COMPUTE, FUSE, OR ACQUIRE? TEST-TIME DECISION-INFORMATION BOTTLENECKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Test-time scaling is usually treated as allocating more computation, or more interaction with the world. We show that this conflates three distinct bottlenecks. Computational regret is imperfect inference over the current interface; representation regret is evidence the model has already observed but its interface discards; observation regret is evidence never observed. The algebra separating them is elementary, but the consequence is not: computational, representation and observation regret require distinct remedies, namely compute, fusion and acquisition respectively. This yields a prediction a compute-versus-interaction view cannot state, namely that performance can improve at exactly fixed evidence. We test the predictions on InfoGap, a benchmark whose floors are designed and, for vision, tunable. Across three VLMs, varying the designed single-view Bayes accuracy moves the observed sampling plateau with it, separating an information ceiling from a capability limit; fixed-view sampling stays at that ceiling while one relevant view lifts accuracy from 0.49 to 0.99. Holding the observed evidence identical and changing only how two channels are combined moves accuracy from 0.50 to 1.00, which acquisition cannot explain. A bottleneck-matching router is a proof of concept rather than a transferable diagnostic: it reaches 0.94 within a known task mixture at half the generated-token cost of Best-of-16, but leave-one-category-out shows it identifies task families, not instances.

1 INTRODUCTION

Test-time scaling improves large models without retraining: sample and vote (Wei et al., 2022; Wang et al., 2023b), rank with a verifier (Cobbe et al., 2021; Lightman et al., 2024), reason longer (Yao et al., 2023a; Muennighoff et al., 2025; DeepSeek-AI et al., 2025), allocate inference compute (Snell et al., 2024; Wu et al., 2024), or repeat until success (Brown et al., 2024), a recipe now migrating to vision-language-action policies that sample and select actions (Kwok et al., 2025; Nakamoto et al., 2024; Dai et al., 2025; Ye et al., 2026). In all of them the scaled quantity is *computation over a fixed information state*: the representation Z of observation and goal is held fixed while inference effort on top of Z grows.

If the information set has already lost what the decision needs, no amount of computation over it can bring it back: an agent whose view makes a successful grasp indistinguishable from a gripper closed behind the object can sample thousands of candidates, but every one inherits the same blindness, and their agreement is confidence in an ambiguity rather than evidence about the world. We make this exact. For any test-time algorithm δ_K conditioning on an interface Z of the observed evidence X_{base}, the excess risk relative to a decision-maker that sees the true state S decomposes (exactly, not as a bound) into three terms (Proposition 1):

$$\underbrace{R_\ell(\delta_K) - R^*(S,G)}_{\text{total regret}} = \underbrace{\big[R_\ell(\delta_K) - R^*(Z,G)\big]}_{\text{computational}} + \underbrace{\big[R^*(Z,G) - R^*(X_{\text{base}},G)\big]}_{\text{representation}} + \underbrace{\big[R^*(X_{\text{base}},G) - R^*(S,G)\big]}_{\text{observation}}. \tag{1}$$

The first term is what more inference over the same interface may fix; the second is evidence the agent *observed* but its interface discards (recoverable by fusion, without touching the world); the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

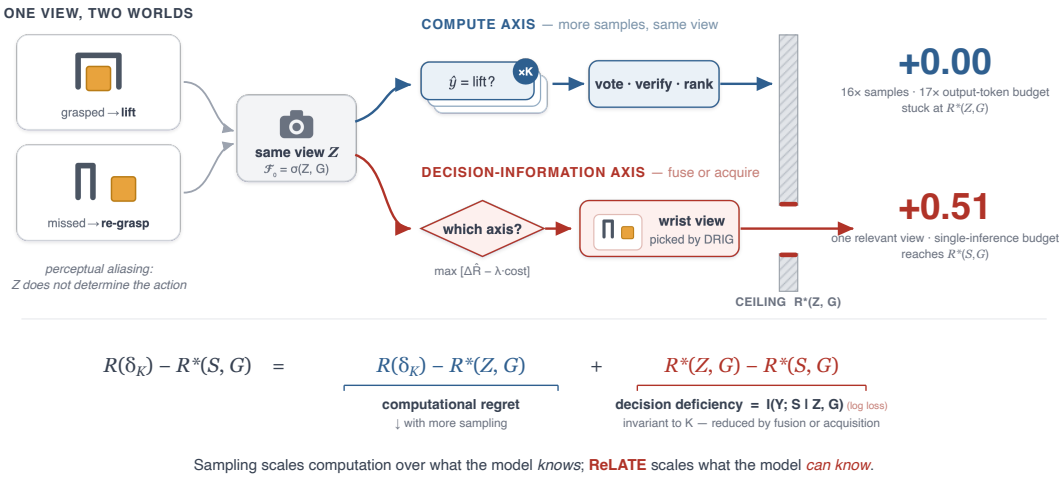


Figure 1: **Three failure modes, three remedies.** *Top left:* perceptual aliasing, two latent states (grasped vs. closed behind the object) render to the same view Z, but demand different actions; the agent's information set does not determine the decision. *Blue path (compute axis):* sampling K candidates, voting, verifying, or re-ranking all condition on the same interface (every candidate inherits the same blindness), so risk is floored at the fixed-representation Bayes risk $R^*(Z, G)$: on InfoGap-Vision, $16\times$ samples ($17\times$ the generated-token budget) buy $+0.00$ accuracy. *Red path (decision-information axis; ReLATE action space, with* fuse *validated separately rather than as a routing arm):* a controller diagnoses whether the binding constraint is compute or information, and if the latter, applies a decision-information operation: *fusion* of already-observed channels (recovering the representation component), or *acquisition* of the observation with the highest *decision-relevant* information gain (DRIG), here a wrist view, expanding the information set to $\mathcal{F}_1 = \mathcal{F}_0 \vee \sigma(O)$ and lowering the floor toward $R^*(S, G)$: one relevant view at a *smaller* token budget buys $+0.51$ accuracy. *Bottom:* the decomposition underlying the picture, displayed in its two-axis grouping (the representation and observation terms of Proposition 1 combined into the red decision-deficiency term, which equals $I(Y; S \mid Z, G)$ under log loss and is invariant to K (measured flat to 10^{-9} on exact POMDPs), and is reduced by fusion of its representation component and acquisition of its observation component, never by more sampling). Quantitative results: Figures 3–5.

third is evidence it has *never* observed (addressable only by acquisition, and reducible only if an informative sensor exists). Three terms, three *different* remedies (sample/verify, fuse, acquire), and this is the structure the paper exploits: grouped one level up, computational regret is the *computation axis* while the representation and observation terms together form the *decision-information axis* (the decision deficiency), reduced by fusion and acquisition respectively. The middle term is what separates this from an internal-versus-external or thinking-versus-interaction dichotomy: evidence the agent already holds can be recovered without acting on the world at all. The identity itself is elementary; our theoretical contribution is a characterization of the deficiency terms (Theorem 1): their sum vanishes for *every* loss exactly when Z is posterior-sufficient for the task, and across loss families it is controlled by a single conditional mutual information $I(Y; S \mid Z, G)$, which we call the *task-conditional information ceiling*, itself splitting by the chain rule into a representation part $I(Y; X_{\text{base}} \mid Z, G)$ and an observation part $I(Y; S \mid X_{\text{base}}, G)$. A no-free-information principle follows (Proposition 2): any statistic computed from candidates that were all generated from the same information set carries zero additional Shannon information about the decision. Sampling more from Z is not the same as acquiring information beyond Z.

What *can* lower the ceiling is evidence not a function of Z alone: *relational* information existing only jointly across channels, and *interventional* information from acting on the world. We operationalize this as ReLATE (Relational Latent-state Acquisition Through Evidence), which compresses observed channels into a decision-sufficient relational state (§E.7) and ranks acquisitions by *decision-relevant information gain* (DRIG, §3.5) rather than generic uncertainty. Our contribution is not a new information identity. It is that test-time error contains three operationally *non-substitutable*

components which compute-versus-interaction views conflate, and that separating them yields predictions those views cannot express.

(C1) Conceptual and theoretical. Proposition 1 splits test-time excess risk into computational, representation and observation regret. The algebra is elementary; what it buys is the isolation of a third regime, *evidence already observed but discarded by the interface*, which a two-way split folds into acquisition. Theorem 1 characterizes the deficiency via a task-restricted Blackwell sufficiency, Proposition 2 shows same-state sampling cannot lower it, and Proposition 3 shows each operation class directly targets only its corresponding component, the other objects held fixed. We claim the decomposition and its consequences, not the underlying identities, which have precedents (§2).

(C2) Testable consequences, and a benchmark built to test them. The decomposition makes three predictions. **(P1)** Increasing K over a fixed interface may reduce computational regret but cannot move an information-limited Bayes floor. **(P2)** Holding the observed evidence *exactly* fixed, changing only how channels are jointly represented can move the attainable risk. **(P3)** Acquisition moves the floor only when the acquired channel carries decision-relevant information. **InfoGap** makes each floor designed, and in the vision case *tunable*, across exact POMDPs, procedural visual reasoning, and channel interventions on LIBERO (Liu et al., 2023). §6.1–6.2 test all three.

(C3) Operationalization. We instantiate bottleneck matching as ReLATE, which estimates each term's marginal risk reduction and trades it against that operation's cost over compute/fuse/acquire/stop; the action space is *derived* from (C1) rather than posited. This is a proof of concept that the decomposition can be operationalized at all, not a claim of transferable bottleneck diagnosis: within a known task mixture the routing component reaches 0.94 success versus 0.75 for Best-of-16, at $0.51\times$ its normalized generated-token cost, while leave-one-category-out shows it identifies task families rather than diagnosing instances.

The formalization earns its place on (P2), which the two-way view cannot state: late voting over two views sits at 0.500 on relational XOR where joint conditioning on the *same* views reaches 1.000 (§6.2). Acquisition cannot explain that contrast, because the σ-algebra of observed evidence is identical in both conditions. One sentence summarizes the paper: *sampling scales computation over what the model knows; ReLATE scales what the model can know.*

2 Related work

Fixed-information test-time scaling. Chain-of-thought (Wei et al., 2022), self-consistency (Wang et al., 2023b), Best-of-N with verifiers (Cobbe et al., 2021; Lightman et al., 2024), search (Yao et al., 2023a), repeated sampling (Brown et al., 2024), long reasoning (DeepSeek-AI et al., 2025; Muennighoff et al., 2025) and compute-optimal inference (Snell et al., 2024; Wu et al., 2024) all scale inference over a fixed encoding; recent theory analyses their limits (Huang et al., 2025b; Beirami et al., 2025; Huang et al., 2025a) but stays *within* a representation.

Closest prior work, and the exact delta. Non-computational gains are argued qualitatively via interaction, perception scaling, adaptive looking and context (Shen et al., 2025b; Jiang et al., 2026b; Choi et al., 2026; Shen et al., 2026; Jiang et al., 2026a). Recent VLM work separates perception from reasoning (Avogaro et al., 2026), jointly adapts visual-context and reasoning scaling (Jeddi et al., 2026), and systematically evaluates conventional test-time scaling on LVLMs (Sammani et al., 2026); these establish multi-axis allocation *operationally*, and our delta is the Bayes-risk decomposition separating discarded observed evidence from genuinely unobserved evidence, which is what distinguishes fusion from acquisition. Five further works need a precise delta. Li & Shi (2026) derive an internal-vs-external ceiling from the same inequality but treat all internal computation as one bucket; we show the external side *splits*, making fusion a third operation. Bay & Yearick (2026) and Majumdar (2026) bound selection and resolvable context, both computational-axis limits at a fixed information set, whereas Δ_ℓ bounds what *any* selection could reach and names the operation removing the residue. Sevetlidis (2026a) defines Bayes-sufficiency for a fixed decision problem and Sevetlidis (2026b) equates a frozen representation's log-loss excess risk with the predictive information it discards—our Theorem 1(ii), for which we claim no priority; we add its use as a *test-time* ceiling, extended to never-observed state, invariant to the sample budget, and tied to a three-way operation choice. Acquisition and decision-relevant gain are established (Liu et al., 2026; Howard, 1966; Lindley, 1956); Jeong & Yoon (2026) is the closest operational relative. We claim neither

the second axis nor acquisition, but its *resolution* into representation and observation; Appendix E.5 gives the full map, including why two-way sampling-or-retrieve routing cannot express fusion.

3 THEORY: THREE BOTTLENECKS, THREE RESOURCES

3.1 SETUP: INFORMATION SETS, NOT REPRESENTATION VECTORS

Let S be the true latent state, G the goal, X_{base} the observations available before any test-time computation, generating the *information set* $\mathcal{F}_0 = \sigma(X_{\text{base}}, G)$. Methods differ in *which* sub-σ-algebra they condition on: Z is the interface statistic a procedure actually uses (a prompt encoding, feature summary, or X_{base} itself), and δ_K is a *fixed-representation algorithm over* Z if it is $\sigma(Z, G, \xi_{1:K})$-measurable with $\xi_{1:K} \perp (S, Y) \mid (Z, G)$, covering self-consistency, Best-of-K, verifier reranking, self-refinement, and long reasoning over a given encoding. Acquiring an observation O^d instead *expands* the set: $\mathcal{F}_{t+1} = \mathcal{F}_t \vee \sigma(O^{d_t})$. Y is the decision target (in VLA: optimal action class, chunk code, subgoal, or success-equivalence class), and $R^*_\ell(W) = \inf_\delta \mathbb{E}[\ell(\delta(W), Y)]$ is the Bayes risk under source W. Unless noted, Y is finite/countable with $H(Y \mid Z, G) < \infty$ (Appendix B: continuous case).

3.2 A THREE-TERM REGRET DECOMPOSITION

Test-time methods act at three levels (compute harder over a fixed interface Z, build a better interface from observed evidence (fusion, moving Z toward X_{base}), or bring new observations into $\mathcal{F}_0$), and these reduce three separate error terms.

Proposition 1 (Three-term regret decomposition)**.** *For every fixed-representation algorithm δ_K over Z with Z $\mathcal{F}_0$-measurable,*

$$R_\ell(\delta_K) - R^*_\ell(S, G) = \underbrace{\left[R_\ell(\delta_K) - R^*_\ell(Z, G)\right]}_{\text{computational regret } \geq 0} + \underbrace{\left[R^*_\ell(Z, G) - R^*_\ell(X_{\text{base}}, G)\right]}_{\text{representation regret } \geq 0} + \underbrace{\left[R^*_\ell(X_{\text{base}}, G) - R^*_\ell(S, G)\right]}_{\text{observation regret } \geq 0}. \tag{2}$$

Under log loss, the last two terms decompose the total deficiency by the chain rule: $I(Y; S \mid Z, G) = I(Y; X_{\text{base}} \mid Z, G) + I(Y; S \mid X_{\text{base}}, G)$.

The identity is elementary (add and subtract twice); its value is that the three terms have three different remedies. *Computational regret* is what more inference over the same interface can fix; the decomposition only lower-bounds it by zero, and whether it *decreases* in K is an algorithmic property, not a consequence of equation 2 (nested classes $\mathcal{A}_K \subseteq \mathcal{A}_{K+1}$ give monotonicity; convergence needs coverage and consistent selection, which our synthetic procedures satisfy, no claim for deployed heuristics). *Representation regret* is decision-relevant information observed but not preserved by the interface (channels summarized separately, relations destroyed by per-channel encoding); fusion reduces it without expanding $\mathcal{F}_0$. *Observation regret* is information never observed (occlusion, perceptual aliasing (Whitehead & Ballard, 1991), missing history); no computation over $\mathcal{F}_0$ can reduce it. The "two axes" group these by *which operation removes them*: computational regret is reduced by more inference over a fixed interface (the *computation axis*), whereas the representation and observation terms together (the *decision-information axis*) are reduced by two distinct information operations, fusion (representation) and acquisition (observation). The sum of the last two is the *decision deficiency* $\Delta_\ell(S \to Z; Y \mid G) := R^*_\ell(Z, G) - R^*_\ell(S, G) \geq 0$; note that fusion, not only acquisition, can reduce it (by recovering its representation component), and only its observation component requires acting on the world.

