ACCURACY-OPTIMAL STOPPING CAN REVERSE TREATMENT EFFECT ESTIMATES FROM LLM LABELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Repeated language model calls can improve the accuracy of outcome labels in randomized studies. We show that stopping optimized for pooled accuracy can nevertheless reverse the reported treatment effect while accuracy rises and the population and response kernel remain fixed. When stopping may depend on treatment, this occurs even with conditionally independent calls, answers better than chance on every record, and oracle posterior decoding. In a construction with at most three calls, we characterize the attainable region of accuracy and effect, identify an open budget interval where every accuracy maximizer reports the wrong sign, and derive the exact accuracy cost of preserving the effect. For any prescribed number of sign reversals, we construct one population and a finite call cap under which price optimal policies produce those reversals as call prices decrease and accuracy strictly increases. A first revision representation characterizes all attainable effect displacements within a call horizon, without requiring independent calls or Bayesian decoding. In the four type model, a known response catalog permits unbiased correction using at most two additional calls. Semisynthetic experiments using measured response catalogs and recorded model answers show that stopping optimized separately by arm produces contrast displacement and confidence interval undercoverage on held out records. Ambiguous polarity pools can exhibit reversals, and policies with nearly identical accuracy can yield opposite effect signs. In the held out comparisons, fitted stopping rules shared by both arms produce much smaller mean absolute displacement with similar mean accuracy.

1 INTRODUCTION

Consider a study that randomizes an advertisement and measures the resulting reactions by asking a language model to read what each consumer wrote. Repeated calls can improve label accuracy, so the provider allocates computation where it yields the greatest improvement. We show that this allocation decision alone can reverse the reported treatment effect while pooled accuracy rises and both the response kernel and the population remain fixed.

The decoder maps the observed answer history to a hard label, while the stopping rule determines whether to purchase another call and, in our formal model, may depend on treatment. Because the objective averages label accuracy across both arms, gains for some record types can outweigh losses for others.

A familiar intuition suggests that this cannot happen: errors that are independent of treatment conditional on the truth attenuate a contrast when true positive rates exceed false positive rates. Yet a difference between two proxy means is not simply a count of mistakes; it is a signed sum of them. An initial comparison can be exactly correct because errors on one side cancel those on the other, and a procedure that removes errors without preserving this balance can break the cancellation. Quantification research has long separated classification accuracy from prevalence estimation, showing that prevalence estimated from an optimal classifier can be biased (Forman, 2008; Hopkins & King, 2010; González et al., 2017) and that a more accurate classifier can be a worse quantifier (Kloos, 2021). Here we hold the decoder fixed and vary only when the procedure stops asking for additional answers.

The mechanism depends on the cost of revising a label. Three agreeing answers make a label expensive, or impossible, to change, whereas two conflicting answers leave it one call from resolution.

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.

The budget therefore purchases the cheaper revisions first. These revisions occur along one of two routes through the history tree. In our construction, the two arms are mirror images: the cheaper route begins at label zero in control and label one in treatment. Revising these labels raises the control mean and lowers the treatment mean, narrowing the gap until it crosses zero. Purchasing the more expensive route as well would restore the balance, but an intermediate budget leaves the asymmetry in place. Throughout an open interval of budgets, this is the accuracy maximizing choice whenever the true effect is smaller than the mass of cheap revisions.

Section 2 makes this mechanism exact with at most three calls. Section 3 then replaces the budget with a falling price of computation. For any prescribed number of sign reversals, it constructs a population and one finite call cap that produce them while accuracy strictly increases throughout. Without requiring independent calls or Bayesian decoding, Section 6 characterizes every displacement within a call horizon through a single event: the first revision of the initial label. When the true posterior is unavailable, we use either external calibration or a known response catalog. Calibration bands on terminal histories exclude the distorted effect from every compatible structural model when the band lies within the stated tolerance and below half the mass of cheap revisions (Section 4). A known response catalog instead repairs the effect exactly, using the smallest worst case number of additional calls and attaining minimum conditional variance within that depth cap (Section 5).

The formal model fixes the response kernel, assumes that calls are independent conditional on the record, and uses a true posterior decoder, while ensuring that every record receives a correct answer more often than not. Section 7 studies measured response catalogs and recorded answers. With the decoder, tilt, and stopping rules fitted on one half of the records and evaluated on the other, arm specific stopping shifts the first call contrast by .06–.12 in mean absolute value on ambiguous pools, and nominal 95 percent coverage falls as low as 3 percent. Neither a shared rule nor a label free stopping rule tuned on each arm's training answers reverses in these comparisons. Across separate requests, marginal frequencies and pairwise agreement are consistent with independent draws from each record's mean answer distribution; completions within a request show excess agreement. Changes in prompts, conversational feedback, and tokenwise greedy generation remain outside the model.

