

Search, Experiments, and the Value of Feedback

Which observation is worth obtaining next?

Complexity, Information, and Cross-Domain Transfer

Why question the most informative experiment?

A deployment succeeds with probability 0.4; success pays +10, failure −10, and abstaining pays zero. Only one test costing 3 can be afforded.

An independent fair binary nuisance is also unknown but has no effect on success or payoff. Learning it reduces uncertainty without changing the action. This suggests valuing tests through decisions.

Two plausible tests answer different questions

Choice	Information gained	Net utility
Reveal nuisance	1 bit; irrelevant to action	-3
Reveal success	0.971 bits about success	+1
Do not test	0 bits	0

Here information is measured about the joint state (success, nuisance). The highest bit count need not buy the best decision.

A worked decision-value calculation

$$\text{VOI} = (0.4 \times 10 + 0.6 \times 0) - 0 - 3 = 1$$

A candidate succeeds with probability 0.4. Deployment pays +10 on success and −10 on failure; not deploying pays zero. Current best utility is zero.

A perfect test allows deployment only after success, giving expected utility 4. A test costing 3 has net value 1.

Value of information is decision-specific

$$\text{VOI}(e) = E_y[\max_a E(u|D, y, e)] - \max_a E(u|D) - c(e)$$

Let $u(a, h)$ be utility of final action a when h is true. An experiment's value is the expected best posterior utility minus the best current utility, less experiment cost.

It can be zero even when information gain is positive.

Expected information gain

$$\text{EIG}(e) = H(h|D) - E_y[H(h|D,y,e)]$$

For target h , candidate experiment e and current data D , average over possible outcomes y . The expected posterior entropy reduction is conditional mutual information.

The expectation uses the current model's predictive distribution; a wrong model can favor the wrong experiment.

Balanced questions reveal more bits

$$h_2(1/2)=1; \quad h_2(1/4)\approx 0.811 \text{ bits}$$

Four equiprobable hypotheses have entropy two bits. A deterministic test splitting them 2:2 yields one bit; a 3:1 split yields about 0.811 bits.

The balanced split is better for immediate information, provided the tests have equal cost and reliability.

Noise changes the value of a balanced test

$$I(\text{partition}; Y) = 1 - h_2(\eta)$$

A balanced binary test with independent flip probability η is a binary symmetric channel. It supplies $1-h_2(\eta)$ bits about the tested partition.

At $\eta=0.1$ the value is about 0.531 bits. An unbalanced but much cleaner test can be preferable.

A history-dependent discovery policy

$$a_t = \pi(H_t, b_t); \quad H_{t+1} = H_t \cup \{(a_t, y_t)\}$$

At step t , history H_t contains actions and outcomes. A policy selects an experiment or computation a_t , observes y_t and updates the history.

The action can test a hypothesis, compare candidates or improve the measurement process itself.

Counterexamples compress failures

Counterexample-guided inductive synthesis proposes a program, asks a verifier for a violation and refines the candidate using that counterexample.

A counterexample is useful because many programs fail on the same input; the feedback excludes a region of program space.

The polynomial example revisited

At $x=3$, $f(x)=x^2$ returns 9 while $g(x)=x^2+x(x-1)(x-2)$ returns 15. Observing the intended output 9 removes g .

A test at $x=1$ would add no information because both hypotheses already agree there.

Active learning chooses what to label

Active learning lets the learner select labels or experiments, often by uncertainty or disagreement. Its gains depend on the query access and the data distribution.

A membership query can ask about any permitted input; pool-based selection chooses only from available examples.

Three acquisition rules choose different points

Point	Class probabilities	Max p / margin	Entropy (bits)
A	.450, .449, .101	.450 / .001	1.371
B	.400, .380, .220	.400 / .020	1.540
C	.410, .295, .295	.410 / .115	1.567

The table gives predictive class probabilities for three unlabeled points. Least confidence chooses B; smallest top-two margin chooses A; maximum entropy chooses C.

Search for quality differs from learning the whole function

Bayesian optimization learns a probability model for costly f . A Gaussian process (GP) specifies jointly Gaussian values at any finite set of inputs, using a mean and covariance kernel k .

Posterior mean predicts value; posterior variance tracks uncertainty under the model. The next query can favor a promising mean or informative uncertainty rather than map f everywhere.

GP-UCB: a finite-domain probabilistic setting

$$\mathbf{x}_t \in \operatorname{argmax}_{\mathbf{x} \in D} [\mu_{t-1}(\mathbf{x}) + \sqrt{\beta_t} s_{t-1}(\mathbf{x})]$$

Let D be finite, $f \sim \text{GP}(0, k)$, $k(\mathbf{x}, \mathbf{x}) \leq 1$, and $y_t = f(\mathbf{x}_t) + \varepsilon_t$ with iid Gaussian noise of known variance $\sigma^2 > 0$.