Theorem 1 (Characterization of decision deficiency)**.** *Let Y be finite or countable with $H(Y \mid Z, G) < \infty$, and let Z be $\sigma(X_{\text{base}}, G)$-measurable and assume $Y \perp X_{\text{base}} \mid (S, G)$, hence also $Y \perp Z \mid (S, G)$. (We state the assumption this way rather than as an unconditional chain $Y \leftrightarrow (S, G) \leftrightarrow X_{\text{base}} \leftrightarrow Z$, because Z may depend on the goal G, and the unconditional chain then fails in general; every downstream node must carry G.) All information quantities are in nats. Then:* ***(i)*** *$\Delta_\ell(S \to Z; Y \mid G) = 0$ for* every *bounded measurable decision problem about Y (arbitrary measurable action spaces, bounded loss) if and only if $p(y \mid Z, G) = p(y \mid S, G)$*

almost surely, i.e., Z is posterior-sufficient for the induced task experiment*, a task-restricted Bayes sufficiency closely related to (but weaker than) Blackwell equivalence of experiments (Blackwell, 1953), which quantifies over all priors.* ***(ii)*** *Under log loss, $\Delta_{\log} = I(Y; S \mid Z, G)$ exactly.* ***(iii)*** *The same conditional mutual information* upper-bounds *the deficiency for* every *bounded loss with $\sup|\ell| \le L$. We stress the scope: $I(Y; S \mid Z, G)$ is exactly the fixed-interface* log-loss *deficiency, whereas the* 0–1 *accuracy ceiling is the separate quantity $A^*(Z, G) = 1 - R^*_{0\text{-}1}(Z, G)$; the two are related only through the bounds in this theorem, and $I(Y; S \mid Z, G) > 0$ does not by itself imply a drop in* 0–1 *accuracy:*

$$\Delta_\ell \;\le\; 2L\,\mathbb{E}\,\mathrm{TV}\big(p(y \mid S, G),\, p(y \mid Z, G)\big) \;\le\; 2L\sqrt{\tfrac{1}{2} I(Y; S \mid Z, G)}. \tag{3}$$

A complementary lower bound *holds in the special case of a deterministic finite-label task under 0–1 loss: when Y is determined by (S, G) and takes M values, Fano's inequality gives $\Delta_{0\text{-}1} \ge \max\{0,\, [I(Y; S \mid Z, G) - \ln 2]/\ln M\}$, so in that case the same quantity that upper-bounds the deficiency also forces it positive once it exceeds one bit. (We do not claim a two-sided bound for general bounded losses.)* ***(iv)*** *No fixed-representation algorithm's regret can fall below Δ_ℓ, but a single acquired observation O with $I(Y; O \mid Z, G) > 0$ strictly reduces the achievable floor under log loss.*

Proof in Appendix A. Part (i) is the nontrivial content (task-conditional sufficiency for *all* losses collapses to posterior matching), and (ii)–(iii) show one scalar $I(Y; S \mid Z, G)$ controls the deficiency across loss families.

Relation to reducible and irreducible uncertainty. Under log loss the Bayes risks are conditional entropies, so equation 2 splits total predictive uncertainty into an irreducible part $H(Y \mid S, G)$ that *no* operation removes plus the three reducible terms (Appendix E.4). This is why the controller needs a stop action and why calibrated total uncertainty is not a sufficient trigger for acquiring evidence: a high value is compatible with all four terms. These reducible terms are epistemic *in spirit* but are not the parameter uncertainty usually called epistemic in Bayesian deep learning.

Corollary 1 (Task-conditional information ceiling)**.** *Under log loss,*

$$C_{\mathrm{info}}(Z; Y \mid G) \;\coloneqq\; \Delta_{\log}(S \to Z; Y \mid G) \;=\; I(Y; S \mid Z, G)\,. \tag{4}$$

The ceiling is a property of the *triple* (representation, task, goal), not of Z alone (the same Z can be sufficient for Y_1 and deficient for Y_2), and quantifies in nats the irreducible log-loss gap of every fixed-representation method. Experimental figures plot accuracy $A^*(Z, G) = 1 - R^*_{0\text{-}1}(Z, G)$; the theory tracks risks.

3.3 No free information from same-state sampling

Suppose K proposals are drawn conditionally on the initial information set, $P_k \sim q_\theta(\cdot \mid Z, G)$, and an arbitrary statistic is computed from them, $U = f(Z, G, P_{1:K})$.

Proposition 2 (No-free-information principle)**.** *If $P_{1:K} \perp (S, Y) \mid (Z, G)$ and $U = f(Z, G, P_{1:K})$ for a (possibly randomized) f whose randomness is independent of (S, Y) (equivalently, if $(S, Y) \leftrightarrow (Z, G) \leftrightarrow (P_{1:K}, U)$ is a Markov chain, which is exactly the sampling semantics of a fixed-representation algorithm (§3.1)), then $U \perp (S, Y) \mid (Z, G)$, hence*

$$I(S; U \mid Z, G) = 0 \qquad \textit{and} \qquad I(Y; U \mid Z, G) = 0. \tag{5}$$

The name is deliberately narrow: this does *not* say more sampling cannot improve accuracy, only that it cannot create decision-relevant information absent from Z, and sampling may still reduce computational regret. It follows from the conditional data-processing inequality (Cover & Thomas, 2006): voting over or reasoning about candidates that all descend from the same information set cannot recreate information $\mathcal{F}_0$ lacks, though it *can* reduce computational regret by surfacing information implicit in Z. Li & Shi (2026) argue this "internal ceiling" qualitatively via the same inequality; stating it as a consequence of equation 2 makes it compose with Theorem 1 (bounding what any selection rule can *achieve*, not just a channel capacity), and extends it to the representation term, where the remedy is fusion rather than interaction.

3.4 NON-SUBSTITUTABILITY OF TEST-TIME RESOURCES

The decomposition becomes a statement about *resources* once we fix what each test-time operation is allowed to change. At an information set $\mathcal{F}$ with interface Z, define three operation classes:

$$\begin{aligned}
\mathcal{O}_{\text{comp}} &: \text{any } \sigma(Z, G, \xi_{1:K})\text{-measurable rule, } \xi_{1:K} \perp (S, Y) \mid (Z, G) && \text{(same } \mathcal{F}\text{, same } Z) \\
\mathcal{O}_{\text{fuse}} &: \text{re-encode } Z \mapsto Z' \text{ with } \sigma(Z, G) \subseteq \sigma(Z', G) \subseteq \sigma(X_{\text{base}}, G) && \text{(same } \mathcal{F}\text{, richer } Z) \\
\mathcal{O}_{\text{acq}} &: \text{expand } \mathcal{F} \mapsto \mathcal{F} \vee \sigma(O^d) && \text{(new evidence)}
\end{aligned}$$

Proposition 3 (Non-substitutability)**.** *Under the assumptions of Proposition 1, and holding the other two objects fixed, each operation class directly targets only its corresponding component (it may still change the* realized *risk of the others (fusing or acquiring can make downstream decoding easier), but not their floors): **(i)** every $\delta \in \mathcal{O}_{\text{comp}}$ satisfies $R_\ell(\delta) \geq R^*_\ell(Z, G)$, and the representation and observation terms of equation 2 do not depend on δ; **(ii)** $\inf_{Z' \in \mathcal{O}_{\text{fuse}}} R^*_\ell(Z', G) = R^*_\ell(X_{\text{base}}, G)$, so fusion can remove representation regret in full but leaves $R^*_\ell(X_{\text{base}}, G) - R^*_\ell(S, G)$ invariant; **(iii)** acquiring O^d lowers the floor below $R^*_\ell(X_{\text{base}}, G)$ only if $I\big(Y; O^d \,\big|\, X_{\text{base}}, G\big) > 0$. Consequently, for each class there are instances on which its own term is zero while another term is strictly positive; no single class is universally sufficient, and a policy restricted to one class cannot reach the optimum on some instance. Part (iii) also fixes the correct reading of the observation term: acquisition is the only operation that* may *reduce it, and it does so only for probes with $I\big(Y; O^d \,\big|\, X_{\text{base}}, G\big) > 0$. If no available sensor carries decision-relevant information the term is irreducible, so "reducible only by acquisition" is a statement about which operation can act, not a guarantee of recovery.*

Proof in Appendix A; each part follows from Proposition 1 and the conditional data-processing inequality, and we claim the *statement*, not the machinery. Part (i) is why sampling plateaus, (ii) why fusing already-observed channels can break a ceiling with no new observation, and (iii) why acquisition must be decision-targeted. InfoGap (§5) realizes all three corner cases by construction, so the proposition is directly testable.

This motivates the principle the rest of the paper follows.

Definition 1 (Bottleneck matching)**.** *Given a budget and per-operation costs c_a, spend the next unit on $a^\star \in \arg\max_{a \in \{\textit{compute}, \textit{fuse}, \textit{acquire}\}} [\widehat{\Delta R}_a - \lambda c_a]$, and **stop** when $\max_a [\widehat{\Delta R}_a - \lambda c_a] \leq 0$. We use this Lagrangian form throughout, matching Algorithm 1 and the utility used in the experiments; it is not equivalent to maximizing the ratio $\widehat{\Delta R}_a / c_a$.*

The question is therefore not *how much* to scale but *what* to scale: because the terms are non-substitutable, the correct intervention is determined by which term binds. ReLATE (§4) is one estimator-plus-controller instantiation of Definition 1.

3.5 ACTIVE INFORMATION EXPANSION

When the decision stays ambiguous the agent may execute a *diagnostic action* d, yielding O^d. Selection happens at the *realized* state, so with $r_\ell(z, g, u) = \inf_a \mathbb{E}[\ell(a, Y) \mid z, g, u]$ the statewise value of information is $r_\ell(z, g, u) - \mathbb{E}_{O^d} r_\ell(z, g, u, O^d, d)$, whose log-loss case is the statewise *decision-relevant information gain*

$$\text{DRIG}(d \mid z, g, u) \;=\; \mathbb{E}_{O^d \mid z, g, u, d}\, D_{\text{KL}}\big(p(Y \mid z, g, u, O^d, d) \,\|\, p(Y \mid z, g, u)\big), \tag{6}$$

which counts information about the decision target, not generic gain about the state (Howard, 1966; Lindley, 1956); under log loss it equals the statewise value of information, and for general bounded losses it controls the attainable risk reduction through Pinsker-type bounds (Theorem 1(iii)). Probes are ranked by $\text{DRIG}(d \mid z, g, u) - \lambda c(d)$. Averaging over states recovers the classical identity $\mathbb{E}[\text{DRIG}] = I\big(Y; O^d \,\big|\, Z, G, U, d\big)$ (Proposition 4); we claim no novelty for it, only its placement inside the same allocator as sampling. The Pinsker consequence (a low-DRIG probe cannot help much even when generic uncertainty is high) and the conditions under which greedy sequential selection inherits standard guarantees (Golovin & Krause, 2011; Krause & Guestrin, 2005) are in Appendices A and A.9; absent those conditions we treat it as a heuristic.

Algorithm 1 ReLATE one-step bottleneck matching (as evaluated)

Require: interface Z, goal G, held evidence E, cost weight λ, pilot budget K_0
draw pilot proposals $P_{1:K_0} \sim q_\theta(\cdot \mid Z, G)$ and form features $\psi(P_{1:K_0})$
$\widehat{\Delta R}(a) \leftarrow$ estimated risk reduction of a for each $a \in \{\mathsf{compute}, \mathsf{fuse}, \mathsf{acquire}\}$ from ψ
$a^* \leftarrow \arg\max_a [\widehat{\Delta R}(a) - \lambda c(a)]$
if $\max_a [\widehat{\Delta R}(a) - \lambda c(a)] \leq 0$ **then**
  return the pilot answer $\{\mathsf{stop}\}$
else if $a^* = \mathsf{compute}$ **then**
  return majority vote over K fresh samples on Z
else if $a^* = \mathsf{fuse}$ **then**
  return decision conditioned jointly on all channels already in E
else
  $d^* \leftarrow \arg\max_d [\widehat{\mathrm{DRIG}}(d \mid Z, G, E) - \lambda c(d)]$; execute d^*; **return** decision on (Z, E, O^{d^*})
end if

4 ReLATE: matching the operation to the bottleneck

ReLATE wraps a *frozen* base VLM or VLA policy. It spends K_0 cheap pilot samples on the current interface, forms diagnostic features from them, estimates each operation's marginal risk reduction, trades it against that operation's cost, and applies Definition 1. We deliberately present and evaluate the *one-step* form, which is what our experiments run:

Every branch returns a decision, and every conditioning variable is declared in REQUIRE. Scope: the algorithm defines four actions, but the learned router of §6.3 selects among compute/acquire/stop only; fuse is validated separately (§6.2), not as a routing arm, so we do not claim a four-way controller end to end. The iterative form (re-diagnosing after each action against a running budget) follows by repetition, but we do not evaluate it and therefore do not claim it; §7 lists this as the main gap between the framework and what is demonstrated. Only lightweight components are trained: the diagnostic map $\psi \mapsto a$, a DRIG predictor for probe ranking, and (on synthetic only) a set-attention relational encoder for the fuse branch; the base model is never fine-tuned. Implementation details, the encoder objective, and the conditional-IB view of fuse are in Appendix D.

5 The InfoGap benchmark

InfoGap makes the ceiling a *designed* quantity in three domains, so a plateau can be attributed to information rather than to tuning. **Synthetic**: factored POMDPs where every Bayes risk and conditional MI is computed by enumeration, giving exact ceilings and the four families used in Table 3 (aliasing, relational XOR, an irrelevant-information trap, and a compute-limited family). **Vision**: procedurally generated two-view scenes on Qwen3-VL-8B (plus Qwen3-VL-4B and Gemma-4-12B for transfer), where a constraint solver verifies *by enumeration* that single-view Bayes accuracy is ≤ 0.6 and joint-view ≥ 0.95 for every instance; the *occlusion-leak* family tunes single-view Bayes accuracy to any target p (§6.2), and *hard-count* is compute-limited by construction. **Embodied**: channel interventions on a frozen OpenVLA-OFT policy in LIBERO, reported as ancillary evidence in Appendix E. Generators, prompts, seeds and the solver are described in Appendix C.

6 Experiments

Setup. Models: Qwen3-VL-8B (primary, vLLM), Qwen3-VL-4B and Gemma-4-12B (transfer), OpenVLA-OFT 7B (embodied). Compute-axis baselines: single inference, self-consistency(K), Best-of-K with a same-model judge; information-axis: random probe/crop, all-views, and a trained DRIG selector. Synthetic quantities are exact by enumeration (504k rows). On the empirical VLM and VLA domains, where $I(Y; U \mid Z, G)$ is not identifiable, we report the *cross-fitted log-loss improvement* $\widehat{\Delta}(U) = \mathrm{CE}(q_0) - \mathrm{CE}(q_1)$; we do *not* call it an MI bound, since $\widehat{\Delta}(U) = I(Y; U \mid Z, G) + D_0 - D_1$ with either-signed bias, a predictive-information proxy (cf. Belghazi et al., 2018; Poole et al., 2019; McAllester & Stratos, 2020) that on synthetic tracks exact

Table 1: **Main result: the information ceiling on InfoGap-Vision** (Qwen3-VL-8B; accuracy with scene-clustered 95% CIs, n=500 scenes/category). *Compute-axis* methods (single, self-consistency, Best-of-16) all sit at the single-view Bayes ceiling (≈ chance) across the tested budgets, up to K=16 (17× the token budget of a single pass). *Decision-information* methods that expand the view break it: a random crop recovers part of the gap, and one decision-relevant view (ReLATE, B=1) reaches near-perfect accuracy at the *single-inference* token cost. "Tokens" is the configured generated-token budget; "axis" marks which axis the method scales.

Method	axis	tokens	occlusion	before/after	relational XOR
Single view (K=1)	–	64	0.496 [.452, .540]	0.480 [.436, .524]	0.485 [.443, .530]
Self-consistency (K=16)	compute	1024	0.496 [.452, .540]	0.480 [.436, .524]	0.500 [.465, .536]
Best-of-16 (verifier)	compute	1088	0.496 [.452, .540]	0.480 [.436, .524]	0.481 [.444, .517]
Random crop (B=1)	info	64	0.761 [.733, .788]	0.742 [.713, .770]	0.747 [.723, .771]
All views (K=1)	info	64	**1.000**	**1.000**	**1.000**
ReLATE acquire (B=1)	info	64	**1.000** [1.0, 1.0]	**0.977** [.967, .985]	**1.000** [1.0, 1.0]
InfoGap (ReLATE−Best-of-16)			**+0.504**	**+0.497**	**+0.519**

MI at $R^2 = 0.999$. State rate uses the variational upper bound $\mathbb{E}\, D_{\mathrm{KL}}(q_\varphi \| p) \geq I(U; E \mid Z, G)$. For the visual experiments with repeated stochastic generations, all significance tests cluster at the *scene* level to avoid pseudo-replication: effect CIs are scene-clustered bootstraps (10k resamples, unit = scene with its generations nested), and paired significance uses a scene-level sign test on per-scene mean outcomes (intervals report scenes and runs). For the mixed router experiment (§6.3), where each instance is a distinct task with one outcome per arm, the independent task instance is the bootstrap unit.

Synthetic validity check. On InfoGap-Synthetic every Bayes risk and conditional MI is exact by enumeration and the three terms dissociate exactly as Proposition 3 requires; this confirms the theory rather than adding independent evidence (Appendix E).

6.1 A REAL VLM HITS THE CEILING; ONE RELEVANT VIEW BREAKS IT

Table 1 is the headline on Qwen3-VL-8B (n=500/category; per-category K-sweeps in Figure 6). The main categories put a single view at *zero* decision information, so the ceiling is chance; its location is trivial by design, but three things are not: it is exactly attained, no sampling or verification escapes it (a verifier could have selected confidently wrong), and the same dissociation appears *away* from chance on synthetic aliasing (ceiling 0.90). The InfoGap of Table 1 (+0.50 to +0.52) is statistically decisive: a scene-level sign test rejects at $p < 10^{-75}$ with *every* discordant scene favoring acquisition (252/260/342 of 500 scenes strictly higher, none lower). The gain is *what the budget buys, not the budget*: Best-of-16 spends 1088 tokens to stay at chance where one relevant view costs 64. **Transfer (E9):** the same InfoGap (+0.44 to +0.52) holds on Qwen3-VL-4B and Gemma-4-12B (Table 8), supporting that the observed ceiling is driven by the information set rather than a peculiarity of one network.