2 A FIXED SAMPLER AND THE PRICE OF A REVISION

Let a fixed raw state Z contain a binary scientific attribute X and initially observed covariates W, including treatment D. Fresh sampling seeds are independent conditional on Z, and every call follows the same kernel:

$$\Pr(R_t = r \mid Z, R_1, \ldots, R_{t-1}) = Q_r(Z). \tag{1}$$

Two records sharing the same value of X may produce different answers because Q depends on the entire raw state. The interface observes $\mathcal{F}_t = \sigma(W, R_1, \ldots, R_t)$ and auxiliary randomness independent of $(Z, R_1, R_2, \ldots)$, but not Z or $Q(Z)$. Its oracle posterior and hard decision are

$$S_t = \mathbb{E}[X \mid \mathcal{F}_t], \qquad t \geq 1, \qquad G_t = \begin{cases} 1, & S_t > 1/2, \\ 0, & S_t < 1/2, \\ G_{t-1}, & S_t = 1/2. \end{cases}$$

At a tie, the decoder retains the preceding label; at an initial tie, set $G_0 = 0$. In the four type construction below, all first call posteriors are strict, so subsequent ties preserve the reflection between the two arms. Here S_t is a true conditional probability rather than a number reported by the model itself. Semantic correctness refers to the probability that $Q(Z)$ assigns to a category with the correct meaning, which is distinct from both the most probable output string and greedy decoding.

A policy releases G_τ at a stopping time $\tau \geq 1$ and spends $C(\tau) = \mathbb{E}(\tau - 1)$ additional calls beyond the first. Under the budget formulation used here, it maximizes the released accuracy $\Pr(G_\tau = X)$ subject to a bound on $C(\tau)$. Section 3 instead studies a price on computation. Policies may retain all observed reports, depend on treatment, and randomize. They satisfy $1 \leq \tau \leq T$ under the finite horizon formulation and $\mathbb{E}\tau < \infty$ under the price formulation. Potential attributes $X(0), X(1)$, with $X = X(D)$, precede the independent assignment $D \sim \text{Bernoulli}(1/2)$. The true and released contrasts are

$$\Delta^* = \mathbb{E}[X(1) - X(0)], \qquad \Delta_\tau = \mathbb{E}[G_\tau \mid D = 1] - \mathbb{E}[G_\tau \mid D = 0]. \tag{2}$$

For independently sampled records under a fixed stopping rule with independent policy randomization, the empirical difference in proxy means converges to Δ_τ, so the analysis below describes the large study limit rather than a particular sample. In a saturated linear mean model, Δ_τ is the population coefficient on treatment, giving its sign a direct structural interpretation. More generally, replacing G_1 with G_τ changes the population score of a binary criterion by $\mathbb{E}[a_\theta(W)(G_\tau - G_1)]$, where $a_\theta(W) = \nabla_\theta \ell(W, 1; \theta) - \nabla_\theta \ell(W, 0; \theta)$ (Appendix B). Thus, any nonzero displacement invalidates a parameter that solved the score equation before stopping.

Four record types. Fix

$$0 < q < p < 1/2, \qquad 0 < s < 1/2. \tag{3}$$

Here $W = D$, so the covariates reveal the orientation but not the record type within an arm. The control arm has the following strictly positive masses and response vectors:

X	mass	Q_A	Q_B	Q_C
0	$1-p-q$	1	0	0
0	q	0	$1-s$	s
1	q	0	s	$1-s$
1	$p-q$	0	0	1.

(4)

The rows consist of a pure negative type that always answers A, two noisy types that mix B and C in mirror image, and a pure positive type that always answers C. The treatment arm complements X and relabels each report as its barred counterpart $\bar{A}, \bar{B}, \bar{C}$, with the semantics also complemented. Reports A and B carry semantic label zero, while C carries semantic label one. Consequently, every record is semantically correct with probability at least $1 - s > 1/2$. Only the resulting mirrored arm marginals enter the statements below. The first report law and the initial anchor are

R_1	A	B	C
mass	$1-p-q$	q	p
S_1	0	s	$1-qs/p$.

(5)

$$\Delta_1 = \Delta^* = \delta := 1 - 2p > 0, \qquad A_0 := \Pr(G_1 = X) = 1 - 2qs. \tag{6}$$

The mean of the first hard label is p in the control arm and $1 - p$ in the treatment arm. Thus, the initial comparison is exactly correct at the population level, although a fraction $2qs$ of the individual labels are incorrect. This cancellation is the balance that stopping can break.

A second call changes no decision. Set

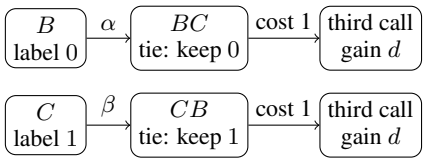
$$\alpha = 2s(1-s), \quad \beta = \frac{q\alpha}{p} < \alpha, \quad L = q\alpha = p\beta, \quad d = \frac{1}{2} - s, \quad k_B = 1 + \frac{1}{\alpha} < k_C = 1 + \frac{1}{\beta}. \tag{7}$$

The cross histories BC and CB each carry mass L and have posterior exactly $1/2$. Conditional on their first reports, they occur with probabilities α and β, respectively. At either tie, the second report leaves the first decision unchanged. The same is true for AA, BB, and CC. Thus, every second call has zero immediate accuracy gain, and a rule that purchases calls only for immediate gain never purchases a second call. The value of the second call lies in reaching a tie, since only at a tie can a third call change the decision.