The upper confidence bound (UCB) is posterior mean plus $\sqrt{\beta_t}$ times posterior standard deviation. Maximize it each round, assuming the kernel, noise model and maximization are correct.

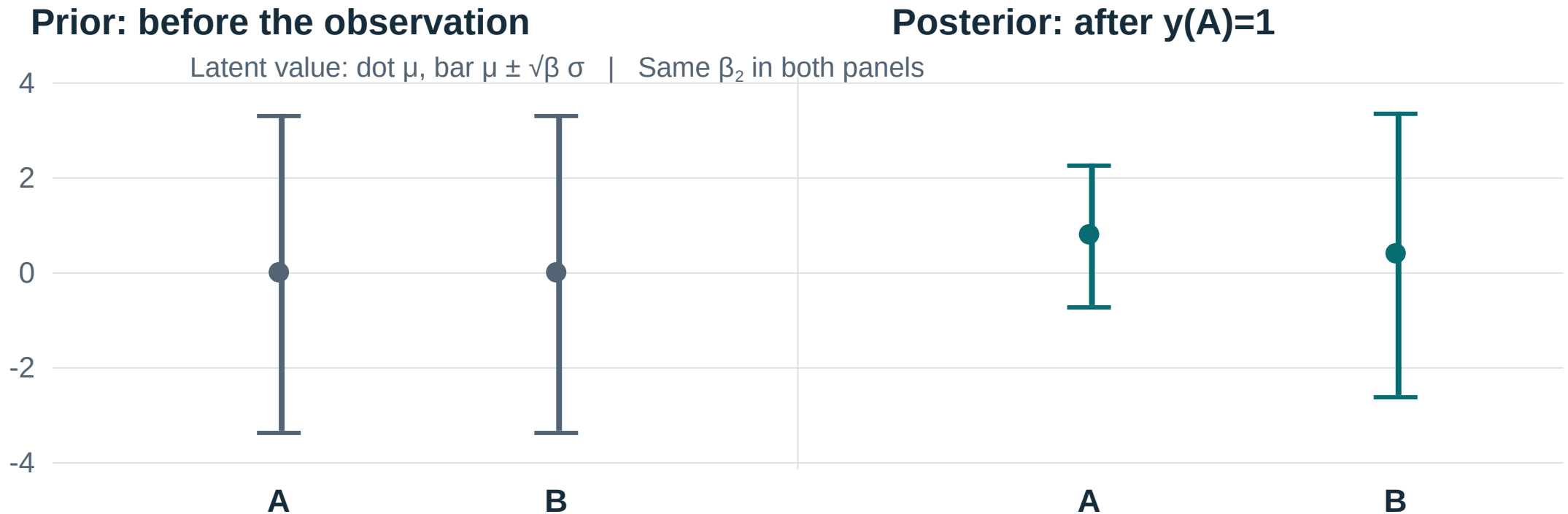
A posterior update changes the next experiment

$$\mu_1=(0.8,0.4); \quad K_1=[[0.2,0.1],[0.1,0.8]]$$

Use two candidates A,B with prior variances 1, covariance 0.5, and noise variance 0.25. Observe $y(A)=1$.

Posterior means are (0.8,0.4), variances (0.2,0.8), and covariance 0.1. With $\delta=0.05$, the theorem's round-2 $\beta \approx 11.146$ gives UCBs (2.293,3.386): query B next.

Posterior uncertainty explains the next query



At the same round-2 $\beta \approx 11.146$, both prior UCBs are 3.339. After $y(A)=1$, A has a higher mean but B retains more uncertainty: B's UCB is 3.386 versus 2.293.

GP-UCB: the exact finite-domain guarantee

$$R_T \leq \sqrt{(C_1 T \beta_T \gamma_T)}$$

Define R_T as cumulative loss relative to the best true f -value. Set $\beta_t = 2 \ln(|D| \pi^2 t^2 / (6\delta))$ and $C_1 = 8 / \ln(1 + \sigma^{-2})$.

Let γ_T maximize $\frac{1}{2} \ln \det(I + \sigma^{-2} K)$ over T -query designs. Then, with probability $\geq 1 - \delta$, the bound holds for every $T \geq 1$.

Why the regret bound contains information

$$\sum_t s_{t-1}^2(x_t) \leq 2\gamma_T / \ln(1+\sigma^{-2})$$

Gaussian tails plus a union bound give simultaneous confidence bands. On that event, optimism implies $r_t \leq 2\sqrt{\beta_t} s_{t-1}(x_t)$.

Cauchy–Schwarz turns cumulative regret into summed posterior variance. The log-determinant identity bounds that sum using γ_T .