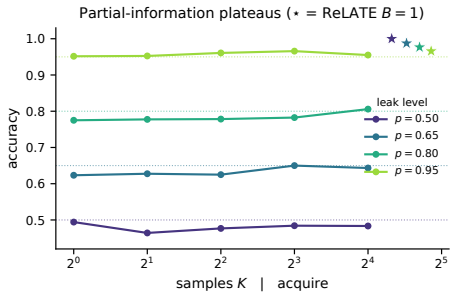


Figure 2: **Partial-information plateaus** (occlusion-leak, Qwen3-VL-8B). Self-consistency (solid) tracks the designed single-view Bayes accuracy p (dotted, within 0.02); acquiring the complementary view (⋆) stays at 0.97–1.00, the two meeting as $p \to 0.95$.

6.2 THE CEILING AWAY FROM CHANCE: PARTIAL-INFORMATION REGIMES

To show the plateau is not an artifact of the chance construction, we build *occlusion-leak*, whose single-view Bayes accuracy is a tunable, exact $p \in \{0.50, 0.65, 0.80, 0.95\}$ (the occluder reveals the decisive region with probability $2p-1$; single-view Bayes accuracy $= p$ by

enumeration, joint $= 1.0$). If the ceiling tracked the construction rather than information, it would not follow p; instead (Figure 2, 400 instances $\times$ 3 seeds/level) self-consistency tracks the designed single-view Bayes bound closely (measured $0.483/0.643/0.806/0.955$ vs. designed $0.50/0.65/0.80/0.95$, each within 0.02), while acquiring the complementary view stays high but not perfect ($1.000/0.988/0.977/0.966$). The InfoGap shrinks smoothly as scarcity eases ($+0.52, +0.35, +0.17, +0.01$; scene-clustered), and at p=0.95 sampling has essentially reached the (now near-unity) ceiling so acquisition's residual advantage is negligible. On a real VLM, more samples help only up to where the information runs out, wherever that point is placed.

Fusion at fixed evidence: what a two-way view cannot explain. Every result above adds or withholds an observation; this one does not. Both channels are already in hand, nothing is acquired, and only the *combination rule* changes. Late voting (each view answered separately, then voted) sits at 0.500 on relational XOR, exactly the single-view baseline, while joint conditioning on the *same* two views reaches 1.000 (0.929 on before/after; n=450/cell, Figure 7). No observation is added, so neither sampling nor acquisition can account for the gain. This is a restricted-class gap, not a bound on representation regret (Appendix E.6).

Decision-relevant vs. generic acquisition (P3). Acquiring the *right* evidence differs from acquiring more: on the targeted-probe family context-only inference is 0.540 at any $K \leq 16$ and a random probe reaches 0.612, where a trained DRIG selector reaches $\mathbf{1.000}$. Opening *all four* probes gives only 0.793, a *computational* cost of irrelevant context, so the two axes interact and acquisition must be decision-targeted (details, and $\rho = 0.78$ against exact gain, in Appendix E).

6.3 Learned bottleneck matching with a pretrained VLM

We run the one-step router of Algorithm 1 on a pretrained VLM over a mixed pool (n_{test}=1800: 40% information-limited, 40% compute-limited, 20% saturated), reading six cheap features from two pilot samples (pilot agreement, mean and minimum self-rated confidence, parse rate, answer length, and acquisition availability, with *no* category identity or oracle information, as enforced by a schema test), and selecting compute/acquire/stop under $u =$ success $- \lambda c$, λ=0.1 fixed before any test evaluation on labels from a disjoint validation split (n_{val}=900).

Matching dominates every fixed policy (Table 4): 0.938 success at half the generated-token cost of Best-of-16, which reaches only 0.748 (paired margin $+0.189$ $[+0.171, +0.209]$). Per regime it selects exactly the operation the decomposition predicts. **Interpretation.** A category-only baseline already reaches 0.931, and under leave-one-category-out the router drops to 0.695 pooled, *below* fixed compute. We therefore read this as *controlled task-family routing*, not transferable instance-level diagnosis. Full setup, the four router families, and the confound controls are in Appendix E (Table 4).

7 Conclusion

Test-time excess risk splits into three bottlenecks (computational, representation and observation), each directly targeted by a distinct operation class: compute, fusion and acquisition. Proposition 3 shows these resources are non-substitutable when the defining information objects are held fixed. The middle term is what a compute-versus-interaction view misses and motivates a fixed-evidence fusion axis, probed operationally in §6.2: changing only the combination rule at identical evidence moves accuracy from 0.500 to 1.000, though that restricted-class contrast is not itself an estimate of representation regret. The ceiling behaves as designed, exactly on enumerable POMDPs and on a real VLM tracking a *tunable* single-view Bayes bound within 0.02. Bottleneck matching reaches 0.94 success versus 0.75 for Best-of-16 within a known mixture, at $0.51\times$ its normalized generated-token cost, but leave-one-category-out drops it to 0.695: the router identifies task families, not transferable bottlenecks. Transferable diagnosis, a controller that also routes fuse, and the iterative loop remain open.

Limitations. The two that most constrain the claims: the vision benchmark is procedural, so natural-image validation is absent, and the router identifies task families rather than diagnosing instances. Appendix F lists all of them. Test-time scaling should ask not how much to spend but *what to spend it on*.

REPRODUCIBILITY STATEMENT

All theoretical claims are proved in Appendix A with stated assumptions. The anonymous supplementary material includes the router pipeline and logs for the router, the primary Qwen3-VL-8B vision experiments, the partial-information sweep, the state-construction frontier, and selected embodied interventions, together with a SHA-256 manifest and reproduction map. It is a verified subset, not the whole pipeline: the multi-model transfer logs, the late-voting records behind Figure 7, and the masked action-sampling and wrist control logs are *not* included, and the plotting scripts for three vision result figures were not retained. Those generators, drivers and logs will be released. Appendix C describes the generation procedures. All reported numbers come from logged runs (504,000 exact-enumeration rows on synthetic; over $450{,}000$ VLM evaluations spanning vision-main, transfer, probe, the $196{,}800$-row partial-information sweep, and $8{,}100$ controller arm-executions across the disjoint 900-instance validation and $1{,}800$-instance test pools; 800 embodied rollouts plus controls), with per-row provenance (task, method, K, B, seed, costs, estimator) archived in the authors' internal experiment repository.

USE OF LARGE LANGUAGE MODELS

Generative AI tools were used as productivity aids for language polishing and software development, including assistance with implementation, debugging, and experiment workflow automation. The research questions, core ideas, methodology, theoretical development, experimental design, analysis, and conclusions were developed by the authors. The authors take full responsibility for the content of the paper.

REFERENCES

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. In *International Conference on Learning Representations (ICLR)*, 2017. arXiv:1612.00410.

Niccolo Avogaro, Nayanika Debnath, Li Mi, Thomas Frick, Junling Wang, Zexue He, Hang Hua, Konrad Schindler, and Mattia Rigotti. SPARC: Separating perception and reasoning circuits for test-time scaling of VLMs. In *International Conference on Machine Learning (ICML)*, 2026. arXiv:2602.06566.

Ruzena Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, 1988.

Ruzena Bajcsy, Yiannis Aloimonos, and John K. Tsotsos. Revisiting active perception. *Autonomous Robots*, 42(2):177–196, 2018.

Yong Yi Bay and Kathleen A. Yearick. When more sampling hurts: The modal ceiling and correlation ceiling of test-time scaling. *arXiv preprint arXiv:2606.28661*, 2026.

Ahmad Beirami, Alekh Agarwal, Jonathan Berant, Alexander D'Amour, Jacob Eisenstein, Chirag Nagpal, and Ananda Theertha Suresh. Theoretical guarantees on the best-of-n alignment policy. In *International Conference on Machine Learning (ICML)*, 2025. arXiv:2401.01879.

Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeswar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and R Devon Hjelm. Mutual information neural estimation. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018. arXiv:1801.04062.

Nils Bertschinger, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. Quantifying unique information. *Entropy*, 16(4):2161–2183, 2014.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Lucy Xiaoyang Shi, James Tanner, Quan Vuong, Anna Walling, Haohuan Wang, and Ury Zhilinsky. π_0: A vision-language-action flow model for general robot control. In *Robotics: Science and Systems (RSS)*, 2025. arXiv:2410.24164.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

Jeannette Bohg, Karol Hausman, Bharath Sankaran, Oliver Brock, Danica Kragic, Stefan Schaal, and Gaurav S. Sukhatme. Interactive perception: Leveraging action in perception and perception in action. *IEEE Transactions on Robotics*, 33(6):1273–1291, 2017.

Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, Pete Florence, Chuyuan Fu, Montse Gonzalez Arenas, Keerthana Gopalakrishnan, Kehang Han, Karol Hausman, Alexander Herzog, Jasmine Hsu, Brian Ichter, Alex Irpan, et al. RT-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning (CoRL)*, 2023. arXiv:2307.15818.

Bradley Brown, Jordan Juravsky, Ryan Ehrlich, Ronald Clark, Quoc V. Le, Christopher Ré, and Azalia Mirhoseini. Large language monkeys: Scaling inference compute with repeated sampling. *arXiv preprint arXiv:2407.21787*, 2024.

Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

Xiao Chen, Quanyi Li, Tai Wang, Tianfan Xue, and Jiangmiao Pang. GenNBV: Generalizable next-best-view policy for active 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. arXiv:2402.16174.

Hyeonbeom Choi, Daechul Ahn, Youhan Lee, Taewook Kang, Seongwon Cho, and Jonghyun Choi. SCALE: Self-uncertainty conditioned adaptive looking and execution for vision-language-action models. In *International Conference on Machine Learning (ICML)*, 2026. arXiv:2602.04208.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.

Mingtong Dai, Lingbo Liu, Yongjie Bai, Yang Liu, Zhouxia Wang, Rui Su, Chunjie Chen, Liang Lin, and Xinyu Wu. Rover: Robot reward model as test-time verifier for vision-language-action model. *arXiv preprint arXiv:2510.10975*, 2025.

DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. *Nature*, 645:633–638, 2025. arXiv:2501.12948.

Wenxuan Ding, Nicholas Tomlin, and Greg Durrett. Calibrate-then-act: Cost-aware exploration in LLM agents. *arXiv preprint arXiv:2602.16699*, 2026.

Marco Federici, Anjan Dutta, Patrick Forré, Nate Kushman, and Zeynep Akata. Learning robust representations via multi-view information bottleneck. In *International Conference on Learning Representations (ICLR)*, 2020. arXiv:2002.07017.

Guhao Feng, Bohang Zhang, Yuntian Gu, Haotian Ye, Di He, and Liwei Wang. Towards revealing the mystery behind chain of thought: A theoretical perspective. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. arXiv:2305.15408.

Ian Fischer. The conditional entropy bottleneck. *Entropy*, 22(9):999, 2020.

Adam Foster, Martin Jankowiak, Eli Bingham, Paul Horsfall, Yee Whye Teh, Tom Rainforth, and Noah Goodman. Variational bayesian optimal experimental design. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. arXiv:1903.05480.

Xingyu Fu, Yushi Hu, Bangzheng Li, Yu Feng, Haoyu Wang, Xudong Lin, Dan Roth, Noah A. Smith, Wei-Chiu Ma, and Ranjay Krishna. BLINK: Multimodal large language models can see but not perceive. In *European Conference on Computer Vision (ECCV)*, 2024. arXiv:2404.12390.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

Virgil Griffith and Christof Koch. Quantifying synergistic mutual information. In Mikhail Prokopenko (ed.), *Guided Self-Organization: Inception*, pp. 159–190. Springer, 2014.

Mohsen Hariri, Weicong Chen, Nahal Shahini, Vikash Singh, Kai Ye, Amirhossein Samandar, Debargha Ganguly, Sreehari Sankar, Yanyan Zhang, Shouren Wang, Jerry Peng, Biyao Zhang, Michael Hinczewski, and Vipin Chaudhary. Test-time scaling in reasoning LLMs: Inference regimes, evaluation, and reproducibility. *arXiv preprint arXiv:2608.04001*, 2026.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, and Laurent Sifre. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022.

Arian Hosseini, Xingdi Yuan, Nikolay Malkin, Aaron Courville, Alessandro Sordoni, and Rishabh Agarwal. V-STaR: Training verifiers for self-taught reasoners. *arXiv preprint arXiv:2402.06457*, 2024.

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, and Máté Lengyel. Bayesian active learning for classification and preference learning. *arXiv preprint arXiv:1112.5745*, 2011.

Ronald A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, 1966.

Yushi Hu, Weijia Shi, Xingyu Fu, Dan Roth, Mari Ostendorf, Luke Zettlemoyer, Noah A. Smith, and Ranjay Krishna. Visual sketchpad: Sketching as a visual chain of thought for multimodal language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. arXiv:2406.09403.

Audrey Huang, Adam Block, Dylan J. Foster, Dhruv Rohatgi, Cyril Zhang, Max Simchowitz, Jordan T. Ash, and Akshay Krishnamurthy. Self-improvement in language models: The sharpening mechanism. *arXiv preprint arXiv:2412.01951*, 2024a.

Audrey Huang, Adam Block, Qinghua Liu, Nan Jiang, Akshay Krishnamurthy, and Dylan J. Foster. Is best-of-n the best of them? coverage, scaling, and optimality in inference-time alignment. *arXiv preprint arXiv:2503.21878*, 2025a.

Baihe Huang, Shanda Li, Tianhao Wu, Yiming Yang, Ameet Talwalkar, Kannan Ramchandran, Michael I. Jordan, and Jiantao Jiao. Sample complexity and representation ability of test-time scaling paradigms. *arXiv preprint arXiv:2506.05295*, 2025b.

Jie Huang, Xinyun Chen, Swaroop Mishra, Huaixiu Steven Zheng, Adams Wei Yu, Xinying Song, and Denny Zhou. Large language models cannot self-correct reasoning yet. In *International Conference on Learning Representations (ICLR)*, 2024b. arXiv:2310.01798.

Ahmadreza Jeddi, Minh Ngoc Le, Amirhossein Kazerouni, Hakki Can Karaimer, Hue Nguyen, Iqbal Mohomed, Michael Brudno, Alex Levinshtein, Konstantinos G. Derpanis, Babak Taati, and Radek Grzeszczuk. AVIS: Adaptive test-time scaling for vision-language models. *arXiv preprint arXiv:2606.11576*, 2026.

Yuhwan Jeong and Kuk-Jin Yoon. Reuse before you retrieve: Diagnosing headroom and complementarity for test-time augmentation of embodied multimodal policies. *arXiv preprint arXiv:2608.17484*, 2026. ECCV 2026 workshop.

Yunfan Jiang, Yevgen Chebotar, Ruijie Zheng, Fengyuan Hu, Yunhao Ge, Jimmy Wu, Tianyuan Dai, Scott Reed, Li Fei-Fei, Yuke Zhu, and Linxi Fan. RoboTTT: Context scaling for robot policies. *arXiv preprint arXiv:2607.15275*, 2026a.

Zheng Jiang, Yiming Chen, Nan He, Jiahui Chen, Chaoyang Li, Houde Qian, and Lifeng Sun. Test-time scaling over perception: Resolving the grounding paradox in thinking with images. *arXiv preprint arXiv:2604.11025*, 2026b.

Andy L. Jones. Scaling scaling laws with board games. *arXiv preprint arXiv:2104.03113*, 2021.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134, 1998.

Ryo Kamoi, Yusen Zhang, Nan Zhang, Jiawei Han, and Rui Zhang. When can LLMs actually correct their own mistakes? A critical survey of self-correction of LLMs. *Transactions of the Association for Computational Linguistics*, 12:1417–1440, 2024. arXiv:2406.01297.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. OpenVLA: An open-source vision-language-action model. In *Conference on Robot Learning (CoRL)*, 2024. arXiv:2406.09246.

Moo Jin Kim, Chelsea Finn, and Percy Liang. Fine-tuning vision-language-action models: Optimizing speed and success. In *Robotics: Science and Systems (RSS)*, 2025. arXiv:2502.19645.

Andreas Krause and Carlos Guestrin. Near-optimal nonmyopic value of information in graphical models. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 324–331, 2005.

Jacky Kwok, Christopher Agia, Rohan Sinha, Matt Foutter, Shulu Li, Ion Stoica, Azalia Mirhoseini, and Marco Pavone. Robomonkey: Scaling test-time sampling and verification for vision-language-action models. *arXiv preprint arXiv:2506.17811*, 2025.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R. Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019. arXiv:1810.00825.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. arXiv:2005.11401.

Bojie Li and Noah Shi. Interaction scaling: Grounding the third axis of test-time compute. *arXiv preprint arXiv:2607.11598*, 2026.

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

Xuanlin Li, Kyle Hsu, Jiayuan Gu, Karl Pertsch, Oier Mees, Homer Rich Walke, Chuyuan Fu, Ishikaa Lunawat, Isabel Sieh, Sean Kirmani, Sergey Levine, Jiajun Wu, Chelsea Finn, Hao Su, Quan Vuong, and Ted Xiao. Evaluating real-world robot manipulation policies in simulation. In *Conference on Robot Learning (CoRL)*, 2024. arXiv:2405.05941.