A third call behaves differently. At a cross history, it returns B or C with equal probability and increases conditional accuracy by d; elsewhere, it cannot change the strict decision. There are therefore exactly two routes to a useful third call. They yield the same benefit at different costs. The B branch carries mass q and reaches a cross history with probability α, while the C branch carries mass p and reaches a cross history with probability β. Hence, $q\alpha = p\beta = L$. Purchasing the third call costs q second calls along the B route and $p > q$ along the C route. Consequently, k_B and k_C are the costs per unit of useful mass along the two routes, and each unit returns an accuracy gain of d. The inequality $k_B < k_C$ drives the mechanism. It holds because the C branch also contains the pure positive records, which always answer C and never generate the contradiction that makes a third call valuable. An accuracy optimizer therefore purchases the cheaper route first, increasing accuracy while decreasing the contrast.

Proposition 2.1 (The three call frontier and its wrong sign interval)**.** *Under* (3)*, let* $b \geq 0$ *bound the expected number of additional calls, and allow all randomized, treatment aware stopping policies with* $1 \leq \tau \leq 3$*. The attainable accuracy and effect pairs are*

$$\{(A_0 + d(u+v),\ \delta - u + v) : 0 \leq u, v \leq L, \qquad k_B u + k_C v \leq b\}. \tag{8}$$



$$\alpha > \beta \implies k_B < k_C$$

Figure 1: The mechanism, with histories oriented toward the control arm. Cross history probabilities are conditional on the first report. The C branch also contains pure positive records, which never contradict themselves, so a useful third call is more costly along that route. Treatment reflects the labels. The effective total route costs per unit of useful third call mass are k_B and k_C, respectively.

Here u and v are the pooled masses of useful third calls executed along the B and C routes, respectively. Every accuracy maximizing policy satisfies $u_* = \min\{L, b/k_B\}$ and $v_* = \min\{L, (b - k_B L)_+/k_C\}$. If $0 < \delta < L$, its effect is negative exactly when

$$k_B\delta < b < (k_B + k_C)L - k_C\delta. \tag{9}$$

Define $H(b) = [\min\{b/k_B, ((k_B + k_C)L - b)/k_C\}]_+$ and $\chi = d(k_C - k_B)/(k_B + k_C)$. Relative to the optimum at budget b without an effect constraint, the smallest accuracy loss is $\chi H(b)$ when imposing $\Delta_\tau = \delta$, and $\chi[H(b) - \delta]_+$ when imposing $\Delta_\tau \geq 0$.

Appendix E states the complete three call frontier (Theorems E.1 and E.2), proved in Appendix K, and shows that both constraint costs are sharp. Equation (8) can be interpreted as allocating a budget across two goods. Accuracy depends on the total useful mass $u + v$ and not on the route that supplies it, whereas the effect depends on the difference $v - u$. The cheaper good therefore decreases the effect. At the budget $b = k_B L$, which exactly exhausts the B route, the accuracy optimal policy attains the minimum effect among all three call policies. Further spending along the C route increases both accuracy and the effect.

For the population parameters

$$p = 9/20, \qquad q = 3/10, \qquad s = 2/5, \tag{10}$$

a budget of $b = .444$ additional calls has a unique accuracy optimum. It raises accuracy from .76 to .7744 but releases an effect of $-.044$ when the true effect is $+.10$. Every accuracy optimum therefore has the wrong sign for budgets between approximately .308 and .626. This interval is nonempty whenever $0 < \delta < L$: the true effect is smaller than the mass that the cheap revisions can move. The call horizon also matters. At the same budget, caps of five and ten calls admit accuracy optima with positive effects (Appendix N.2).

3 Reversals along a falling price of computation

A budget is not how computation is typically purchased. We therefore replace the budget constraint with a price and maximize $J_c(\tau) = \Pr(G_\tau = X) - c\mathbb{E}(\tau - 1)$ for $c > 0$ over admissible stopping times with finitely many expected calls, while holding the population, kernel, information, and decoder fixed. As c decreases, this objective traces how a provider responds when computation becomes cheaper.

The optimal rule combines a symmetric boundary m with a cutoff h for unrevised C reports. After a B excludes the pure positive type, it continues until $\#C - \#B$ reaches $-m$ or m; before the first B, it stops after h consecutive C reports. It stops on A, and treatment uses the reflected rule. As the price falls, the boundary m can increase before the cutoff h, temporarily unbalancing revisions. Appendix C gives the explicit policies and outcomes.

