

C.5 REGISTERED GATE AND COMPLETE OUTCOME

Table 3: **Where the ordered certificate stops at a converged budget.** Thresholds were fixed before the sealed responses were materialized and are unchanged here; only the executor and compiler budget differs from the registered run. Every absolute check now passes, including the oracle prerequisite that stopped the 1,200-update run at 1.2238/3.5175. The two remaining × **STOP**s have one cause: the diagonal restriction of the fused information matrix matches or beats the full rank-two matrix.

Check	Requirement	Observed outcome
ABSOLUTE CHECKS		
Blind-direction error	$\leq 10^{-12}$	✓ PASS 0.0
Each factor-effect RMS	$> 10^{-4}$	✓ PASS
Chart relative residual	< 0.20	✓ PASS 0.1023
Oracle NMSE, dev / test	< 0.80	✓ PASS 0.0756 / 0.1844
Fused information $\lambda_{\min}$	$> 10^{-5}$	✓ PASS 0.5867
True commutator RMS	$> 10^{-3}$	✓ PASS 0.6202
COMPARISONS		
Fusion beats both singles	ES and NMSE	✓ PASS
Fusion beats wrong entity	ES and NMSE	✓ PASS
Fused ES beats diagonal	strict	× STOP 0.0615 / 0.0546
Fused Δ beats diagonal	strict	× STOP 0.1816 / 0.1740

The full conjunction also requires both partial beliefs to improve over the prior; fusion to improve over both partial beliefs, population, wrong-entity, point, and diagonal controls; fusion NMSE to improve over both partial NMSEs; posterior trace to decrease; and the fused commutator to improve over zero, population, point, and diagonal predictions. At the registered 1,200-update budget, oracle execution, fused response improvement, and commutator superiority all stop. At the converged budget, 14 of the 16 checks pass; the exceptions are the two strict comparisons against the diagonal control, so the status remains `formal_gate_failed` on a far narrower ground.

Table 4: **Complete held-out response comparison** on unseen entity values and $\{AB, BA\}$, at the converged 30,000-update budget. Fusion improves on both partial beliefs in every column; the **diagonal restriction** nonetheless matches or beats it, which is the sole remaining gate failure. Bars run to NMSE 1.05; the direct diagnostic (44.9) is clipped, marked $\gg$. Lower is better.

Representation or control	Joint ES		Commutator	
	↓	NMSE ↓		↓
BASELINE				
Population prior	0.10239	1.0000		0.37388
CHART BELIEFS				
Rank-2 factor A	0.07909	0.5492		—
Rank-2 factor B	0.06594	0.4066		—
Rank-2 fused	0.06147	0.3193		0.18165
REGISTERED CONTROLS				
Diagonal fused	0.05456	0.2873		0.17396
Point fused	0.08484	0.3736		0.20158
Observable wrong fusion	0.10217	0.8398		—
Direct diagnostic	0.96737	44.873	$\gg$	0.47015
Zero order	—	—		1.00000