Hunter Lightman, Vineet Kosaraju, Yura Burda, Harri Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2305.20050.

Dennis V. Lindley. On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005, 1956.

Anjie Liu, Ziqin Gong, Yan Song, Yuxiang Chen, Xiaolong Liu, Hengtong Lu, Kaike Zhang, Chen Wei, and Jun Wang. The perceptual bandwidth bottleneck in vision-language models: Active visual reasoning via sequential experimental design. In *International Conference on Machine Learning (ICML)*, 2026. arXiv:2605.01345.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. LIBERO: Benchmarking knowledge transfer for lifelong robot learning. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, 2023. arXiv:2306.03310.

Joseph T. Lizier, Nils Bertschinger, Jürgen Jost, and Michael Wibral. Information decomposition of target effects from multi-source interactions: Perspectives on previous, current and future work. *Entropy*, 20(4):307, 2018.

Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. Self-refine: Iterative refinement with self-feedback. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. arXiv:2303.17651.

Subhabrata Majumdar. Limits of reliability and scaling in language models. *arXiv preprint arXiv:2607.14112*, 2026.

David McAllester and Karl Stratos. Formal limitations on the measurement of mutual information. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108 of *Proceedings of Machine Learning Research*, pp. 875–884, 2020.

Fanqing Meng, Jin Wang, Chuanhao Li, Quanfeng Lu, Hao Tian, Jiaqi Liao, Xizhou Zhu, Jifeng Dai, Yu Qiao, Ping Luo, Kaipeng Zhang, and Wenqi Shao. MMIU: Multimodal multi-image understanding for evaluating large vision-language models. *arXiv preprint arXiv:2408.02718*, 2024.

William Merrill and Ashish Sabharwal. The expressive power of transformers with chain of thought. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2310.07923.

Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke Zettlemoyer, Percy Liang, Emmanuel Candès, and Tatsunori Hashimoto. s1: Simple test-time scaling. *arXiv preprint arXiv:2501.19393*, 2025.

Mitsuhiko Nakamoto, Oier Mees, Aviral Kumar, and Sergey Levine. Steering your generalists: Improving robotic foundation models via value guidance. In *Conference on Robot Learning (CoRL)*, 2024. arXiv:2410.13816.

Octo Model Team, Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, Jianlan Luo, You Liang Tan, Lawrence Yunliang Chen, Pannag Sanketi, Quan Vuong, Ted Xiao, Dorsa Sadigh, Chelsea Finn, and Sergey Levine. Octo: An open-source generalist robot policy. In *Robotics: Science and Systems (RSS)*, 2024. arXiv:2405.12213.

Yury Polyanskiy and Yihong Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2025.

Ben Poole, Sherjil Ozair, Aäron van den Oord, Alexander A. Alemi, and George Tucker. On variational bounds of mutual information. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, pp. 5171–5180, 2019.

Ofir Press, Muru Zhang, Sewon Min, Ludwig Schmidt, Noah A. Smith, and Mike Lewis. Measuring and narrowing the compositionality gap in language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023. arXiv:2210.03350.

Tom Rainforth, Adam Foster, Desi R. Ivanova, and Freddie Bickford Smith. Modern bayesian experimental design. *Statistical Science*, 39(1):100–114, 2024.

Allen Z. Ren, Anushri Dixit, Alexandra Bodrova, Sumeet Singh, Stephen Tu, Noah Brown, Peng Xu, Leila Takayama, Fei Xia, Jake Varley, Zhenjia Xu, Dorsa Sadigh, Andy Zeng, and Anirudha Majumdar. Robots that ask for help: Uncertainty alignment for large language model planners. In *Conference on Robot Learning (CoRL)*, 2023. arXiv:2307.01928.

Allen Z. Ren, Jaden Clark, Anushri Dixit, Masha Itkina, Anirudha Majumdar, and Dorsa Sadigh. Explore until confident: Efficient exploration for embodied question answering. In *Robotics: Science and Systems (RSS)*, 2024. arXiv:2403.15941.

Jon Saad-Falcon, E. Kelly Buchanan, Mayee F. Chen, Tzu-Heng Huang, Brendan McLaughlin, Tanvir Bhathal, Shang Zhu, Ben Athiwaratkun, Frederic Sala, Scott Linderman, Azalia Mirhoseini, and Christopher Ré. Shrinking the generation-verification gap with weak verifiers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2506.18203.

Fawaz Sammani, Tzoulio Chamiti, and Nikos Deligiannis. On test-time scaling for vision-language models. In *European Conference on Computer Vision (ECCV)*, 2026. arXiv:2606.28864.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. *arXiv preprint arXiv:2302.04761*, 2023.

Burr Settles. Active learning literature survey. Technical Report 1648, University of Wisconsin–Madison, Department of Computer Sciences, 2009.

Vasileios Sevetlidis. Bayes-sufficient representations in supervised learning. *arXiv preprint arXiv:2606.04045*, 2026a.

Vasileios Sevetlidis. Loss-shift transfer via bayes quotients. *arXiv preprint arXiv:2606.13178*, 2026b.

Haozhan Shen, Kangjia Zhao, Tiancheng Zhao, Ruochen Xu, Zilun Zhang, Mingwei Zhu, and Jianwei Yin. Zoomeye: Enhancing multimodal llms with human-like zooming capabilities through tree-based image exploration. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025a. arXiv:2411.16044.

Junhong Shen, Hao Bai, Lunjun Zhang, Yifei Zhou, Amrith Setlur, Shengbang Tong, Diego Caples, Nan Jiang, Tong Zhang, Ameet Talwalkar, and Aviral Kumar. Thinking vs. doing: Agents that reason by scaling test-time interaction. *arXiv preprint arXiv:2506.07976*, 2025b.

Maohao Shen, Kaiwen Zha, Zexue He, Zhang-Wei Hong, Siru Ouyang, J. Jon Ryu, Prasanna Sattigeri, Suhas Diggavi, and Gregory Wornell. Decocted experience improves test-time inference in LLM agents. *arXiv preprint arXiv:2604.04373*, 2026.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling LLM test-time compute optimally can be more effective than scaling model parameters. *arXiv preprint arXiv:2408.03314*, 2024.

Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop (ITW)*, pp. 1–5. IEEE, 2015.

Naftali Tishby, Fernando C. Pereira, and William Bialek. The information bottleneck method. In *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, pp. 368–377, 1999.

Fei Wang, Xingyu Fu, James Y. Huang, Zekun Li, Qin Liu, Xiaogeng Liu, Mingyu Derek Ma, Nan Xu, Wenxuan Zhou, Kai Zhang, Tianyi Lorena Yan, Wenjie Jacky Mo, Hsiang-Hui Liu, Pan Lu, Chunyuan Li, Chaowei Xiao, Kai-Wei Chang, Dan Roth, Sheng Zhang, Hoifung Poon, and Muhao Chen. Muirbench: A comprehensive benchmark for robust multi-image understanding. *arXiv preprint arXiv:2406.09411*, 2024.

Hongru Wang, Boyang Xue, Baohang Zhou, Tianhua Zhang, Cunxiang Wang, Huimin Wang, Guanhua Chen, and Kam-Fai Wong. Self-DC: When to reason and when to act? self divide-and-conquer for compositional unknown questions. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*, pp. 6510–6525, 2025. arXiv:2402.13514.

Peiyi Wang, Lei Li, Zhihong Shao, R.X. Xu, Damai Dai, Yifei Li, Deli Chen, Y. Wu, and Zhifang Sui. Math-shepherd: Verify and reinforce LLMs step-by-step without human annotations. *arXiv preprint arXiv:2312.08935*, 2023a.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc Le, Ed Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations (ICLR)*, 2023b. arXiv:2203.11171.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. arXiv:2201.11903.

Steven D. Whitehead and Dana H. Ballard. Learning to perceive and act by trial and error. *Machine Learning*, 7:45–83, 1991.

Paul L. Williams and Randall D. Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.

Penghao Wu and Saining Xie. V*: Guided visual search as a core mechanism in multimodal LLMs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. arXiv:2312.14135.

Yangzhen Wu, Zhiqing Sun, Shanda Li, Sean Welleck, and Yiming Yang. Inference scaling laws: An empirical analysis of compute-optimal inference for problem-solving with language models. *arXiv preprint arXiv:2408.00724*, 2024.

Siheng Xiong, Oguzhan Gungordu, James C. Kerce, and Faramarz Fekri. Adaptive information control for search-augmented LLM reasoning. *arXiv preprint arXiv:2602.01672*, 2026.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023a. arXiv:2305.10601.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. ReAct: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*, 2023b. arXiv:2210.03629.

Wen Ye, Peiyan Li, Tingyu Yuan, Yuan Xu, Xiangnan Wu, Chaoyang Zhao, Jing Liu, Nianfeng Liu, Yan Huang, and Liang Wang. E-TTS: A new embodied test-time scaling framework for robotic manipulation. In *European Conference on Computer Vision (ECCV)*, 2026. arXiv:2606.27268.

Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah D. Goodman. STaR: Bootstrapping reasoning with reasoning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. arXiv:2203.14465.

Guibin Zhang, Fanci Meng, Guancheng Wan, Zherui Li, Kun Wang, Zhenfei Yin, Lei Bai, and Shuicheng Yan. Latentevolve: Self-evolving test-time scaling in latent space. *arXiv preprint arXiv:2509.24771*, 2025a.

Jiarui Zhang, Mahyar Khayatkhoei, Prateek Chhikara, and Filip Ilievski. Using visual cropping to enhance fine-detail question answering of BLIP-family models. *arXiv preprint arXiv:2306.00228*, 2023.

Lunjun Zhang, Arian Hosseini, Hritik Bansal, Mehran Kazemi, Aviral Kumar, and Rishabh Agarwal. Generative verifiers: Reward modeling as next-token prediction. In *International Conference on Learning Representations (ICLR)*, 2025b. arXiv:2408.15240.

Xintong Zhang, Zhi Gao, Bofei Zhang, Pengxiang Li, Xiaowen Zhang, Yang Liu, Tao Yuan, Yuwei Wu, Yunde Jia, Song-Chun Zhu, and Qing Li. Adaptive chain-of-focus reasoning via dynamic visual search and zooming for efficient vlms. *arXiv preprint arXiv:2505.15436*, 2025c.

A PROOFS

Throughout, all random variables are defined on a common probability space and take values in standard Borel spaces, so regular conditional distributions exist and the simulation arguments below are well defined; Y takes values in a standard Borel space $\mathcal{Y}$ (finite or countable with $H(Y \mid Z, G) < \infty$ for all log-loss statements; the continuous extension is Appendix B); losses are measurable and bounded where stated; all logarithms are natural (nats) unless a figure axis states bits. We use the joint law $p(s, g, x, y)$ with $Y \perp X_{\text{base}} \mid (S, G)$ and Z $\sigma(X_{\text{base}}, G)$-measurable, hence $Y \perp Z \mid (S, G)$ (the decision target depends on the world only through (S, G); the agent's statistic depends on the world only through observations, but may also depend on the goal). We avoid writing this as an unconditional Markov chain because Z may be goal-dependent. *Measurability convention (important):* a fixed-representation algorithm over Z is, by the definition in §3.1, measurable with respect to $\sigma(Z, G, \xi_{1:K})$ with $\xi_{1:K} \perp (S, Y) \mid (Z, G)$; it may *not* access parts of X_{base} outside $\sigma(Z, G)$. Without this restriction the computational-regret term of Proposition 1 can be negative (e.g., if X_{base} contains a channel that reveals Y but Z discards it, an algorithm reading that channel beats the Z-Bayes risk), which is precisely why the three-term decomposition separates representation regret from observation regret.

A.1 PROOF OF PROPOSITION 1

The identity is a telescoping sum: add and subtract $R^*_\ell(Z, G)$ and $R^*_\ell(X_{\text{base}}, G)$. Nonnegativity of the first term: δ_K is $\sigma(Z, G, \xi_{1:K})$-measurable with $\xi_{1:K} \perp (S, Y) \mid (Z, G)$, hence a randomized decision rule based on (Z, G). For any such rule, $\mathbb{E}[\ell(\delta_K, Y)] = \mathbb{E}\big[\mathbb{E}[\ell(\delta_K, Y) \mid Z, G, \xi_{1:K}]\big] \geq \mathbb{E}\big[\inf_a \mathbb{E}[\ell(a, Y) \mid Z, G]\big] = R^*_\ell(Z, G)$, where the inner conditional risk does not depend on ξ by conditional independence. Nonnegativity of the second term: Z is $\mathcal{F}_0$-measurable, so any rule based on (Z, G) is also a rule based on (X_{base}, G), giving $R^*_\ell(X_{\text{base}}, G) \leq R^*_\ell(Z, G)$. Nonnegativity of the third term: any decision rule $\delta(X_{\text{base}}, G)$ can be simulated by a decision-maker who observes (S, G): draw $X \sim p(x \mid s, g)$, apply $\delta(X, G)$. Hence $R^*_\ell(S, G) \leq R^*_\ell(X_{\text{base}}, G)$. The chain-rule statement: under log loss the three Bayes risks are $H(Y \mid Z, G) \geq H(Y \mid X_{\text{base}}, G) \geq H(Y \mid S, G)$, and $I(Y; S \mid Z, G) = [H(Y \mid Z, G) - H(Y \mid X_{\text{base}}, G)] + [H(Y \mid X_{\text{base}}, G) - H(Y \mid S, G)] = I(Y; X_{\text{base}} \mid Z, G) + I(Y; S \mid X_{\text{base}}, G)$, using $H(Y \mid S, G) = H(Y \mid S, G, X_{\text{base}}) = H(Y \mid S, G, X_{\text{base}}, Z)$ from the Markov chain. □

A.2 PROOF OF THEOREM 1

Part (i). ($\Leftarrow$) If $p(y \mid Z, G) = p(y \mid S, G)$ a.s., then for any loss ℓ and any action a, $\mathbb{E}[\ell(a, Y) \mid Z, G] = \mathbb{E}[\ell(a, Y) \mid S, G]$ a.s. (both are integrals of $\ell(a, \cdot)$ against the same posterior), so the conditional Bayes envelopes $a \mapsto \inf_a \mathbb{E}[\ell(a, Y) \mid \cdot]$ coincide pointwise a.s.; taking expectations, the infimal risks are equal, $R^*_\ell(Z, G) = R^*_\ell(S, G)$, for every ℓ (no minimizer need exist, we argue via the infima, not attained optimal actions). ($\Rightarrow$) Suppose $p(y \mid Z, G) \neq p(y \mid S, G)$ on a set of positive probability. Because $\mathcal{Y}$ is standard Borel, there is a measurable set $B \subset \mathcal{Y}$ with $q_Z := \Pr(Y \in B \mid Z, G) \neq q_S := \Pr(Y \in B \mid S, G)$ on a positive-probability event A. Take the bounded loss $\ell_B(\hat{q}, Y) = (\hat{q} - \mathbf{1}_{Y \in B})^2$ (Brier loss on the event B), for which the Bayes rule given information W is $\Pr(Y \in B \mid W)$ with risk $\mathbb{E}[\mathrm{Var}(\mathbf{1}_{Y \in B} \mid W)]$. The key step is $q_Z = \mathbb{E}[q_S \mid Z, G]$. This does *not* follow from any inclusion $\sigma(Z, G) \subseteq \sigma(S, G)$, which fails in general because Z is built from a noisy X_{base}; it follows by conditioning and then applying $Y \perp Z \mid (S, G)$:

$$\Pr(Y \in B \mid Z, G) = \mathbb{E}\big[\Pr(Y \in B \mid S, Z, G) \,\big|\, Z, G\big] = \mathbb{E}\big[\Pr(Y \in B \mid S, G) \,\big|\, Z, G\big].$$

