

Verifier Information & Missing-Outcome Bounds

Cisco Caceres · 7 October 2026 Pacific · Authored methods supplement

These elementary derivations are not claimed as novel. Model-assisted independent checks reviewed the proof; they are not human peer review. The finite examples use authored probability laws, not model outputs. They provide no empirical support for the proposed verifier controller and change no frozen study. The numerical helper is illustrative and is not certified for extreme rational inputs.

A finite-action conditional-information ceiling

Prospective authored methods draft. This is an elementary decision-theoretic bound, not a novelty claim, an empirical finding, or an amendment to frozen H1. No model calls, task execution or external spending were used.

Assume a finite latent outcome state W , finite current observation X and finite additional verifier observation Z with a specified joint probability law. A nonempty finite action set A is identical before and after observing Z . A policy may condition only on the observations it has received. For each action let $u(a,x,w)$ lie in $[0,1]$; the utility function and action set do not change when Z is revealed. Expectations refer to the specified population law, not an inferred target guarantee. Zero-probability observation cells are omitted.

For fixed candidate selection, W can be the candidate correctness vector, A the fixed candidate identities plus optional abstention, and $u(a,x,w)$ the selected candidate's binary correctness. Abstention has success utility zero for this endpoint. A differently valued abstention changes the utility endpoint. X may include already permitted observations. Z is an additional signal, not permission to query hidden labels or expand the candidate bank.

Let $p_x(w) = P(W=w|X=x)$ and $q_{xz}(w) = P(W=w|X=x,Z=z)$. Define

$$\begin{aligned} V_0 &= E_X \max_a \sum_w p_X(w) u(a,X,w) \\ V_1 &= E_{XZ} \max_a \sum_w q_{XZ}(w) u(a,X,w) \\ D &= V_1 - V_0 \\ T &= E_{XZ} TV(q_{XZ}, p_X), \quad TV(q,p) = (1/2) \sum_w |q(w) - p(w)| \\ I &= I(W;Z|X) = E_{XZ} KL(q_{XZ} || p_X) \end{aligned}$$

KL and I use natural logarithms; I is measured in nats. Then

$$0 \leq D \leq \min\{1 - V_0, T, \sqrt{I/2}\}.$$

In particular, conditional independence $W \perp Z | X$ implies $D=0$. The converse need not hold: a signal can reveal outcome information irrelevant to the available decisions. Mutual information and posterior variation provide upper bounds, not a predicted gain or an identification formula for utility.

The same bounds hold conditionally for each positive-probability $X=x$, using $V_0(x)$, $E[TV|x]$, and $I(W;Z|X=x)$. For utilities in any common finite interval $[L,H]$, multiply the TV and information bounds by $H-L$, and replace the first ceiling by $H-V_0$. When $H=L$, the gain is zero.

Proof. A policy with Z can ignore it and use an optimizer for X , so $V_1 \geq V_0$; equivalently, conditional expectation followed by the convex maximum gives the inequality. Both values are at most one, so $D \leq 1 - V_0$.

For each (x,z) , choose a maximizer a_z for q_{xz} . The coarse optimal value is at least its value under p_x , hence

$$\begin{aligned} \max_a E_q[u(a,x,W)] - \max_a E_p[u(a,x,W)] \\ &\leq E_q[u(a_z,x,W)] - E_p[u(a_z,x,W)] \\ &\leq TV(q,p). \end{aligned}$$

The last inequality follows by splitting $q-p$ into positive and negative parts and using $0 \leq u \leq 1$: the total positive mass is TV . Taking expectations yields $D \leq T$. Pinsker's inequality, with natural logarithms, gives $TV(q,p) \leq \sqrt{KL(q||p)/2}$. Concavity of the square root then gives $T \leq \sqrt{I/2}$. Here p_x is the mixture of q_{xz} over z , so a positive posterior mass never falls outside the support of p_x . Affine utility rescaling proves the $[L,H]$ statement. For completeness, the finite-law Pinsker step follows by grouping states into $B = \{w: q(w) > p(w)\}$ and its complement. The log-sum inequality reduces $KL(q||p)$ to the binary relative entropy $d(q(B)||p(B))$. For fixed t in $(0,1)$, $d(s||t)$ has value and first derivative zero at $s=t$ and second derivative $1/[s(1-s)] > 4$, so $d(s||t) > 2(s-t)^2$. Since $q(B)-p(B)=TV(q,p)$, the asserted inequality follows, including endpoints by continuity or infinite divergence. The log-sum inequality itself follows from convexity of $v \log v$ applied within each group with p -normalized weights. This completes the proof.

For the correctness endpoint, $R_0=1-V_0$ and $R_1=1-V_1$ are Bayes error risks, so $D=R_0-R_1$. The proposition bounds the improvement possible from extra information under the assumed law and fixed decisions. It does not bound a learned controller's achieved gain from below. For arbitrary learned policies with values $L_0 \leq V_0$ and $L_1 \leq V_1$, the correct comparison is

$$L_1 - L_0 = D + (V_0 - L_0) - (V_1 - L_1).$$

Thus a learned gain need not be $\leq D$: it can also reflect a different approximation to the coarse Bayes policy. The ceiling concerns matched Bayes-optimal decisions, not every empirical policy comparison.