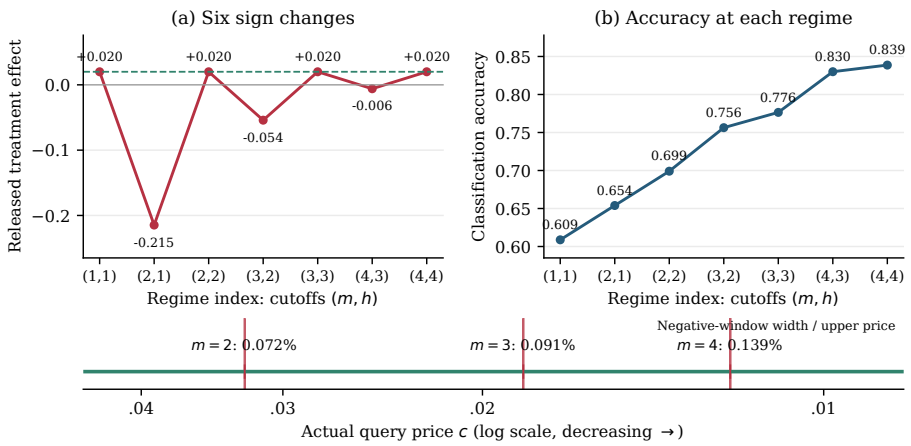


Figure 2: Exact outcomes for one fixed population with $(p, q, s) = (.49, .489, .40)$. Panels (a) and (b) show the seven selected regimes at prices strictly inside their respective intervals; the selected prices are not equally spaced. The lower strip uses the actual logarithmic price scale. The red windows retain their true widths, and each percentage reports the window width divided by its upper price. Immediate release is denoted by $(m, h) = (1, 1)$. The boundary m moves before the cutoff h, decreasing the effect, while a subsequent increase in h restores it. Both transitions improve accuracy. All values are exact kernel calculations rather than simulation estimates.

Theorem 3.1 (Arbitrarily many optimal price sign cycles)**.** *For every $s \in (0, 1/2)$ and integer $N \geq 1$, there exist $0 < q < p < 1/2$ and $2N + 1$ strictly decreasing positive prices such that each price has a unique optimal stopping policy, up to unreachable histories. Their effect signs alternate, beginning and ending positive, and their accuracies strictly increase along the sequence. The population, response kernel, information, and decoder remain fixed across prices, and every record is semantically correct with probability at least $1 - s$.*

The population depends on N, and additional cycles require a smaller true effect in this construction; negative price windows can be very narrow. For a fixed population, accuracy tends to one and every optimal effect converges to $\delta > 0$ as the price tends to zero.

Figure 2 illustrates six sign changes for $(p, q, s) = (.49, .489, .40)$, with true effect .02 and recordwise semantic correctness at least .60; the negative price windows have width approximately 2×10^{-5}. Appendix C gives the selected prices and exact calculations.

3.1 One finite cap and a tolerance for suboptimality

The optimal policies at the selected prices have positive Bellman action gaps and stopping times with geometric tails. Appendix H uses these properties to transfer the sign cycles to one finite cap and to policies with sufficiently small objective regret.

Corollary 3.2 (Arbitrary finite cycles under one common finite cap)**.** *For every $s \in (0, 1/2)$ and integer $N \geq 1$, there exist one four type population, one finite cap T, and $2N + 1$ strictly decreasing prices such that every capped price optimum has alternating positive and negative effects, beginning and ending with a positive effect. Accuracy strictly increases at every successive price. At each selected price, the sign persists for all policies within a positive objective regret tolerance. The population and the cap may depend on N.*

For the seven prices used in Figure 2, Appendix N certifies that a cap of 168 total calls and objective regret at most 10^{-9} preserve all seven signs and all six accuracy increases. This is a sufficient certificate under the fixed kernel, not a minimal cap.

4 CALIBRATION BANDS CAN EXCLUDE THE DISTORTED EFFECT

The price path fixes one law and varies the policy. Identification reverses this perspective: it fixes the observed stopped law and asks which latent completions could have generated it. If every compatible completion has a strictly positive effect, then a reader who distrusts the released proxy can still identify the sign using calibration information about terminal histories rather than individual records. We use elementary conditional mean bounds from the partial identification literature (Horowitz & Manski, 1995). Related calibrated score bounds in Chen et al. (2026, Proposition 2 and Theorem 3) condition on the truth, whereas our bounds condition on the terminal history. The result is a worked exclusion calculation rather than a general method.

Consider the accuracy optimum from Proposition 2.1 at the budget $b_* = k_B L$, which exactly exhausts the cheap route. Let P_{obs} denote the observed law of treatment, stopped reports, reference scores, and released labels. For $\xi \geq 0$, let $\mathcal{C}_\xi(P_{\mathrm{obs}})$ contain every law of binary potential truths and potential stopped records that induces P_{obs}, preserves independent treatment randomization, and places each arm and terminal history mean $\mathbb{E}_P[X(d) \mid H_\tau(d) = h]$ within ξ of the recorded reference score $s_{\mathrm{ref}}(d,h)$. Define $\Theta_{I,\Delta}(\xi) = \{\mathbb{E}_P[X(1) - X(0)] : P \in \mathcal{C}_\xi(P_{\mathrm{obs}})\}$ as in Definition L.1. The scores used here are oracle terminal posteriors. In an application, the calibration band would come from human coding within each terminal history, rather than confidence reported by the model itself. The completion class is otherwise permissive: it imposes no reference prior, no i.i.d. kernel, and no requirement that the decoder remain Bayes optimal. Define $\xi_0 = \min\{qs/p,\ s^2/[s^2 + (1-s)^2]\}$.