With that identity, the law of total variance gives

$$R^*_{\ell_B}(Z, G) - R^*_{\ell_B}(S, G) \;=\; \mathbb{E}\big[(q_S - q_Z)^2\big] \;>\; 0 \quad \text{on } A,$$

so decision deficiency is strictly positive for ℓ_B, contradicting deficiency zero for all bounded losses. Note the action here is a probability estimate $\hat{q} \in [0, 1]$, not an element of $\mathcal{Y}$, this is why the theorem statement quantifies over bounded measurable decision problems with arbitrary action spaces (equivalently, bounded proper scoring rules for events of Y). On the relation to Blackwell: $p(y \mid Z, G) = p(y \mid S, G)$ a.s. says the experiment $(Z, G) \mapsto p(y \mid z, g)$ is *posterior-sufficient for the induced task experiment at the given prior*; classical Blackwell equivalence (Blackwell, 1953) quantifies over all priors and all decision problems, so our condition is

the task-restricted, fixed-prior analogue, not Blackwell equivalence itself. **Part (ii).** Under log loss over posteriors for Y, the Bayes-optimal predictor given W is $p(y \mid w)$ with risk $H(Y \mid W)$. Then $\Delta_{\log} = H(Y \mid Z, G) - H(Y \mid S, G)$. By $Y \perp Z \mid (S, G)$, $H(Y \mid S, G) = H(Y \mid S, G, Z)$, whence $\Delta_{\log} = H(Y \mid Z, G) - H(Y \mid S, G, Z) = I(Y; S \mid Z, G)$. **Part (iii).** For any bounded loss with $\sup|\ell| \le L$ and any fixed action, changing the predictive distribution from $p(y \mid Z, G)$ to $p(y \mid S, G)$ changes the conditional risk by at most $2L\,\mathrm{TV}(p(\cdot \mid S, G), p(\cdot \mid Z, G))$; optimizing actions on both sides and taking expectations yields the first inequality. The second follows from Pinsker's inequality and Jensen's inequality applied to the concave map $t \mapsto \sqrt{t}$:

$$\begin{aligned}\mathbb{E}\,\mathrm{TV}\big(p(y \mid S, G),\, p(y \mid Z, G)\big) &\le \mathbb{E}\sqrt{\tfrac{1}{2} D_{\mathrm{KL}}\big(p(y \mid S, G) \| p(y \mid Z, G)\big)} \\ &\le \sqrt{\tfrac{1}{2}\,\mathbb{E}\, D_{\mathrm{KL}}(\cdot\|\cdot)} = \sqrt{\tfrac{1}{2} I(Y; S \mid Z, G)},\end{aligned}$$

using $\mathbb{E}\, D_{\mathrm{KL}}\big(p(y \mid S, G) \| p(y \mid Z, G)\big) = I(Y; S \mid Z, G)$, which holds because $q_Z = \mathbb{E}[q_S \mid Z, G]$ (the mixture form of conditional MI). The Fano direction is Appendix A.8. **Part (iv).** The floor claim is Proposition 1 plus part (ii). For the strict-reduction claim: under log loss, $R^*_{\log}(Z, O, G) = H(Y \mid Z, O, G) = H(Y \mid Z, G) - I(Y; O \mid Z, G) < H(Y \mid Z, G) = R^*_{\log}(Z, G)$ whenever $I(Y; O \mid Z, G) > 0$. □

A.3 Proof of Corollary 1

Immediate from Theorem 1(ii) with the definition $C_{\mathrm{info}}(Z; Y \mid G) = \Delta_{\log}$. □

A.4 Proof of Proposition 2

By assumption $P_{1:K} \perp (S, Y) \mid (Z, G)$ and $U = f(Z, G, P_{1:K})$ with f's internal randomness independent of (S, Y). Hence $(P_{1:K}, U)$ is a measurable function of variables that are jointly independent of (S, Y) given (Z, G), so $(P_{1:K}, U) \perp (S, Y) \mid (Z, G)$, i.e. $(S, Y) \leftrightarrow (Z, G) \leftrightarrow (P_{1:K}, U)$ is a Markov chain. Conditional independence of U from the pair (S, Y) gives both marginal statements directly: $I(S; U \mid Z, G) = 0$ and $I(Y; U \mid Z, G) = 0$. (The weaker hypothesis $P_{1:K} \perp S \mid (Z, G)$ alone does *not* suffice: with (Z, G) degenerate, $S \perp Y$ and $P = Y$ one has $P \perp S$ yet $I(Y; P) > 0$.) □

A.5 Proof of Proposition 3

(i) For $\delta \in \mathcal{O}_{\mathrm{comp}}$ the bound $R_\ell(\delta) \ge R^*_\ell(Z, G)$ is the first nonnegativity argument in the proof of Proposition 1. The second and third terms of equation 2 are functionals of $(Z, X_{\mathrm{base}}, S, G)$ and the loss only (δ does not appear in them), so they are invariant across $\mathcal{O}_{\mathrm{comp}}$. **(ii)** Admissible re-encodings satisfy $\sigma(Z, G) \subseteq \sigma(Z', G) \subseteq \sigma(X_{\mathrm{base}}, G)$. For any such Z', every (possibly randomized) rule measurable w.r.t. $\sigma(Z', G)$ is also measurable w.r.t. $\sigma(X_{\mathrm{base}}, G)$, hence $R^*_\ell(Z', G) \ge R^*_\ell(X_{\mathrm{base}}, G)$. The choice $Z' = X_{\mathrm{base}}$ is admissible ($\sigma(Z, G) \subseteq \sigma(X_{\mathrm{base}}, G)$ holds because Z is $\mathcal{F}_0$-measurable), and attains the bound, giving the stated infimum. The observation term $R^*_\ell(X_{\mathrm{base}}, G) - R^*_\ell(S, G)$ does not involve Z', so it is invariant across $\mathcal{O}_{\mathrm{fuse}}$. **(iii)** Under log loss $R^*_{\log}(X_{\mathrm{base}}, O^d, G) = H\big(Y \,\big|\, X_{\mathrm{base}}, O^d, G\big) = H(Y \mid X_{\mathrm{base}}, G) - I\big(Y; O^d \,\big|\, X_{\mathrm{base}}, G\big)$, so the floor strictly decreases iff $I\big(Y; O^d \,\big|\, X_{\mathrm{base}}, G\big) > 0$; if O^d is $\mathcal{F}$-measurable this vanishes, which is Proposition 2 applied at $\mathcal{F}$. The "only if" also holds for an arbitrary Y-decision loss, not just log loss: if $I\big(Y; O^d \,\big|\, X_{\mathrm{base}}, G\big) = 0$ then $p(y \mid X_{\mathrm{base}}, O^d, G) = p(y \mid X_{\mathrm{base}}, G)$ a.s., so the Bayes act and hence the Bayes risk of every such loss are unchanged. **Separation.** It remains to exhibit, for each class, an instance where its own term vanishes and another is positive. (a) Take $Z = X_{\mathrm{base}} = S$ and Y a hard function of S: representation and observation regret are 0, computational regret is positive for a finite-budget decoder. (b) Take $X_{\mathrm{base}} = (E_1, E_2)$ with E_1, E_2 i.i.d. uniform bits, $S = X_{\mathrm{base}}$, $Y = E_1 \oplus E_2$, and let the interface retain one channel, $Z = E_1$: observation regret is 0 (since Y is a function of X_{base}) while representation regret is $H(Y \mid Z) - H(Y \mid X_{\mathrm{base}}) = \ln 2 > 0$. We stress what this example does *not* say: a per-channel encoding $Z = (h_1(E_1), h_2(E_2))$ with h_i injective preserves $\sigma(X_{\mathrm{base}})$, so its representation regret is 0 and an unrestricted Bayes decoder recovers Y. Positive representation regret requires the interface to actually discard the joint, not merely to present it channelwise. (c) Take $Z = X_{\mathrm{base}}$ a single view of an aliased state: computational and representation regret are 0 while observation regret is positive. Any policy confined to one class leaves the binding term untouched in the corresponding instance, hence is strictly suboptimal there. □

A.6 DECISION-SUFFICIENCY (§E.7)

If $Y \perp E \mid (Z, G, U)$ then $p(y \mid z, g, u, e) = p(y \mid z, g, u)$, so the Bayes rule given (Z, G, U) achieves the risk of the Bayes rule given (Z, G, E, U), which is at most $R^*(Z, G, E)$; the reverse inequality holds because U is a stochastic function of (E, Z, G) (simulation argument as in Proposition 1). Under log loss, $H(Y \mid Z, G, U) = H(Y \mid Z, G, E)$ gives $I(Y; U \mid Z, G) = I(Y; E \mid Z, G)$. □

A.7 PROOF OF PROPOSITION 4

Immediate from the definition of conditional mutual information: $I\big(Y; O^d \,\big|\, Z, G, U, d\big) = H(Y \mid Z, G, U, d) - H\big(Y \,\big|\, Z, G, U, O^d, d\big)$, and $H(Y \mid Z, G, U, d) = H(Y \mid Z, G, U)$ under the stated condition (either d is a measurable function of (Z, G, U), or its randomization is independent of Y given (Z, G, U)). The second term is the expected post-probe log-loss Bayes risk. The identity assumes the probe does not alter the state relevant to Y (or that its effect is folded into the observation model); probes that change dynamics require the sequential treatment of Appendix A.9. □

A.8 0–1 LOSS BOUNDS

All entropies and mutual informations in this subsection are in nats (natural logarithm), consistent with the main text. **Deficiency lower bound (Theorem 1(iii), Fano direction).** Suppose Y is determined by (S, G) and takes M values. Then $R^*_{0\text{-}1}(S, G) = 0$, so $\Delta_{0\text{-}1} = P^*_e(Z, G)$, the Bayes 0–1 error from (Z, G). Fano's inequality gives $H(Y \mid Z, G) \le h(P^*_e) + P^*_e \ln(M - 1) \le \ln 2 + P^*_e \ln M$, and since $H(Y \mid S, G) = 0$ implies $H(Y \mid Z, G) = I(Y; S \mid Z, G)$, rearranging yields $\Delta_{0\text{-}1} \ge [I(Y; S \mid Z, G) - \ln 2]/\ln M$; since the bound can be vacuous we state it as $\Delta_{0\text{-}1} \ge \max\{0, [I(Y; S \mid Z, G) - \ln 2]/\ln M\}$. **Post-probe bounds.** For Y uniform on a decision space of size M, any estimator from (Z, G, U, O^d) has error $P_e \ge [H(Y \mid Z, G, U) - I\big(Y; O^d \,\big|\, Z, G, U, d\big) - \ln 2]/\ln M$ (Fano, in nats; the familiar "-1" form holds when all quantities are in bits). The gain in Bayes accuracy from observing O^d is at most the expected total-variation distance between prior and posterior over Y, which by Pinsker's inequality and Jensen (concavity of $\sqrt{\cdot}$) is at most $\sqrt{\frac{1}{2} I(Y; O^d \mid Z, G, U, d)}$ (the same argument as Theorem 1(iii)). Full derivations follow Cover & Thomas (2006) and Polyanskiy & Wu (2025).

A.9 SEQUENTIAL PROBE SELECTION

Full statement of the active-expansion section. When the decision remains ambiguous, the agent executes a *diagnostic action* d (glance at the wrist camera, crop, lift the gripper to test whether the object follows, re-query history), yielding O^d. The controller selects d at the *current* realized state (z, g, u), so we work with the statewise conditional Bayes risk $r_\ell(z, g, u) = \inf_a \mathbb{E}[\ell(a, Y) \mid z, g, u]$ and the statewise value of information

$$\mathrm{VOI}_\ell(d \mid z, g, u) \;=\; r_\ell(z, g, u) \;-\; \mathbb{E}_{O^d \mid z, g, u, d}\, r_\ell(z, g, u, O^d, d), \tag{7}$$

whose log-loss special case is the statewise *decision-relevant information gain*

$$\mathrm{DRIG}(d \mid z, g, u) \;=\; \mathbb{E}_{O^d \mid z, g, u, d}\, D_{\mathrm{KL}}\big(p(Y \mid z, g, u, O^d, d) \,\|\, p(Y \mid z, g, u)\big), \tag{8}$$

which, unlike generic information gain $\mathbb{E}\, D_{\mathrm{KL}}(p(S \mid \cdot, O^d) \| p(S \mid \cdot))$ (Lindley, 1956), counts only bits that can move the optimal decision, the decision-theoretic value of information in the sense of Howard (1966). Algorithm 1 ranks probes by $\mathrm{DRIG}(d \mid z, g, u)$ at the realized state; averaging over states recovers the global identity $\mathbb{E}_{Z,G,U}[\mathrm{DRIG}(d \mid Z, G, U)] = I\big(Y; O^d \,\big|\, Z, G, U, d\big)$ used in Proposition 4.

Proposition 4 (Value-of-information identity, state-averaged)**.** *Suppose the choice of d is a measurable function of (Z, G, U), or randomized with randomness independent of Y given (Z, G, U). Then under log loss the state-averaged VOI equals the conditional mutual information, $\mathbb{E}_{Z,G,U,d}[\mathrm{VOI}_{\log}(d \mid Z, G, U)] = \mathbb{E}_{Z,G,U,d}[\mathrm{DRIG}(d \mid Z, G, U)]$ (the expectation includes the randomization of d):*

$$H(Y \mid Z, G, U) - \mathbb{E}_{O^d}\big[H\big(Y \,\big|\, Z, G, U, O^d, d\big)\big] \;=\; I\big(Y; O^d \,\big|\, Z, G, U, d\big)\,. \tag{9}$$

This is the classical value-of-information identity (Howard, 1966) for the decision target, not a new result; probe selection maximizes the statewise Lagrangian $\mathrm{DRIG}(d \mid z, g, u) - \lambda\, c(d)$, the myopic one-step optimum for log-loss (no claim for multi-step or 0–1 reward). Its novelty is placement inside the same allocator as sampling. For 0–1 loss over M decisions, Pinsker gives (per state) $\Delta\mathrm{Acc}(d \mid z, g, u) \le \sqrt{\frac{1}{2}\,\mathrm{DRIG}(d \mid z, g, u)}$ (Appendix A): a low-DRIG probe cannot help much even if generic uncertainty is high, the reason to prefer DRIG over entropy heuristics. All acquisition uses one criterion, the Lagrangian score $s(a) = \widehat{\Delta R}(a) - \lambda c(a)$; the agent *closes* when $\max_a s(a) \le 0$, else *reopens* by acquiring evidence of positive priced value; for sequential probes, under the conditional-independence and adaptive-submodularity conditions stated in Appendix A.9, greedy selection inherits the standard $(1-1/e)$ approximation guarantees (Golovin & Krause, 2011; Krause & Guestrin, 2005); otherwise we treat it as a heuristic.

The decomposition makes one qualitative, model-free prediction: an *information-limited regime* where acquisition's marginal value dominates computation's (one relevant observation outvalues an order of magnitude more samples) and a *compute-limited regime* where the reverse holds. We posit no parametric "scaling law", the axes interact (irrelevant context inflates computational burden, §6.2), so an additively separable model would mislead. Section 6 maps both regimes directly.

We claim no new approximation guarantee for greedy probe selection. For completeness: if (a) probe outcomes are conditionally independent given (Y, Z, G), and (b) the expected marginal DRIG of any probe is non-increasing in the set of already-executed probes under every realization (adaptive submodularity, which condition (a) alone does not imply and which must be checked per environment), then the DRIG objective is adaptive monotone and adaptive submodular, and the results of Golovin & Krause (2011) give a $(1 - 1/e)$ guarantee for cardinality budgets; for knapsack (cost) budgets the guarantee degrades and requires either the standard partial-enumeration variant or a benefit-cost greedy with a $\frac{1}{2}(1 - 1/e)$-type constant. In InfoGap-Synthetic we verify condition (b) numerically per task family; on non-synthetic VLM/VLA domains we treat greedy DRIG/cost as a heuristic and report its regret against exhaustive probe search where enumerable.

B Continuous-action extension

For continuous Y (e.g., action chunks) in a metric space $(\mathcal{Y}, \rho)$ that is *totally bounded* (or, without loss for our simulators, with Y restricted to a compact subset), and with $\ell(\hat{y}, y) = h(\rho(\hat{y}, y))$ for h Lipschitz with constant L_h: fix $\epsilon > 0$. Total boundedness guarantees a *finite* measurable partition $\mathcal{P}_\epsilon$ of $\mathcal{Y}$ into cells of diameter at most ϵ; let $Y^{(\epsilon)}$ be the index of the cell containing Y. All identities in §3 hold for the discrete variable $Y^{(\epsilon)}$, and the induced risks satisfy $|R^*_\ell(W) - R^*_{\ell,\epsilon}(W)| \le L_h\,\epsilon$ for every information source W, so the deficiency of §3 is recovered up to $2L_h\,\epsilon$ as $\epsilon \to 0$. Choose the partitions *nested* and refining ($\mathcal{P}_{\epsilon_{k+1}}$ refines $\mathcal{P}_{\epsilon_k}$, $\epsilon_k \downarrow 0$) so that $\sigma(Y^{(\epsilon_k)}) \uparrow \mathcal{B}(\mathcal{Y})$ generates the Borel σ-algebra; then $I\big(Y^{(\epsilon_k)}; S \,\big|\, Z, G\big)$ is non-decreasing in k and converges to $I(Y; S \mid Z, G)$ by monotone convergence of mutual information along generating filtrations (finite whenever the limit is). The choice of ϵ trades discretization bias in the risk against estimation variance in the MI proxy; we report sensitivity to cover granularity here.

C InfoGap generation details

Average success on standard suites conflates the two regret terms. InfoGap controls the information ceiling by construction, across three domains.

InfoGap-Synthetic. Exact POMDPs where every mutual information and Bayes risk is computable in closed form. *Task A (state aliasing)*: distinct latent states emit identical observations but demand different actions. *Task B (relational XOR)*: each evidence source alone gives 0 bits; the joint relation gives 1. *Task C (irrelevant-information trap)*: probe A reveals 3 bits about the latent state but none about the action; probe B reveals 1 bit that determines the action, generic EIG picks A, DRIG must pick B. *Task D (compute-limited)*: the state is sufficient but search is hard, so sampling *should* beat acquisition; a controller that always expands information fails here.