A good search history can end with the wrong candidate

Candidate	True mean	Queries	Successes	Observed mean
A	.60	9	6	.667
B	.50	1	1	1.000

True means are $\mu_A = .60$ and $\mu_B = .50$. Nine queries to A and one to B incur cumulative regret .10 and include the best candidate. The displayed Bernoulli observations make empirical-mean selection choose B. Its final simple regret is .10.

Diminishing returns can justify greedy adaptation

Adaptive submodularity means that an action's expected marginal gain does not increase as more outcomes are observed, in the prescribed partial-realization order.

With adaptive monotonicity and a cardinality budget, adaptive greedy obtains a $1-1/e$ approximation in the standard maximization setting.

Complementary tests defeat a greedy intuition

$$I(H;A)=I(H;B)=0; \quad I(H;A,B)=1 \text{ bit}$$

Two binary measurements may individually reveal nothing about a target but jointly reveal it. Let target $H=A \text{ XOR } B$, with independent fair bits A and B .

Observing A alone or B alone gives zero information about H ; observing both gives one bit.

Replication and exploration solve different uncertainties

Replication estimates noise and stabilizes a promising candidate's value. Exploration tests a new region or mechanism.

When measurement error dominates, another replicate may be more useful than another proposal. When model uncertainty dominates, replication can be wasteful.

A screen needs independent confirmation

20 null candidates: 64.2% any screen pass; 4.9% any double pass

If 20 independent null candidates each pass a 5% false-positive screen, the chance that at least one passes is $1 - 0.95^{20} \approx 0.642$.

If every screen-positive candidate receives an independent 5% confirmation, the chance of any double false positive is $1 - (1 - 0.05^2)^{20} \approx 0.049$.

A valid rule for repeated confirmation looks

$$\sum_{t=1}^{\infty} 0.05/[t(t+1)] = 0.05$$

At look t , use a valid null p-value and threshold $\alpha_t = 0.05/[t(t+1)]$. Thresholds start at 0.025, 0.00833 and 0.00417.

Stop and reject at the first $p_t \leq \alpha_t$. Since the thresholds sum to 0.05, any false rejection has probability at most 0.05.

The p-values must remain valid for the actual sampling and candidate-selection rule.

Agreement at one input suggests a new experiment

At $x=0$, repeated measurements fluctuate around zero. Two explanations fit: H_0 has mean response 0 everywhere; H_1 has mean response x .

To construct a test, make each verbal explanation generate an outcome distribution: $y=\mu(x)+\varepsilon$, with $\varepsilon \sim N(0,1)$. At $x=0$ their predictions coincide.

The rival rules construct the next measurement

Candidate input	Predicted means H_0 / H_1	What it can reveal
Repeat $x=0$	0 / 0	Noise, but no model separation
Test $x=1$	0 / 1	Some model separation
Test $x=2$	0 / 2	Largest separation on $[0,2]$

Suppose $0 \leq x \leq 2$, with equal cost and noise at every input. The candidate inputs follow from where the two rules disagree. If extremes are noisier or infeasible, the choice can change.

A noisy result changes belief, without settling truth

$$\text{Posterior odds} = \text{prior odds} \times \text{likelihood ratio}$$

At $x=2$, observing $y=2$ gives likelihood ratio $e^2 \approx 7.39$ for H_1 against H_0 . Starting from equal odds, $P(H_1|y) = 7.39/8.39 \approx 0.881$.

If $y=1$, the two likelihoods match and the posterior remains 0.5. A surprising result can instead motivate another model. The next example shows why a narrow posterior can still mislead.

A narrow posterior can miss the true response

Fit the constant model $y=\theta+\varepsilon$ with $\varepsilon\sim N(0,1)$ and prior $\theta\sim N(0,1)$. Twenty-five observations at $x=0$, with mean zero, give posterior $\theta\sim N(0,1/26)$.

Its latent-response 95% interval is approximately ± 0.384 at every x . If the true mean is x , the mean at $x=2$ is 2: the interval misses because the model forbids a slope.

Stopping depends on the scientific objective

A stopping rule can require a validated threshold, a confidence interval, a proof certificate or low expected value from another experiment.

Stopping after a favorable noisy result without accounting for repeated looks can invalidate an ordinary fixed-sample error guarantee.