Proposition 4.1 (Sharp terminal band exclusion in the same experiment)**.** *For $0 \leq \xi \leq \xi_0$, the sharp treatment effect region in this completion class is*

$$\Theta_{I,\Delta}(\xi) = [\delta - 2\xi,\ \delta + 2(p+q)\xi]. \tag{11}$$

The released proxy $\delta - L$ lies below this region exactly when $L > 2\xi$. For (10) and $\xi = .02$, the region is $[.06, .13]$, whereas the proxy effect is $-.044$. Thus, every compatible structural model has a positive treatment effect, excluding the negative coefficient obtained from the proxy regression.

Within the stated range $\xi \leq \xi_0$, two thresholds govern the conclusion: every compatible effect is positive once $\xi < \delta/2$, while excluding the released number requires $\xi < L/2$. Further restrictions shrink the region, and the displayed endpoints need not remain sharp (Appendix L).

5 REPAIRING THE RELEASE WHEN THE RESPONSE CATALOG IS KNOWN

Calibration bands can identify the sign while leaving the effect's magnitude uncertain. If the interface releases the true posterior, the tower property gives $\mathbb{E}[S_\tau \mid D] = \mathbb{E}[X \mid D]$ under any admissible stopping rule, so its mean contrast already recovers the effect. The distortion instead arises from hard labels. Here the posterior and mixing weights are unavailable, but the response vectors are known.

Retain the four response vectors of (4) and their reflections, with s and the orientation known, but allow arbitrary unknown positive mixing weights in each arm separately. Neither p nor q, nor a prior on the truth, nor the initial validity of the effect is assumed. A production policy stops at $\tau \geq 1$ with $\mathbb{E}\tau < \infty$ and retains a history H_τ of reports, initial covariates and policy information; for the sharp claims, these covariates reveal only the orientation. Fresh calls use seeds independent of Z and H_τ, and barred reports are oriented to control first.

Theorem 5.1 (A posterior free correction with sharp additional depth)**.** *For a control oriented production history, define $\widetilde{V}$ as follows. If the history contains A, set $\widetilde{V} = 0$ without another call. If it contains B, buy one fresh call, put $Y = \mathbf{1}\{R_{\tau+1} = C\}$, and set $\widetilde{V} = \frac{Y-s}{1-2s}$. The remaining supported histories consist entirely of C reports. Buy two fresh calls in that case, put $J_j = \mathbf{1}\{R_{\tau+j} = B\}$, $j = 1, 2$, and set $\widetilde{V} = 1 - \frac{J_1 J_2 - s(J_1+J_2)/2}{(1-s)(1-2s)}$. Use $V = \widetilde{V}$ in control and $V = 1 - \widetilde{V}$ in treatment. Then, for every compatible raw type and almost every production history,*

$$\mathbb{E}[V \mid Z, H_\tau] = X, \qquad -\frac{2s}{1-2s} \leq V \leq 1 + \frac{2s}{1-2s}. \tag{12}$$

The additional call depths are respectively 0, 1, and 2. Each is the smallest possible worst case depth for a correction that is unbiased for every raw type still compatible with that history. Moreover,

among corrections that are unbiased for every raw type compatible with the history and obey the same depth cap, the displayed correction minimizes conditional variance under each compatible raw type. This variance assertion does not optimize over procedures that allow a larger or unbounded random number of calls.

The three depths have a simple interpretation. An A identifies the record; a B excludes the pure positive type and leaves a symmetric binary channel, which one fresh call inverts. A history consisting only of C reports remains compatible with both noisy types and the pure positive type; the resulting three unbiasedness constraints require a quadratic function of two fresh calls. After a prefix containing B, $\widetilde{V} = \frac{Y-s}{1-2s}$ is the inverse channel correction of Natarajan et al. (2013), while Lee & Valiant (2021) already optimize bounded depth unbiased outputs for noisy coin populations. The new ingredient is the surviving pure positive type, which forces the quadratic form after a prefix consisting only of C reports. The correction also preserves any integrable binary criterion, not only the contrast (Corollary M.1), although the corrected criterion need not be convex. Excursions outside $[0, 1]$ are unavoidable: while both noisy types remain compatible, no exact recordwise correction can remain within the label range, even without a deterministic call cap, provided stopping is almost surely finite under each compatible type (Proposition F.1; Jacob & Thiery, 2015).

Exactness is not free. At the accuracy optimum in (10), with $b = .444$, repair uses 1.2 fresh calls per record and the corrected contrast has variance $16.51/n$ over n records evenly split across arms, compared with $.99/n$ when the truths are observed (Appendix N).