InfoGap-Vision. Procedurally generated two-view problems (E_1, E_2, Y) with exact ground truth: unlike existing multi-image suites (Fu et al., 2024; Meng et al., 2024; Wang et al., 2024), a constraint

solver verifies by enumeration that single-view Bayes accuracy is ≤ 0.6 and joint-view ≥ 0.95 for every instance. Main categories: *occlusion complementarity*, *before/after transition* (token on a circular track), *relational XOR*; plus *targeted probe* (irrelevant-information trap: a context view points at one of four boxes, only the pointed-at box is decision-relevant) and *occlusion-leak* (a parametric family with single-view Bayes accuracy tunable to any p, §6.2). Runs use 500 instances/main-category/cell.

InfoGap-Embodied. Information-channel interventions on LIBERO (Liu et al., 2023) with OpenVLA-OFT (Kim et al., 2025) as frozen base policy. The wrist camera disambiguates grasp/contact states; holding policy, compute, and initialization fixed, we vary only the information set: *full*, *masked* (wrist blanked), and matched-distribution controls (§E.8). A state-intervention API (92-d simulator state) supports paired inits; rollouts follow the official protocol (all four suites $\times$ 10 tasks $\times$ 10 inits). Controlled simulation isolates the information channel while keeping the policy and dynamics fixed, complementing sim-based policy evaluation (Li et al., 2024).

Synthetic. Each POMDP is specified by a factored latent state $S \in \{0,1\}^n$ ($n \leq 12$), an emission kernel per evidence channel ($|E_i| \leq 8$), and a decision map $Y = g(S)$. Exact posteriors and all conditional MI quantities are computed by enumeration; the identity $I(Y;S \mid Z,G) = R^*_{\log}(Z,G) - R^*_{\log}(S,G)$ is asserted to 10^{-9} on every generated instance. Task instances are sampled to hit target ceilings $C_{\text{info}} \in \{0.2, 0.5, 0.8\}$ nats; for the binary-label families (relational XOR, trap) the maximum achievable ceiling is $H(Y) = \ln 2 \approx 0.693$ nats, so the 0.8 target is clipped to $\ln 2$ there (the aliasing family, with four labels, reaches 0.8 exactly). Task B additionally includes a 3-source parity variant and an *additive control* ($Y = (Y_1, Y_2)$, $E_i = Y_i$) in which the joint evidence advantage is positive but the relational gap is zero, used by E7 to attribute gains to the correct mechanism. **Vision.** Scenes are rendered procedurally (deterministic given category and instance seed; two 512^2 views per instance). Each category's constraint solver verifies by exact enumeration over the latent scene factors that single-view Bayes accuracy is ≤ 0.6 and joint-view accuracy is ≥ 0.95. One design note: the before/after category uses a *circular* track, a linear track leaks the movement direction at its endpoints, pushing single-view accuracy above the cap; on a ring every slot has both neighbors, making single-view exactly 0.5 while the ordered pair is deterministic. Prompts describe the two-view task structure (occlusion semantics, cross-view combination, ring arithmetic); we verified that joint-view accuracy reaches ≈ 1.0 under these prompts while single-view remains at chance under the *same* prompts, so the measured gap is informational, not a prompting artifact. **Embodied.** Information-channel interventions on LIBERO: the wrist camera input of OpenVLA-OFT is either passed through (full), blanked (masked), or replaced with a noise image of matched shape (noise control). A state-intervention API exposes the 92-dimensional simulator state (object poses via qpos addressing) for paired initializations; rollouts follow the official evaluation protocol (settle steps, 220-step horizon). All four suites $\times$ 10 tasks $\times$ 10 initializations per condition.

D Implementation details

Models. Qwen3-VL-8B (primary), Qwen3-VL-4B, Gemma-4-12B served via vLLM (temperature 0.7 for sampling methods, 0 for single inference; answers extracted by a constrained format with $\geq 95\%$ parse rate). Embodied policy: `openvla-7b-oft-finetuned-libero-*` checkpoints, frozen, run in the canonical OpenVLA-OFT environment. The relational encoder for the fuse branch is trained with the conditional-IB objective (used as a training objective for the injected state, not as a regularizer on the frozen base model):

$$\mathcal{L} = \underbrace{-\log p(Y \mid Z,U,G)}_{\mathcal{L}_{\text{decision}}} + \lambda_{\text{f}} \underbrace{[-\log p(O_{t+1} \mid Z,U,A_t)]}_{\mathcal{L}_{\text{future}}} + \lambda_{\text{IB}} D_{\text{KL}}\big(q_\varphi(U \mid E,Z,G) \,\|\, p(U \mid Z,G)\big) + \lambda_{\text{ax}}\, \mathcal{L}_{\text{axis}}. \tag{10}$$

Relation encoder. 4-layer set-attention module (Lee et al., 2019), 16 stochastic latent tokens of dimension 64, trained with equation 10, λ_{IB} annealed $10^{-4} \to 10^{-2}$; the KL rate term converges and the λ_{IB} sweep of Figure 5a uses the exact per-instance KL as the rate on synthetic tasks. **DRIG predictor.** On synthetic: a 2-layer MLP trained to rank probes against exact decision-relevant gain; held-out Spearman $\rho = 0.78$. In the live VLM loop (targeted probe): a probe-selection network trained on context geometry with held-out probe-selection accuracy 1.0; no VLM query and no hand-written rule is used at selection time. **Costs.** We log the generated-token budget per call as 64

tokens per sampled answer (a fixed decoding cap), so a method's cost is $64K$ for K samples ($+64$ for a Best-of-K verifier pass); Table 1's "tokens" column is this configured budget. Under it the compute axis spends strictly more than acquisition (1088 for Best-of-16 vs. 64 for one acquire-then-answer pass, a $17\times$ gap) yet stays at chance. This budget counts *generated output tokens only* (not input/visual-encoding tokens, VLM-call count, or wall-clock); environment steps and wall-clock are logged separately for embodied runs.

D.1 Controller cost model, utility, and splits (§6.3)

Cost. Every policy pays the same two $K{=}2$ *pilot* samples up front; the cost c of a trajectory is a *normalized generated-token count* (pilot + the generated tokens of the chosen arm), scaled so the fixed-compute reference (Best-of-16) has cost 1.00. In these units the three arms cost compute 1.00, acquire 0.35 (pilot + one complementary view + one greedy re-inference), stop 0.125 (pilot only). These per-arm costs are *configured*, not measured per instance: each arm's generated-token budget is fixed by its construction, so `cost_norm` is constant across instances for a given arm (the released label files show exactly $1.00/0.35/0.125$). Variation across policies therefore comes from *which* arm is selected, not from per-instance token variation. This cost counts *generated output tokens only*. It excludes input processing, visual encoding, acquisition latency, and wall-clock time, and both repeated sampling and acquisition incur input and vision-encoding work (sampling re-encodes the prompt per call; acquisition processes an extra image). It should therefore not be read as total serving cost, and the sign of the total-cost difference between the two axes is *not identified* by this metric alone. We therefore report it explicitly as "normalized generated-token cost," one axis of cost rather than the whole; Hariri et al. (2026) make the same point that test-time comparisons are sensitive to which inference regime and cost axis is reported. Table 2 states every cost axis we can and cannot account for. Model calls follow from the arm definitions above and favour acquire even more strongly than tokens do (3 calls vs. 18); per-call input and vision encoding, which favour compute, were never logged (the `costs_flops_est`, `costs_env_steps` and `costs_wallclock_s` fields are identically zero in our runs), so we report them as unmeasured rather than estimate them. The two unmeasured axes push in opposite directions, which is exactly why the sign of the total-cost difference is not identified.

Table 2: **Cost accounting, and its limits.** Generated tokens are measured per run. Model calls are *derived* from the arm definitions ($K_0{=}2$ pilot samples for every policy; compute adds 16 fresh samples, acquire adds one greedy re-inference, stop adds none), not separately logged. The remaining axes were not instrumented; we list them so the omission is explicit rather than implied.

Cost axis	compute	acquire	stop	**status**
Normalized generated tokens	1.00	0.35	0.125	measured per run
Model calls	18	3	2	derived from arm definitions
Images encoded per call	1	≤ 2	1	derived; favours compute
Input/prompt tokens	—	—	—	not logged
Vision-encoder FLOPs	—	—	—	not logged (field is zero)
Wall-clock latency	—	—	—	not logged (field is zero)

Utility and the hindsight-oracle gap. The routing label / utility is $u = \text{success} - \lambda c$ with $\lambda = 0.1$ fixed on validation before any test evaluation. The *hindsight oracle* picks, per instance, the arm with the highest *realized* u over the cached arm outcomes; it is an *optimistic upper reference*, not a Bayes-optimal (expected-utility) oracle, and the validation labels inherit the same realized-best definition, adding label noise. At $\lambda{=}0.1$ the best fixed policy is Best-of-16 on *both* accuracy (0.748) and utility (0.648); the router reaches 0.938/0.886 and the hindsight oracle 0.963/0.931, i.e. 88% of the accuracy gap $\frac{0.938-0.748}{0.963-0.748}$ and 84% of the utility gap $\frac{0.886-0.648}{0.931-0.648}$. The ordering is stable for $\lambda \in [0.05, 0.3]$; at large λ the cheap stop arm becomes competitive and the fraction shifts. **The router is effectively a lookup table.** The six features take only 30 distinct values over the 1800 test instances, so any consistent learner converges to a per-cell decision: an explicit per-cell utility lookup fit on validation scores 0.937/0.885 and agrees with the balanced random forest on 99.7% of instances, while an *unweighted* forest collapses onto the majority action and scores 0.746. We therefore do not claim a learned model is needed, only that matching the operation to the diagnosed

Table 3: **The regret terms dissociate on InfoGap-Synthetic** (accuracy). Bayes risks and mutual informations *within* each generated task are computed by exact enumeration (no sampling error); the reported 95% CIs reflect variation *across* sampled task instances of each family. Sampling (Best-of-K, K=64) helps only the compute-limited family; exact-DRIG probing (B=1–2) helps only the information-limited families; the ReLATE-lite controller *nearly* matches the better axis on each (a small residual routing error, e.g. 0.973/0.987 vs. 1.000).

Method	A (aliasing)	B (rel. XOR)	C (trap)	D (compute-lim.)
Single inference	0.900 [0.881, 0.918]	0.961 [0.954, 0.968]	0.955 [0.942, 0.967]	0.494 [0.463, 0.525]
Best-of-K	0.900 [0.881, 0.918]	0.961 [0.954, 0.968]	0.955 [0.942, 0.967]	**1.000** [1.000, 1.000]
Exact-DRIG probe	**1.000** [1.000, 1.000]	**1.000** [1.000, 1.000]	**1.000** [1.000, 1.000]	0.494 [0.463, 0.525]
ReLATE-lite (controller)	**1.000** [1.000, 1.000]	0.973 [0.971, 0.975]	**1.000** [1.000, 1.000]	0.987 [0.984, 0.989]

bottleneck is what produces the gain, and that with these features the mapping is small enough to tabulate. Closing the residual 0.025 requires richer diagnostics, not a stronger classifier. All numbers, CIs, the confusion matrix and per-instance actions are produced by `train_router_v2.py` from `r8_labels_{val,test}_v2.jsonl`. **Splits.** The router is trained on a *separate* validation pool (n_{val}=900: 120+120+120+360+180 per category) whose instance seeds are disjoint from the test pool (n_{test}=1800: 240×3 + 720 + 360); the router (a 300-tree random forest with balanced class weights, `random_state`= 0) and the uncertainty-heuristic threshold are fit on validation only, and every number in Table 4/Figure 4 is on the held-out test pool. Labeling both pools exhaustively (three arms each) is $(900 + 1800) \times 3 = 8{,}100$ arm executions; the test pass then *reuses* the cached test-arm outcomes so that all six comparators c1–c6 (including the oracle) are scored on identical per-instance arm results, no arm is re-run at test time. **Confound.** Category identity and oracle labels are withheld from the router features (enforced by a schema test that fails if either is present); whether the six features nonetheless separate task family (so that routing partly reduces to family classification) is a confound we do not fully rule out and discuss in the §6.3 limitations.

E Cross-model transfer and additional results

E.1 Exact ceilings on synthetic POMDPs

On the compute-limited family Best-of-K moves $0.494 \to 1.000$ while exact-DRIG probing stays at 0.494; on information-limited families the ordering reverses. This is the dissociation Proposition 3 requires, exact by enumeration.

On InfoGap-Synthetic every Bayes risk and conditional MI is exact by enumeration, so the ceilings are known rather than estimated (Table 3, Appendix E); since $Z = X_{\text{base}}$ here, representation regret is zero by construction and the deficiency is pure observation regret, invariant to the sample budget. The terms dissociate exactly as Proposition 3 requires: on the compute-limited family Best-of-K rises $0.494 \to 1.000$ (K=1 to 64) while exact-DRIG probing without extra samples stays at 0.494, on the information-limited families sampling is flat in K while probing solves them, and ReLATE-lite routes to the better arm on every family with a small residual routing error (0.973 and 0.987 vs. 1.000 for the oracle arm). This is a validity check on the theory, not independent evidence.

E.2 Router qualifications in full

Per-regime behaviour and the cost it does not save. The router selects the predicted operation in each regime: information-limited 1.000 (vs. 0.546 for fixed compute), compute-limited 0.865 (vs. 0.093 for always-acquire), saturated 0.958. It captures 88% of the accuracy gap and 84% of the utility gap to the hindsight oracle (the best realized arm per instance, an optimistic reference). On the saturated third it is *not* the cheapest policy: it splits between acquire (208) and stop (152) at cost 0.255, so always-stop attains slightly higher utility there (0.938 vs. 0.933) despite lower success, over-acquiring where stopping would do.

Three qualifications. (i) Class balance matters: an unweighted learner collapses onto stop and scores 0.746, no better than Best-of-16; balanced weighting gives 0.938. We report both. (ii) No

Table 4: **Bottleneck matching vs. fixed strategies** (held-out test pool, n_{test}=1800; router fit on a disjoint n_{val}=900). Utility = success − 0.1 cost on normalized *generated-token* cost; Δ is the router's success advantage (95% CI, instance bootstrap). The *within-cell oracle* is the best action per distinct feature value on this cached sample, empirical for these features, not an absolute ceiling. The *regime-only baseline* is the task-family confound control; here the per-category argmax picks the same action in every category, so a category-only baseline coincides with it exactly. Arm outcomes are cached, and rows are emitted from `r8_router_v2_report.json` by `emit_table3.py` (Appendix E).

Policy	success	gen.-tok. cost	utility	Δ vs. router (95% CI)
Fixed compute (Best-of-16)	0.748	1.000	0.648	$+0.189$ $[+0.171, +0.209]$
Always acquire	0.629	0.350	0.594	$+0.309$ $[+0.288, +0.331]$
Always stop	0.573	0.125	0.560	$+0.365$ $[+0.341, +0.389]$
Random (3-way)	0.657	0.503	0.607	$+0.281$ $[+0.259, +0.302]$
Uncertainty heuristic	0.527	0.722	0.455	$+0.411$ $[+0.388, +0.434]$
Regime-only baseline	0.931	0.610	0.870	
Bottleneck router (ours)	**0.938**	0.513	**0.886**	—
same, unweighted	0.746	0.326	0.714	$+0.192$ $[+0.172, +0.212]$
per-cell lookup	0.937	0.515	0.885	$+0.001$ $[+0.000, +0.003]$
Within-cell oracle (cached test sample)	0.938	—	0.886	-0.000
Hindsight oracle	0.963	0.322	0.931	-0.026 $[-0.033, -0.018]$

Table 5: **Leave-one-category-out: no transfer to an unseen bottleneck family.** Router trained on four categories, routing the held-out one; in-distribution is the same router trained on all five. Pooled LOCO success is *below* fixed compute (0.748), with the failure concentrated on *hard-count*, the only compute-limited family, which the router sends to acquire (452/720). Per-category rows in Appendix E; generated by `train_router_v2.py`.

held out	n	LOCO succ	util	in-dist succ	util
four visual families (each)	1080	0.938–1.000	0.905–0.965	0.958–1.000	0.933–0.965
hard-count (compute-limited)	720	**0.299**	0.272	0.865	0.785
pooled	1800	**0.695**	0.666	0.938	0.886

learned model is required, the six features take only 30 distinct values, an explicit per-cell lookup scores 0.937 and agrees with the forest on 99.7% of instances, so the router matches the within-cell oracle (0.938) simply because the mapping is tabulable. That reference is empirical for these features, not an absolute ceiling. We read this as support for the *principle* (matching the operation to the diagnosed bottleneck is what produces the gain), and not as evidence for a sophisticated controller; closing the residual 0.025 needs richer diagnostics, not a stronger classifier. (iii) Decisive for how this should be read: *what the router learns is task-family structure, not per-instance bottleneck diagnosis.* (a) A *category-only* baseline applying one fixed action per category, fit on validation, already reaches 0.931/0.870, within 0.007 success and 0.016 utility of the router. (b) The regime is recoverable from the six features alone at 0.81 accuracy, and the router matches the realized-best action on only 54% of instances. (c) Most tellingly, under *leave-one-category-out* training the router does *not* transfer: pooled success falls to 0.695 (utility 0.666), *below* fixed compute, and on held-out *hard-count* (the only compute-limited family), it collapses to 0.299 (vs. 0.865 in-distribution, oracle 0.908), routing 452/720 compute-limited instances to acquire because that is what the visual families taught it (Table 5, Appendix E). We therefore claim only *controlled task-family routing*: within a known mixture the operation predicted by the decomposition is selected and beats every fixed policy, but generalization to an unseen bottleneck family is not demonstrated. Instance-level diagnosis would require same-renderer counterfactuals in which only the binding term changes; our attempt at that benchmark did not pass its calibration gate (§7).