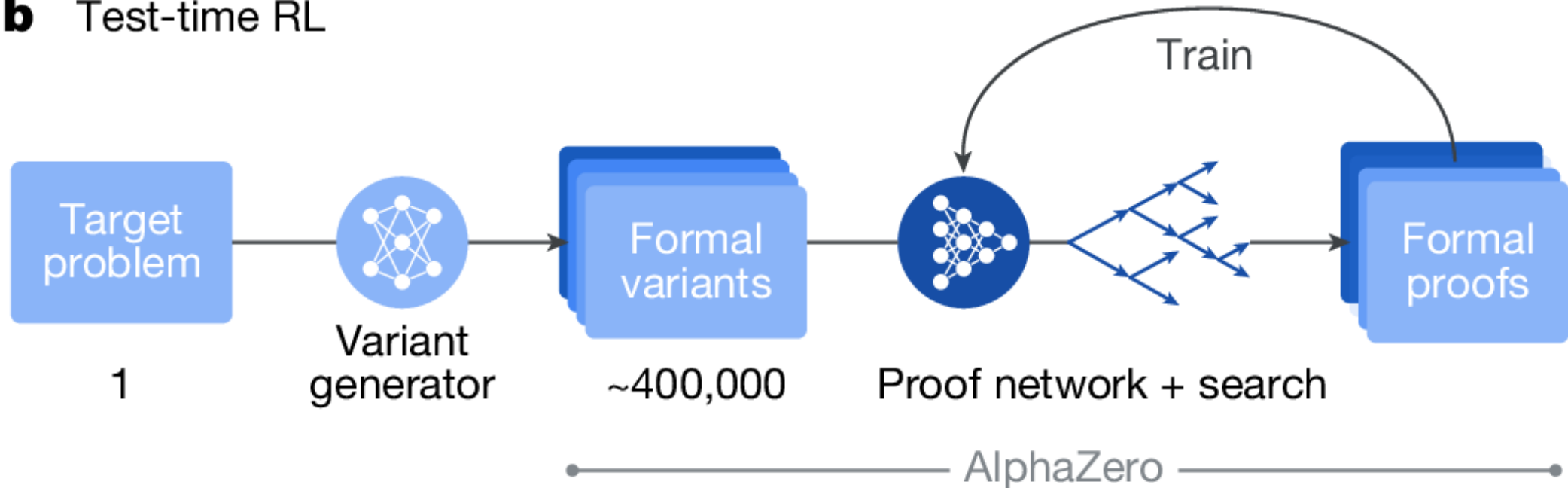
2026: learning during search changes the proposal law

Test-Time Training to Discover adapts model weights to a particular problem, unlike search using a frozen generator. It prioritizes obtaining a strong solution.

The comparison must count training as well as sampling and evaluation. Biological data denoising is an algorithmic task, not direct mechanism validation.

AlphaProof: adaptation changes the search distribution

b Test-time RL



The test-time reinforcement learning (RL) loop trains on variants of a target problem. The diagram shows a learning mechanism; success still depends on the formal task and Lean proof checker.

Feedback quality has several dimensions

Domain	Useful feedback	Main vulnerability
Construction	Violating constraint / certificate	Feasibility mistaken for optimality
Program	Counterexample input	Incomplete specification
Molecule	Assay and selectivity measurements	Proxy and context mismatch
Mechanism	Intervention response	Unmodeled alternatives

Feedback can be accurate but sparse, informative but expensive, or detailed but aligned with the wrong objective.

Conclusion: feedback has value through changed decisions

The best next action depends on the target, uncertainty, cost and opportunities that follow its outcome.

Next: which feedback dependencies can be parallelized, and which require sequential experimental rounds.

References

Dennis V. Lindley (1956). On a Measure of the Information Provided by an Experiment. Annals of Mathematical Statistics 27, 986–1005; journal article

Niranjan Srinivas, Andreas Krause, Sham Kakade and Matthias Seeger (2010). Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. ICML 2010, 1015–1022; conference paper

Daniel Golovin and Andreas Krause (2011). Adaptive Submodularity: Theory and Applications in Active Learning and Stochastic Optimization. JAIR 42, 427–486 (2011); inspected arXiv v5, 6 December 2017

Claude E. Shannon (1948). A Mathematical Theory of Communication. Bell System Technical Journal 27, 379–423, 623–656; journal article

References · continued

Sumit Gulwani, Oleksandr Polozov and Rishabh Singh (2017). Program Synthesis. Foundations and Trends in Programming Languages 4, 1–119; survey monograph

Burr Settles (2009). Active Learning Literature Survey. University of Wisconsin–Madison Technical Report 1648; survey

Wassily Hoeffding (1963). Probability Inequalities for Sums of Bounded Random Variables. Journal of the American Statistical Association 58, 13–30; journal article

S. E. Luria and M. Delbrück (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics 28(6), 491–511; published primary research

References · continued

Yarin Gal and Hao Fei (n.d.). AI Scientist for Scientific Discovery with LLM. University of Oxford MSc research-project brief; accessed 1 October 2026

Mert Yuksekgonul et al. (2026). Learning to Discover at Test Time. arXiv:2601.16175v2, 5 February 2026; preprint

Thomas Hubert et al. (2025). Olympiad-level formal mathematical reasoning with reinforcement learning. Nature; journal article, 2025