The catalog itself is the load bearing assumption, and the effect of catalog error grows with the production cap T. A rowwise total variation error ϵ bounds the contrast bias by $2(1 + s)(T + 2)\epsilon/(1 - 2s)$, which provides little protection when T is large. The growth is not an artifact of the bound. Under a production policy that stops at the first A or B, or after $T = 1000$ calls, perturbing only the pure positive row to $(0, \epsilon, 1 - \epsilon)$ with $\epsilon = .001$ drives the repaired effect to $-.279$ against a true effect of $.10$, even though every semantic mode remains correct (Section O.1).

6 Every displacement that stopping can produce

The four type population is one instance of a more general representation that requires neither i.i.d. reports nor Bayesian decoding and characterizes every increment a release gate can generate when only the horizon constrains it. Let $\mathcal{H}_t$ be the information available to the gate, with $G_t \in \{0, 1\}$ adapted to it, and fix an integrable $\mathcal{H}_1$ measurable $a \in \mathbb{R}^K$, taking $a = a_\theta(W)$ at fixed θ for score substitution. Policies may randomize and stop at any $\tau \in \{1, \dots, T\}$. Define the first revision and its exposure by

$$\sigma = \inf\{t \geq 2 : G_t \neq G_1\}, \quad H_T = \mathbf{1}\{\sigma \leq T\}, \quad r_T = \mathbb{E}[H_T \mid \mathcal{H}_1], \quad \Gamma = a(1 - 2G_1), \tag{13}$$

where $\inf \varnothing = \infty$.

Theorem 6.1 (First revision representation)**.** *The complete set of structural increments over all such policies is the compact convex set*

$$\begin{aligned}\mathcal{Z}_T &:= \{\mathbb{E}[a(G_\tau - G_1)] : 1 \leq \tau \leq T\} \\ &= \{\mathbb{E}[\lambda \Gamma r_T] : \lambda \textit{ is measurable with respect to } \mathcal{H}_1,\ 0 \leq \lambda \leq 1\}.\end{aligned}$$

Every point has a realization that, at time one, chooses with probability λ to wait until $\sigma \wedge T$, and otherwise stops immediately. For every direction $u \in \mathbb{R}^K$, its support function and directional width are exactly

$$h_T(u) := \sup_{z \in \mathcal{Z}_T} u^\top z = \mathbb{E}[(u^\top \Gamma)_+ H_T], \qquad h_T(u) + h_T(-u) = \mathbb{E}[|u^\top a| H_T]. \tag{14}$$

A support maximizing policy waits for the first revision precisely on the initial event $\{u^\top \Gamma > 0\}$ and stops immediately elsewhere.

This representation preserves the structural increment, but need not preserve the accuracy or cost of the original policy; initial observability of a is essential. Even when labels at deterministic times converge to the truth, early releases can sustain a nonzero displacement region. Appendix D gives sharp error bounds, a hidden weight counterexample, and exact effect intervals for the four type population.

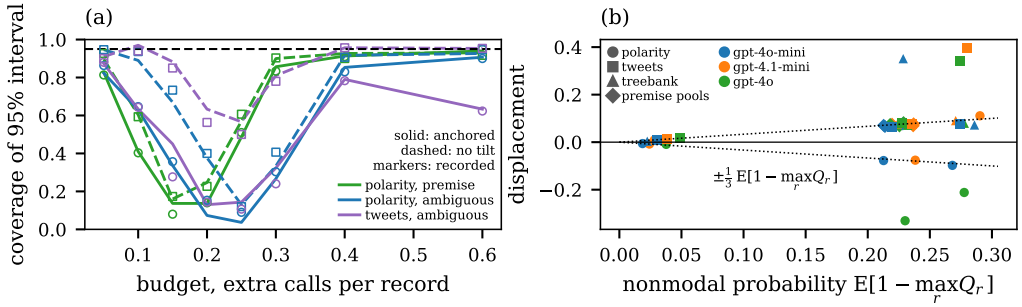


Figure 3: Stopping optimized separately by arm, $\Delta^* = .04$, horizon 5. (a) Coverage of nominal 95 percent intervals at the eight budgets in Table 18, using the optimum with the smallest effect, 300 replications of 1000 records per arm and `gpt-4o-mini`; anchored results are solid and natural results are dashed. Markers repeat the simulation with fresh draws taken from the recorded answers of Section P.5. (b) For each pool and model, signed displacement with the largest absolute magnitude among the computed accuracy optima over the 41 budgets in $[0, 2]$, plotted against nonmodal answer probability under the anchored design at temperature one. Premise pools are marked separately; dotted lines show plus and minus one third of that probability.

7 The mechanism on measured response distributions

We measured response kernels on three corpora of 3000 texts: the sentence polarity corpus of Pang & Lee (2005), the sentiment tweets of Barbieri et al. (2020) and the treebank of Socher et al. (2013). Each text was annotated once by `gpt-4o-mini`, `gpt-4.1-mini` and `gpt-4o`, each model asked for one word from {negative, neutral, positive}, or {negative, mixed, positive} on the polarity corpus. Renormalizing first token probability mass over those words gives a response vector $Q(z)$, and simulations draw independently from this catalog. For diagnostics we also recorded, for every record of four ambiguous pools, twenty separate requests and one request for twenty completions. Marginal frequencies and pairwise agreement across separate requests are consistent with independent draws from each record's mean answer distribution; completions within a request show excess agreement (Section P.5). On tweets and treebank texts, the truth is positive against the rest, including neutral texts.