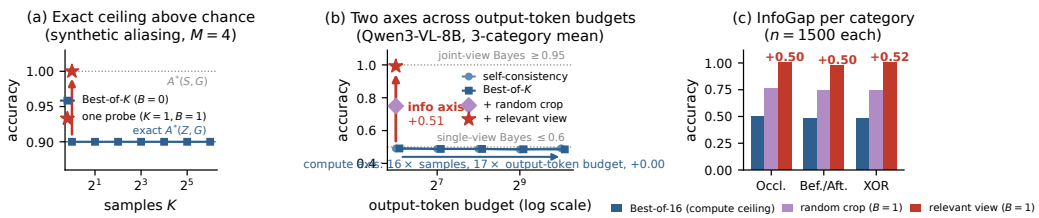


Figure 3: **The ceiling, measured** (all panels plot *accuracy*; $A^* = 1 - R^*_{0\text{-}1}$ denotes Bayes accuracy). (a) The ceiling is not chance: on the synthetic aliasing family (M=4 labels, exact enumeration), Best-of-K is pinned at the exactly computed fixed-representation Bayes accuracy $A^*(Z, G) = 0.90$ from K=1 to 64, while a single probe reaches $A^*(S, G) = 1.0$. (b) The same dissociation on Qwen3-VL-8B (InfoGap-Vision, 3-category mean): the compute axis (16× samples, 17× the generated-token budget) buys $+0.00$ from the single-view Bayes accuracy bound; one decision-relevant view at a smaller budget buys $+0.51$ (star); a random crop (diamond) recovers only part of the gap. (c) InfoGap per category (500 scenes / 7,500 runs each; scene-clustered CIs; scene-level sign test $p < 10^{-75}$, all discordant scenes favoring acquisition).

E.3 Decision-relevant acquisition: details

What a positive DRIG does and does not imply. Under log loss the posterior *is* the action, so DRIG > 0 is exactly the value of the observation. Under a general bounded loss it is not: a posterior may move from $0.8/0.2$ to $0.7/0.3$ with a positive information gain while the Bayes-optimal action is unchanged. We therefore read DRIG as information about the decision target and, for non-log losses, claim only the Pinsker upper bound of Theorem 1(iii) rather than a guaranteed change of action. The targeted-probe family separates *acquiring the right information* from *acquiring more*. Context-only inference (any $K \leq 16$) is 0.540; a random box (B=1) reaches 0.612; a **trained DRIG selector** (a probe-selection net on context geometry, no rule, held-out selection accuracy 1.0) reaches **1.000**, $+0.388$ over generic at identical cost. Opening *all four* boxes reaches only 0.793, worse than the single relevant probe: for a Bayes-optimal agent more information cannot hurt, so this drop is *computational*, irrelevant context inflates the finite model's burden, i.e. the two axes interact, which is exactly why a bottlenecked, decision-targeted interface (§E.7) is needed over maximal acquisition. On synthetic Task C the exact ordering matches (DRIG/oracle solve at one probe, generic/random need two), and the learned DRIG predictor tracks exact gain at Spearman $\rho = 0.78$ on held-out POMDPs, discounting the 3-bit decision-irrelevant trap probe (Figure 5b). A rate–relevance analysis (Figure 5a) confirms a compact decision-sufficient state loses essentially none of the predictive relevance of raw evidence at a fraction of the cost.

E.4 Uncertainty reading of the decomposition

Relation to reducible and irreducible uncertainty. Under log loss the Bayes risks are conditional entropies, so equation 2 and the chain rule give a reading of the agent's total predictive uncertainty in terms of *what could remove each part*:

$$R_{\log}(\delta_K) = \underbrace{H(Y \mid S, G)}_{\text{irreducible}} + \underbrace{\left[R_{\log}(\delta_K) - R^*_{\log}(Z, G)\right]}_{\text{computational}} + \underbrace{I(Y; X_{\text{base}} \mid Z, G)}_{\text{representation}} + \underbrace{I(Y; S \mid X_{\text{base}}, G)}_{\text{observation}}. \tag{11}$$

Each term has a distinct remedy (more inference, fusion, acquisition), except the first, which *none* of them can remove: even a decision-maker who sees the true state faces residual stochasticity in Y. This is why the controller needs a stop action, and why calibrated total uncertainty is not a sufficient trigger for acquiring evidence: high uncertainty is compatible with all four terms. *Caveat on terminology:* the two middle-to-right terms are reducible and thus epistemic *in spirit*, but they are not the parameter uncertainty $p(\theta \mid \mathcal{D})$ usually called epistemic in Bayesian deep learning; they measure task-relevant information the decision-maker lacks *about the latent state* at a fixed model. We therefore call them representation and observation *deficiency* rather than epistemic uncertainty, and note only that equation 11 refines the usual epistemic/aleatoric split by separating the reducible part into a component fixable without acting on the world and a component that requires acting.

E.5 EXTENDED RELATED WORK

The condensed version appears in §2; the paragraphs below give the detailed positioning.

Test-time scaling with fixed representations. Chain-of-thought (Wei et al., 2022), self-consistency (Wang et al., 2023b), Best-of-N with verifiers (Cobbe et al., 2021; Lightman et al., 2024; Wang et al., 2023a; Zhang et al., 2025b; Hosseini et al., 2024), search (Yao et al., 2023a), repeated sampling (Brown et al., 2024), bootstrapped and long reasoning (Zelikman et al., 2022; DeepSeek-AI et al., 2025; Muennighoff et al., 2025), and compute-optimal inference (Snell et al., 2024; Wu et al., 2024) all scale inference over a fixed encoding, echoing training-side scaling laws (Jones, 2021; Kaplan et al., 2020; Hoffmann et al., 2022). Recent theory analyzes their sample complexity and limits (Huang et al., 2025b; Beirami et al., 2025; Huang et al., 2025a; 2024a; Feng et al., 2023; Merrill & Sabharwal, 2024) but stays *within* a representation; our decomposition equation 1 isolates what such methods can never recover. Self-refinement (Madaan et al., 2023) conditions only on model-generated text (consistent with unreliable intrinsic self-correction (Huang et al., 2024b; Kamoi et al., 2024)), so by Proposition 2 it operates under the ceiling, as do verifier ensembles (Saad-Falcon et al., 2025). LatentEvolve (Zhang et al., 2025a) may improve the *internal* use of an existing representation, but it does not distinguish representation deficiency from missing-observation deficiency, diagnose which term binds, or select the corresponding remedy, whereas our decomposition identifies fusion or acquisition as the appropriate operation depending on which deficiency component dominates. Two concurrent works identify ceilings (modal/correlation ceilings of selection (Bay & Yearick, 2026) and reliability ceilings from resolvable context (Majumdar, 2026)), both on the computational axis; our deficiency Δ_ℓ instead bounds what *any* selection could achieve and names fusion or acquisition as the remedy.

Non-computational scaling, and task-dependent sufficiency. That test-time gains can come from something other than more inference is argued qualitatively in several recent works (scaling interaction (Shen et al., 2025b), an internal-vs-external ceiling argued via a data-processing inequality (Li & Shi, 2026), scaling perception (Jiang et al., 2026b), adaptive looking in VLAs (Choi et al., 2026), context as an axis (Shen et al., 2026; Jiang et al., 2026a)), and, closest operationally, diagnosing whether extra sampling suffices or an external prior is needed (Jeong & Yoon, 2026; Wang et al., 2025; Xiong et al., 2026; Ding et al., 2026). On the theory side, Sevetlidis (2026a) defines Bayes-sufficiency for a fixed decision problem and Sevetlidis (2026b) proves that under log loss a frozen representation's excess risk equals the predictive information it discards, the identity our Theorem 1(ii) uses for the Z-vs-X_{base} term, for which we claim no priority. We do not claim the existence of a second axis; we claim its *resolution* into representation and observation, which are repaired by different operations, and an action space derived from that split. Appendix E.5 gives the detailed comparison and Table 7 maps which operations each line of work scales.

Active perception and acquisition. Active/interactive perception (Bajcsy, 1988; Bajcsy et al., 2018; Bohg et al., 2017), value of information (Howard, 1966; Krause & Guestrin, 2005), Bayesian experimental design (Lindley, 1956; Chaloner & Verdinelli, 1995; Foster et al., 2019; Rainforth et al., 2024), active learning (Settles, 2009; Houlsby et al., 2011), POMDP planning (Kaelbling et al., 1998), next-best-view (Chen et al., 2024), and confidence-guided exploration (Ren et al., 2024; 2023) formalize acquiring observations; on VLMs, guided visual search/cropping (Wu & Xie, 2024; Zhang et al., 2023; Shen et al., 2025a; Zhang et al., 2025c), visual tool use (Hu et al., 2024), and retrieval/tools (Lewis et al., 2020; Yao et al., 2023b; Schick et al., 2023; Press et al., 2023) acquire task-relevant evidence, including concurrent active visual reasoning as sequential BED (Liu et al., 2026). Neither acquisition nor decision-relevant (rather than generic) information gain is our novelty, Liu et al. (2026) already cast active visual evidence gathering as sequential BED with a task-relevant gain proxy, and Jeong & Yoon (2026) already diagnose sample-recoverability versus needing an external prior. The structural difference is expressive, not merely operational: a two-way *sample-or-retrieve* controller partitions instances by whether the current information set suffices, so both of its branches are moves along our computational and observation axes and neither can be fuse, which changes the interface at a *fixed* information set. Any policy whose action space lacks that third option leaves representation regret untouched by construction (Proposition 3(ii)), which is why the split into three terms and not two is what makes the action space complete. ReLATE sits *upstream*: it decides, before acquiring, which of three regret terms binds, and its probe ranking is one branch of that allocator rather than a standalone acquisition algorithm (§3.5).

VLA test-time scaling. RoboMonkey (Kwok et al., 2025), value-guided action selection (Nakamoto et al., 2024), VLA reward-model verification (Dai et al., 2025), and embodied test-time scaling (Ye et al., 2026) diversify and select action candidates for VLA policies (Brohan et al., 2023; Octo Model Team et al., 2024; Kim et al., 2024; Black et al., 2025; Kim et al., 2025). These methods improve candidate selection but do not ask why candidates sharing one state abstraction fail together, a failure mode rooted in perceptual aliasing (Whitehead & Ballard, 1991) and lossy state abstraction (Li et al., 2006), and exactly the regime InfoGap isolates.

Information bottlenecks. Our encoder instantiates a conditional information bottleneck (Tishby et al., 1999; Tishby & Zaslavsky, 2015; Alemi et al., 2017; Fischer, 2020; Federici et al., 2020), here a component (not the contribution) deployed at test time to synthesize a decision-sufficient state added to a frozen policy; the decision-sufficiency view descends from Blackwell (1953).

Non-computational test-time scaling axes. That test-time gains can come from something other than more inference is now argued in several places, and we do *not* claim that observation as our contribution. Scaling the *interaction* horizon is proposed as complementary to per-step compute (Shen et al., 2025b); Li & Shi (2026) group reasoning and Best-of-N as "internal" scaling and argue via a data-processing inequality that only external instruments open a new channel; Jiang et al. (2026b) scale *perception* (width/depth of evidence gathering) rather than the answer; Choi et al. (2026) modulate how a VLA looks, not just what it does; and context itself is treated as a scaling axis (Shen et al., 2026; Jiang et al., 2026a). Closest operationally, Jeong & Yoon (2026) diagnose for embodied policies whether extra sampling can recover performance (*recoverable headroom*) or an external prior is needed, and adaptive retrieval controllers choose whether and how much to acquire (Wang et al., 2025; Xiong et al., 2026; Ding et al., 2026). Table 7 (Appendix E) summarizes which operations each line of work scales and whether it routes among them. Our contribution is not the existence of a second axis but its *resolution*: (i) a numbered risk-level decomposition into computational, representation, and observation regret that says *which* operation removes each term, in particular that *fusion* of already-observed evidence can lower the ceiling with no new observation, a distinction absent when the split is internal-vs-external or thinking-vs-interaction; (ii) a characterization of the resulting deficiency (Theorem 1) rather than a channel-capacity argument; and (iii) an action space $\{\mathsf{compute}, \mathsf{fuse}, \mathsf{acquire}\}$ *derived* from that decomposition rather than posited.

Task-dependent sufficiency and excess risk. The representation half of our deficiency has a direct precedent: Sevetlidis (2026a) defines Bayes-sufficiency for a *fixed* supervised decision problem (relevant information is loss-dependent, ordered by refinement of a Bayes quotient), and Sevetlidis (2026b) proves that under log loss the excess risk of a frozen representation equals the predictive information it discards, $I(Y; X \mid H)$, the same identity our Theorem 1(ii) uses for the Z-vs-X_{base} term, and we claim no priority for it. What we add is not that identity but its use as a *test-time* ceiling: we extend it past the representation to *unobserved* state ($I(Y; S \mid Z, G)$, which no representation learning can recover), show it is invariant to the sample budget, and connect it to an acquisition decision.

E.6 Joint evidence and the relational gap

What the measured gap does and does not bound. The relational experiment varies only the decision rule at fixed evidence, so it measures the separable-vs-joint gap, which is what fusion alters in deployment. It lower-bounds representation regret only when the deployed interface genuinely discards the joint: a channelwise but lossless interface has zero representation regret even though late voting still fails on it (Proposition 3(b)).

Let $E_1, \ldots, E_m$ be evidence channels beyond Z (global/wrist cameras, history, predicted next observation, crops). Two quantities matter and must not be conflated:

$$\begin{aligned} \mathrm{JEA} &= I(Y; E_{1:m} \mid Z, G) - \max_i I(Y; E_i \mid Z, G)\,, \\ \mathrm{RelGap} &= \inf_{\delta \in \mathcal{H}_{\text{sep}}} \mathbb{E}\,\ell(\delta, Y) - \inf_{\delta \in \mathcal{H}_{\text{joint}}} \mathbb{E}\,\ell(\delta, Y) \;\geq\; 0. \end{aligned} \tag{12}$$

The *joint evidence advantage* JEA is the extra information from all channels vs. the best one, but $\mathrm{JEA} > 0$ does not certify synergy (independent channels with disjoint bits, $Y = (Y_1, Y_2)$, $E_i = Y_i$, give $\mathrm{JEA} > 0$ with no interaction). The *relational gap* RelGap isolates what *representing relations* buys: $\mathcal{H}_{\text{sep}}$ is an explicitly restricted class, additive logits $\arg\max_y [b_y + \sum_i h_{i,y}(E_i)]$ or

late voting over per-channel predictors, both excluding cross-channel interaction and both $\subseteq \mathcal{H}_{\text{joint}}$, so $\text{RelGap} \geq 0$ by construction. $\text{RelGap} > 0$ says joint structure must be *computed*, not merely collected: on uniform binary XOR $Y = E_1 \oplus E_2$ no rule in the additive class attains zero error (an additive-logit threshold on $E_1 + E_2$ tops out at accuracy $3/4$, and the late-voting protocol used in our experiments stays at chance), whereas joint conditioning is exact. This certifies a gap *for the stated class*, not universal synergy; our experiments use $\mathcal{H}_{\text{vote}}$. PID frames this as synergy (Williams & Beer, 2010; Bertschinger et al., 2014; Griffith & Koch, 2014) but admits competing definitions (Lizier et al., 2018), so we use only these operational quantities. InfoGap contains both regimes (additive: $\text{JEA} > 0$, $\text{RelGap} = 0$; parity: $\text{RelGap} > 0$) so gains attribute to the right mechanism.

E.7 Relational state and the conditional-IB frontier

Forwarding all raw evidence is expensive, so we compress it into a state $U \sim q_\varphi(u \mid E, Z, G)$ under a rate bound. If $Y \perp E \mid (Z, G, U)$ then U is *decision-sufficient* ($R^*(Z, G, U) = R^*(Z, G, E)$; under log loss $I(Y; U \mid Z, G) = I(Y; E \mid Z, G)$), and learning one is a conditional information-bottleneck problem,

$$\min_{q(u|E,Z,G)} I(U; E \mid Z, G) \quad \text{s.t.} \quad I(Y; E \mid Z, G) - I(Y; U \mid Z, G) \leq \epsilon, \tag{13}$$

used here as the training objective for a relational-state encoder whose output is *dynamically injected* at test time, rather than as a regularizer on the frozen base model itself. Its rate–relevance trade-off $F_Z(r) = \sup_{I(U;E \mid Z,G) \leq r} I(Y; U \mid Z, G)$ (Eq. 14, the conditional-IB curve (Tishby et al., 1999; Fischer, 2020)) we call the *relational information frontier*; we claim no novelty for the curve, only its use at test time and the measurement of where practical constructions sit on it (§6).

$$F_Z(r) = \sup_{q(u|E,Z,G):\, I(U;E \mid Z,G) \leq r} I(Y; U \mid Z, G)\,. \tag{14}$$

E.8 Sensitivity to a decision-relevant channel in an embodied policy

Baseline caveat. Our full-observation LIBERO-10 success is 0.26, far below the ≈ 0.95 reported for OpenVLA-OFT with per-suite checkpoints and the official evaluation protocol; we did not reproduce that baseline, so these numbers show *relative* sensitivity to removing a channel and should not be read as an information-ceiling measurement or compared against published LIBERO results.