Charging the verifier changes the comparison. If acquisition is mandatory, costs c in the same utility units, and leaves these actions and correctness utilities unchanged, its net incremental value is $D-c$. It may be negative even though $D \geq 0$. A dollar charge cannot be subtracted from a success probability without a prospectively specified utility conversion. With expected costs use the corresponding expected charge; cost-constrained reachability, optional acquisition, multiple sequential queries or feedback-conditioned generation require an explicit action/observation model and their own optimization. The free-information bound alone does not justify those policies.

A correctness-vector law is necessary for this application: pairwise verifier covariance, marginal false acceptance/rejection, or a variance-equivalent count alone need not determine the posterior utility. Candidate-score calibration alone likewise does not establish the selected-policy risk. Estimating the full law or an information bound from source data requires separate estimator, support, calibration and uncertainty assumptions. Plug-in estimates are not automatically valid target bounds. These prerequisites are not supplied by this draft.

Authored examples in the companion analytic module show a tight TV ceiling (a perfect binary signal raises guessing accuracy from $1/2$ to 1), zero incremental value for an independent signal, positive conditional information with zero decision value, and a cost larger than a real free-information benefit. Tests enumerate small finite laws using exact rational probabilities and check the analytic bounds. No real verifier or candidate bank is used.

The prepared ridge controller ranks source-predicted incremental continuation value per worst-case quote. It is a particular learned heuristic, not the Bayes optimizer in this proposition and not an established optimal value-of-information policy. A practical contribution still needs source-only fitting/calibration, action-support checks, matched baselines, frozen resource accounting, held-out task-cluster inference and authentic

observations. This draft supplies only the modest conditional-information proposition; it does not establish frontier-model transfer, full portfolio allocation, metric novelty or H1 support.

Fixed-bank bounds with missing execution outcomes

Authored mathematical note, 7 October 2026 Pacific. This is an elementary finite-data bound, not a model experiment, novel theorem, protocol amendment or qualification receipt.

For each task t , a fixed bank has candidate truth labels Y_{tj} in $\{0,1\}$. Known successful or failed execution fixes a label only under the declared suite. Missing, refused, ambiguous or timed-out evaluation leaves a label unknown. This note does not identify suite truth with universal program correctness.

Two already fixed policies select candidate indices b_t and s_t , or abstain. Abstention has endpoint value zero. Let A_t contain every binary assignment of unknown labels consistent with the observed labels. For each assignment a , compute $B_t(a)$, $S_t(a)$, $O_t(a)=\max_j Y_{tj}(a)$, and $D_t(a)=S_t(a)-B_t(a)$. Define $l_t=\min_a D_t(a)$ and $u_t=\max_a D_t(a)$. Then $\text{mean}_t l_t \leq \text{mean}_t D_t \leq \text{mean}_t u_t$. These bounds are sharp when unknown labels can vary independently across tasks and no additional cross-task constraints are imposed: each task's minimizing or maximizing assignment can be combined with all others.

Shared candidate identity matters. If both policies select the same candidate, D_t is exactly zero even when its truth is unknown. If selected indices differ and both labels are unknown, the interval is $[-1,1]$; treating the two policies' correctness intervals separately would miss the exact cancellation for a shared index. If one abstains, the bounds follow the other selected label with its sign. Known labels collapse the relevant interval.

Bank oracle O_t has lower bound one if any known candidate passes, otherwise zero. Its upper bound is zero only when every candidate has a known failing label; otherwise it is one. Empty banks have oracle zero. Oracle headroom $H_t=O_t-B_t$ must be bounded jointly over A_t , rather than by subtracting independent oracle and baseline intervals. If baseline is known correct, H_t is zero. If no candidate outside the baseline can be correct, H_t is also zero even when the baseline outcome is unknown. These are logical bounds, not a prediction that a controller can realize oracle headroom at a cost cap.

For a cost-qualified action oracle, enumerate only prospectively admissible action sequences and preserve their revealed-information and charged-cost constraints. Unknown action costs or missing reachability evidence do not justify adding a sequence to that set. An outcome-informed oracle remains offline analysis, never controller input.

These are identification intervals conditional on a retained finite bank and observed execution dispositions, not confidence intervals for a task population. Sampling uncertainty, repeated seeds and task clustering require a separate analysis. Outcome-dependent exclusions or replacing missing labels by failure change the estimand and must not be hidden. Policies must already be fixed independently of the missing hidden outcomes; feedback-conditioned generation and label-informed policy selection need a different design.

No candidate data, model calls, native execution or external spending were used.

Reproduce the illustrative examples

Download `finite_value.py` and `test_finite_value.py` into one directory. Run `python3 -B -m unittest test_finite_value` and repeat with `python3 -O -B -m unittest test_finite_value`. The seven tests include 35 small binary laws; they do not use task data or run models. Rational probabilities are exact, while logarithms use floating-point arithmetic. These illustrative checks do not certify extreme numerical inputs or replace the proof.