Treatment is designed, not estimated. Each arm is a mixture over the pool with prevalence $\pi_0 = .48$ and $\pi_1 = .52$, so $\Delta^* = .04$ is known exactly. The anchored design tilts the within class weights of the arms in opposite directions by the amount making the first call contrast equal the true effect, the empirical counterpart of (6); the natural design applies no tilt. The decoder uses the pooled population and does not observe treatment. Stopping is optimized separately by arm, as $W = D$ permits in Section 2; a common rule is compared below.

Whole corpora offer little scope for revision. Their nonmodal answer probability, $\mathbb{E}[1 - \max_r Q_r(Z)]$, is .019 to .049, and no optimum on the tested grid moves the first call contrast by more than .018. This is not Section 6's $\Pr(\sigma \leq T)$, the probability that a hard label changes within the horizon. Records whose most likely answer has probability below .95 comprise 8 to 20 percent of each corpus and have nonmodal answer probability .21 to .24. On these ambiguous pools, every model reverses on the polarity corpus under the anchored design, with all numerical optima negative at some tested budgets. The tweets and treebank instead reach effects above .10 from a truth of .04, with no negative optimum on the grid (Figure 3(b)).

Replaying catalog optimal policies on recorded separate requests changes released effects by at most .01. Fitting and evaluating stopping on disjoint halves of each record's requests retains polarity reversals of $-.033$ and $-.032$ under `gpt-4o-mini` and `gpt-4.1-mini`, respectively. Reoptimizing on recorded `gpt-4o` answers instead gives a positive effect. Direction is fragile: on every ambiguous polarity pool, policies within .001 of optimal accuracy attain both signs (Section P.5).

Among ambiguous records that answer correctly more often than not, as Section 2 assumes, no tested optimum reverses. Yet with .15 to .20 extra calls per record, stopping lifts a valid .04 to between .106 and .122 in all nine kernels; on the 159 polarity records satisfying the premise under `gpt-4o-mini`,

Table 1: Stopping fixed on half of each pool and evaluated on the recorded answers of the other half, $b = .25$, $\Delta^* = .04$, 50 halvings; mean absolute displacement $|\Delta_\tau - \Delta_1|$, share of halvings with a negative effect, and coverage of a nominal 95 percent interval with 1000 records per arm. Accuracy optima fitted separately by arm and as one common rule, and a label free rule that asks a middle first answer again with a probability tuned on each arm's training answers to meet the budget (Section P.7).

		fitted, by arm			fitted, common			label free, by arm		
pool	design	|disp|	< 0	cov.	|disp|	< 0	cov.	|disp|	< 0	cov.
polarity, `gpt-4o-mini`	anchored	0.099	0.74	0.13	0.007	0.00	0.87	0.004	0.00	0.89
	natural	0.107	0.00	0.17	0.002	0.00	0.83	0.002	0.00	0.82
polarity, `gpt-4.1-mini`	anchored	0.104	0.54	0.08	0.008	0.00	0.90	0.004	0.00	0.91
	natural	0.109	0.00	0.19	0.002	0.00	0.78	0.002	0.00	0.77
polarity, `gpt-4o`	anchored	0.102	0.18	0.06	0.007	0.00	0.91	0.005	0.00	0.93
	natural	0.118	0.00	0.27	0.002	0.00	0.80	0.002	0.00	0.80
tweets, `gpt-4o-mini`	anchored	0.075	0.00	0.10	0.004	0.00	0.94	0.003	0.00	0.95
	natural	0.074	0.00	0.51	0.002	0.00	0.76	0.001	0.00	0.74

pooled accuracy rises from .891 to .924 at $b = .15$. In these experiments, reversals occur only under anchoring and on pools containing records that violate the premise, whereas inflation occurs even among records that satisfy it (Sections P and P.7).

With catalog and policy fixed, nominal 95 percent coverage falls to 4–14 percent under anchoring and 13–57 percent without tilt (Figure 3(a)); recorded answer replays show similar losses (Section P.5).

To assess out of sample behavior, we fit the decoder, tilt, price, and stopping policy on one half of each pool and evaluate them on the other half (Table 1). Across fifty random splits at $b = .25$, arm specific stopping has mean absolute contrast displacement of .06–.12 on every ambiguous pool, within .02 of the corresponding mean for stopping refit on the held out half at the fitted policy's actual call cost. Coverage is .03–.22 under the anchored design and .17–.69 without tilt. Across the tested budgets $b \geq .25$, training and held out effect signs agree in 96–100 percent of splits; held out signs agree with those of the matched cost refit in 96 percent. At $b = .25$, reversals occur in 74, 54, and 18 percent of anchored splits for the three polarity pools, in at most 2 percent of the other anchored splits, and in no natural design splits. A common rule fitted on the training half, and a label free rule that requests a second answer after a middle first answer with probability tuned on each arm's training answers to meet the budget, have mean absolute contrast displacement at most .01 and no observed reversals. The fitted arm specific and common rules differ in mean accuracy by at most .003 (Section P.7).