InfoGap-Embodied intervenes on the information channels of a frozen OpenVLA-OFT policy (single-task smoke sanity check: $10/10$ on LIBERO-spatial; we did not reproduce the published LIBERO-10 baseline). Masking the wrist camera (the channel that disambiguates grasp and contact states), while holding the policy, compute, and initializations fixed collapses success on every suite (Table 6): $0.97 \to 0.07$ on spatial, $0.96 \to 0.09$ on goal, $0.98 \to 0.31$ on object, $0.26 \to 0.00$ on long-horizon LIBERO-10 (n=100 per cell; 800 rollouts). Object degrades least because its targets are largest in the agent view; LIBERO-10 collapses completely.

Additional sampling does not recover the missing channel (R10). Is the collapse an information ceiling or just under-search? Sweeping the action-sample budget $K \in \{1, 2, 4, 8\}$ under masking (spatial and goal, n=100/cell, two selectors) leaves success flat then *worse*: spatial masked accuracy is 0.05–0.14 at K=1 and exactly 0.00 at K=4, 8 vs. 0.99 full-wrist; goal is identical ($0.10 \to 0.00$). Under the tested budgets ($K \leq 8$) and both selectors, more samples over a blind state do not recover performance, and past K=2 they *hurt*, selecting among uniformly-blind candidates amplifies a confident wrong choice (the embodied echo of the all-probes result). This is the action-space analogue of the fixed-state plateau.

Not a blank/noise input artifact (R11). Blank/noise wrists are off-manifold, so their failure could be OOD shock. Three *matched-distribution* controls (spatial, n=50) probe this: a real frame from a *different* episode (*other*, 0.02) and one 40 steps *stale* (0.00) are on-manifold real images yet fail as hard as blank (0.16)/noise (0.14), while *blur* (0.96) nearly matches real (0.98). Destroying high-frequency detail but keeping coarse layout preserves the decision-relevant bits; a temporally or semantically irrelevant real frame does not. This rules out a *marginal-image* OOD explanation; it does not fully rule out a cross-view inconsistency artifact (a real wrist frame paired with the current agent view is a joint the policy never saw), which a same-timestep decision-region occlusion control would isolate, left to future work.

Table 6: **Embodied InfoGap** (OpenVLA-OFT on LIBERO; success rate, n=100/cell unless noted). *Top:* masking the wrist channel at fixed policy and compute collapses success on every suite. *Middle (R10):* under masking, sweeping the action-sample budget K does not recover success under the tested budgets and past K=2 degrades it. *Bottom (R11):* matched-distribution wrist controls (n=50, spatial) show the collapse is not a blank/noise marginal-image artifact, temporally/semantically irrelevant real frames also fail.

	spatial	**goal**	**object**	**LIBERO-10**
Full (agent + wrist)	0.97	0.96	0.98	0.26
Masked (wrist blanked)	0.07	0.09	0.31	0.00
InfoGap	**+0.90**	**+0.87**	**+0.67**	**+0.26**
R10 masked action-sampling (K, medoid vote):				
K=1 / 2 / 4 / 8	0.05 / 0.08 / 0.00 / 0.00	0.10 / 0.10 / 0.00 / 0.00	—	—
R11 wrist controls (spatial, n=50): real **0.98**, blur 0.96 | blank 0.16, noise 0.14, other 0.02, stale 0.00				

Table 7: **What each line of work scales, and whether it decides.** *compute*: more inference over a fixed information set. *fuse*: recover decision-relevant evidence *already observed* but discarded by the interface. *acquire*: obtain observations outside the current information set. *risk decomp.*: a formal decomposition of excess risk into separately-removable terms. *diagnose & route*: choose among the operations per instance. ✓ present, ∼ partial/qualitative, blank absent. Prior work covers individual cells; the decomposition that *derives* the action space is what we add.

	compute	**fuse**	**acquire**	**risk decomp.**	**route**
Sampling / verifiers / search	✓				
Test-time interaction	✓		✓		
Interaction scaling	✓		✓	∼	
Scaling perception	✓	∼	✓		∼
Adaptive looking (VLA)	✓				∼
Reuse-vs-retrieve	✓		✓		✓
Active visual BED			✓		
Adaptive retrieval	✓		✓		✓
Task-dependent sufficiency		✓		✓	
ReLATE (ours)	✓	✓	✓	✓	✓

Rate–relevance detail (E5). VLM state constructions on relational XOR (n=120; cross-fitted proxy, 95% bootstrap CIs): raw concatenation rate 98.1, relevance 0.967 [0.928, 1.008]; free-form hypothesis 59.8, 0.970 [0.908, 1.023]; summary 20.8, 1.009 [0.970, 1.047]; oracle summary 3.0, 1.019 [0.979, 1.060]; taxonomy label 1.0, 1.024 [0.982, 1.067]. Learned bottleneck on synthetic (exact KL rate): $\lambda_{\mathrm{IB}} = 10^{-4}, 3{\cdot}10^{-4}, 10^{-3}, 3{\cdot}10^{-3}, 10^{-2}$ give rates $160.9, 81.6, 38.9, 20.2, 8.3$ bits and relevances $0.226, 0.225, 0.209, 0.198, 0.147$ bits.

Embodied acquisition-content control (E6-embodied). LIBERO-spatial, n=50 per condition: blank wrist 0.16, noise wrist 0.14, real wrist 0.98. The noise condition acquires an observation of the correct shape and modality but with no decision-relevant content; it is statistically indistinguishable from acquiring nothing.

Failed cross-architecture attempt ($\pi_{0.5}$). We attempted a second VLA architecture (flow-matching $\pi_{0.5}$ via its public stack) for embodied transfer. The model loads and consumes a maskable wrist channel, and we identified and fixed a real checkpoint-loading bug in its library (tied `embed_tokens` omitted from the checkpoint, leaving the language backbone randomly initialized). Despite the fix, the policy returns 0% success under its canonical evaluation pipeline in our environment, so no $\pi_{0.5}$ InfoGap is reported and embodied transfer covers OpenVLA-OFT's four suites only.

Table 8: **Cross-model transfer (E9).** InfoGap (accuracy of ReLATE B=1 minus the Best-of-16 compute ceiling) per category and model. n=500/category for Qwen3-VL-8B, n=150/category for the transfer models.

Model (family)	occlusion	before/after	rel. XOR
Qwen3-VL-8B (Qwen)	+0.50	+0.51	+0.52
Qwen3-VL-4B (Qwen)	+0.45	+0.44	+0.49
Gemma-4-12B (Gemma)	+0.49	+0.50	+0.51

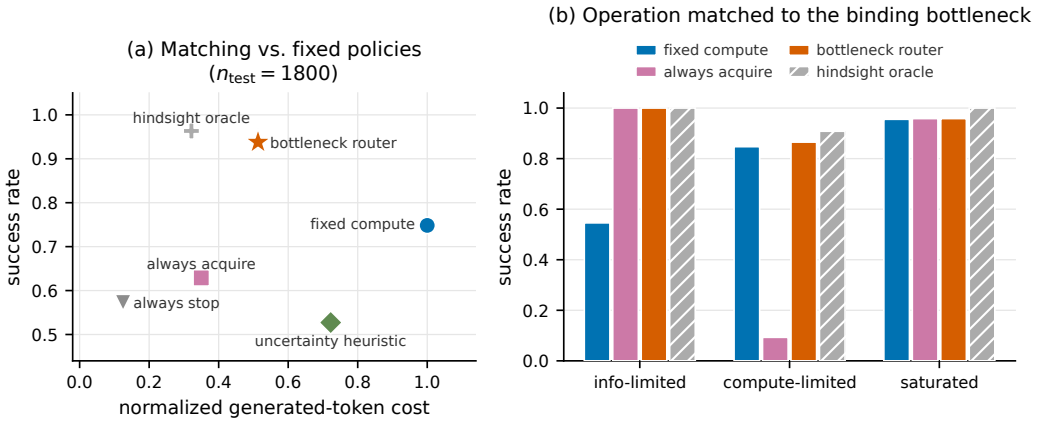


Figure 4: **Bottleneck matching** (held-out test pool, n_{test}=1800; numbers in Table 4). (a) Success vs. normalized generated-token cost: matching ($\star$) sits above every fixed policy and near the hindsight oracle at half the generated-token cost of fixed compute. (b) Per regime the selected operation is the one the decomposition predicts, acquire when information-limited, compute when compute-limited, stop when saturated. Regenerated from `r8_router_v2_predictions.jsonl`.

F Limitations

Synthetic numbers are exact by construction, a validity check rather than independent evidence, and the relational encoder is validated on synthetic only; vision scenes are procedural, with single-view accuracy verified at chance under the *same* prompt but natural-image versions left to future work. On the controller side, the same-renderer counterfactual benchmark for instance-level diagnosis did not pass its calibration gate, only the one-step router is evaluated rather than the iterative loop, DRIG is estimated from oracle probe rollouts that are costly outside simulation, and the hindsight oracle is optimistic with validation labels inheriting its definition. Embodied results are ancillary: simulated, one architecture, published LIBERO-10 baseline not reproduced. Cost counts generated output tokens only, so the total-cost sign is unidentified (Table 2).

G Pipeline validation

All reported numbers were generated from logged outcome files in our internal pipeline, and the released supplementary material contains a verified reproducible subset of those scripts and logs. No reported result was manually entered in the internal analysis pipeline. That pipeline enforces the following invariants; those marked below are checkable directly in the released subset.

- **Labels are derived, not stored.** Routing labels are recomputed from the executed arm outcomes under the utility u = success $- \lambda c$, and `train_router_v2.py` asserts that every stored label equals the recomputed utility argmax before training. A stale or hand-edited label fails the run rather than propagating silently.

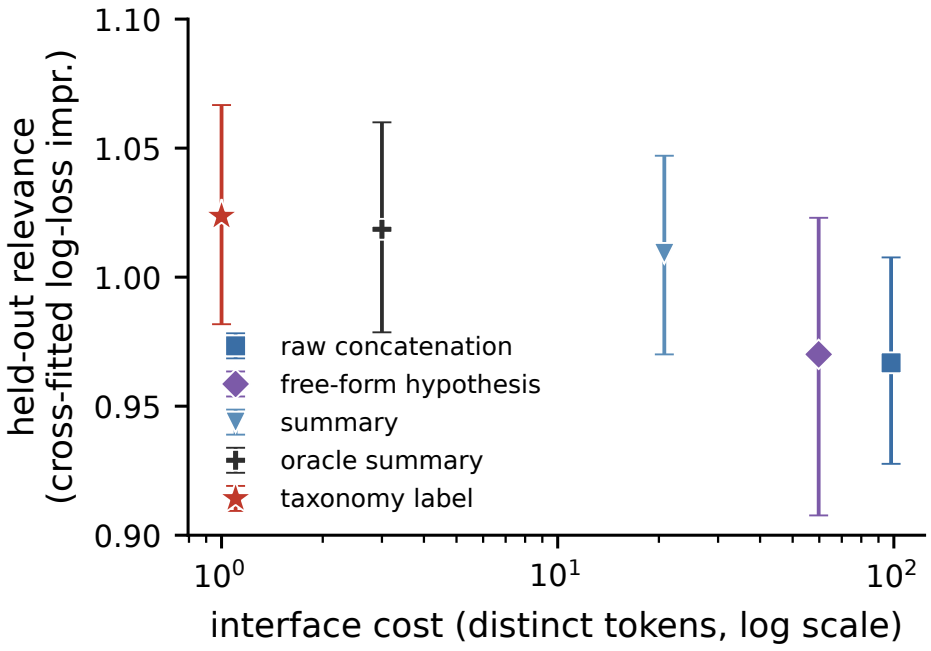


Figure 5: **Cost–relevance of five state constructions** (InfoGap-Vision, relational XOR, n=120). Each marker is one construction's *measured* point with a 95% bootstrap CI, not a fitted curve; x is its distinct-token interface cost on a log scale (*not* a Shannon rate) and y the cross-fitted held-out log-loss improvement (a predictive proxy, *not* an MI bound). Compact constructions (taxonomy, oracle summary; 1–3 tokens) retain essentially all of the relevance of raw concatenation at cost 98. Values tabulated in Appendix E (E5).

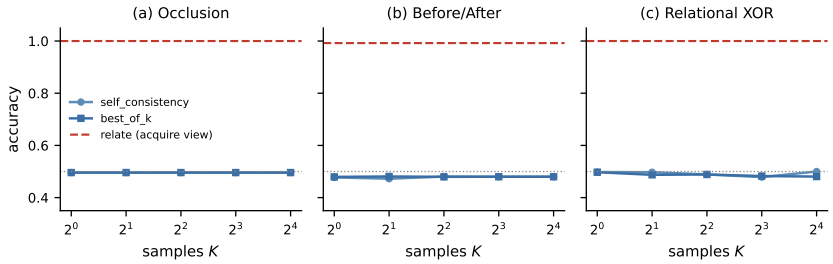


Figure 6: **Full K-sweep per category** (Qwen3-VL-8B, n=500/category). Self-consistency and Best-of-K are flat from K=1 to 16 on all three categories; the dashed line is ReLATE with the acquired view. This is the per-category detail behind Figure 1.

- **Tables are generated.** The rows of Table 4 are the verbatim output of `emit_table3.py` applied to `r8_router_v2_report.json`; `check_table3.py` compares the manuscript block against that output as an ordered list, so drift, reordering or duplication fail.
- **Figures are derived from logged files, not constants.** No figure in this paper is generated from a closed-form expression; each figure script loads its values from a named artifact, so re-running a figure after re-running its experiment updates it. The released supplement contains a *verified subset* of these scripts rather than the complete pipeline: three vision figures were produced by scripts that were not retained, and their late-voting records are not among the released logs.

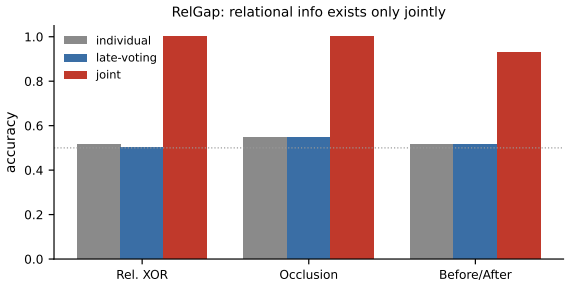


Figure 7: **RelGap on a real VLM (E7).** Individual view, late voting (both views seen, answered separately, then voted), and joint context, per category (n=450/cell). Late voting stays at the individual baseline (exactly 0.500 on relational XOR), while the joint context solves the task: the decision-relevant information exists only jointly.

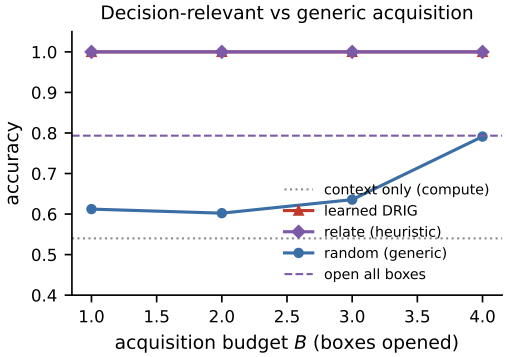


Figure 8: **Decision-relevant vs. generic acquisition in the live VLM loop** (targeted probe, Qwen3-VL-8B). The trained DRIG selector and the ReLATE heuristic reach 1.0 at one probe; random acquisition needs the full budget and opening all four boxes (0.79) is *worse* than the one relevant probe.

- **Per-row provenance.** Logged rows carry task, method, K, B, seed, costs and estimator, so any reported aggregate can be traced to the rows that produced it.
- **Determinism.** The router pipeline fixes its seeds and reproduces the reported report file bit for bit on the pinned dependency versions.

Verifying the manuscript against these artifacts did surface defects in earlier intermediate files, in answer parsing and in label propagation; all reported results were regenerated after those fixes, and the assertions above are what now prevent the same classes of error. One consequence is visible in the paper: we do not plot regret in nats, because the logged per-instance regret fields were not computed under the measurability convention of §3.1, and we report the decomposition qualitatively instead of recomputing those fields.

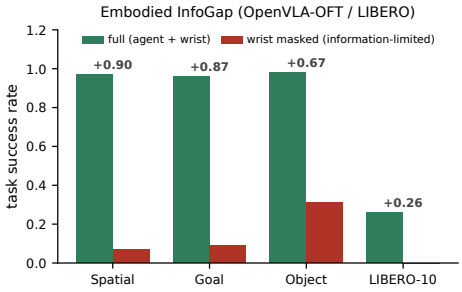


Figure 9: **Embodied InfoGap** per LIBERO suite (OpenVLA-OFT, n=100/cell): full vs. wrist-masked success. The wrist channel is decision-relevant in every suite; LIBERO-10 collapses to 0.00 when it is removed.