A shared rule uses the same continuation probability at each history in both arms. Its accuracy optimum never reverses on the seven comparison grids. In the natural design, the shared decoder gives $\Delta_\tau = \Delta^*\{2\Pr(G_\tau = X) - 1\}$ at pooled prevalence one half, so $|\Delta_\tau| \leq \Delta^*$. Fixed rules can still inflate the anchored effect: five calls raise .04 to .066 on the 239 ambiguous polarity records under `gpt-4o-mini` (Sections P.3 and P.6).

Scope. Arm specific optimization requires treatment information that an anonymous supplier may lack. LLM stopping rules (Aggarwal et al., 2023; Wan et al., 2025; Huang et al., 2026; Li et al., 2026) optimize objectives distinct from the contrast a study reports (Appendix A). A fixed release time selects one point of $\mathcal{Z}_T$; its displacement vanishes only when weighted revisions cancel (Theorem 6.1). Calibration bands and a known response catalog provide remedies under their stated assumptions; further limitations appear in Section P.14.

AI USE STATEMENT

Large language models (LLMs) assisted with language editing (e.g., grammar, spelling, and word choice), implementing standard computational methods, preparing visualizations for submission, and facilitating the execution of experiments. LLMs also served as objects of measurement: in the main catalog collection, each text in the three corpora was submitted once to each of three commercial model endpoints, and the returned first-token log probabilities for the three permitted answer words formed the measured response kernel. Recorded AI-generated answers were also used in the empirical diagnostics and simulations.

LLMs did not conceive the research, determine the methodology or experimental design, or develop the theoretical proofs. The authors made the research and methodological decisions, interpreted the results, reviewed the LLM-assisted work, and take full responsibility for the manuscript.

REPRODUCIBILITY STATEMENT

We provide the code for replicating our study in the anonymized Git repository: https://anonymous.4open.science/r/proxy-stopping-0CF3.

REFERENCES

Pranjal Aggarwal, Aman Madaan, Yiming Yang, et al. Let's sample step by step: Adaptive-consistency for efficient reasoning and coding with llms. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 12375–12396, 2023.

Anastasios N Angelopoulos, Stephen Bates, Clara Fannjiang, Michael I Jordan, and Tijana Zrnic. Prediction-powered inference. *Science*, 382(6671):669–674, 2023.

Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the association for computational linguistics: EMNLP 2020*, pp. 1644–1650, 2020.

Laura Battaglia, Timothy Christensen, Stephen Hansen, and Szymon Sacher. Inference for regression with variables generated by ai or machine learning, 2025.

Joachim Baumann, Paul Röttger, Aleksandra Urman, Albert Wendsjö, Flor Miriam Plaza del Arco, Johannes B. Gruber, and Dirk Hovy. Large language model hacking: Quantifying the hidden risks of using llms for text annotation, 2025. URL https://arxiv.org/abs/2509.08825.

Jacob Carlson and Melissa Dell. A unifying framework for robust and efficient inference with unstructured data. *arXiv preprint arXiv:2505.00282*, 2025.

Lingjiao Chen, Jared Davis, Boris Hanin, Peter Bailis, Ion Stoica, Matei Zaharia, and James Zou. Are more llm calls all you need? towards the scaling properties of compound ai systems. *Advances in Neural Information Processing Systems*, 37:45767–45790, 2024.

Xiaohong Chen, Ashesh Rambachan, and Elie Tamer. Partial identification from llm prompts. *arXiv preprint arXiv:2606.15031*, 2026.

Yeounoh Chung, Sanjay Krishnan, and Tim Kraska. A data quality metric (dqm): how to estimate the number of undetected errors in data sets. *arXiv preprint arXiv:1611.04878*, 2016.

Naoki Egami and Sooahn Shin. Debiased inference for ai-generated data without gold-standard labels: Identification via multiple imperfect measurements. *arXiv preprint arXiv:2608.18294*, 2026.

Naoki Egami, Musashi Hinck, Brandon Stewart, and Hanying Wei. Using imperfect surrogates for downstream inference: Design-based supervised learning for social science applications of large language models. *Advances in Neural Information Processing Systems*, 36:68589–68601, 2023.

Erik Ekström and Yuqiong Wang. Bayesian sequential composite hypothesis testing in discrete time. *ESAIM: Probability and Statistics*, 26:265–282, 2022.

James Fiedler. Bias and uncertainty in llm-as-a-judge estimation. *arXiv preprint arXiv:2605.06939*, 2026.

George Forman. Quantifying counts and costs via classification. *Data Mining and Knowledge Discovery*, 17(2):164–206, 2008.

Meyer Abraham Girshick. Contributions to the theory of sequential analysis, ii, iii. *The Annals of Mathematical Statistics*, pp. 282–298, 1946.

Pablo González, Alberto Castaño, Nitesh V Chawla, and Juan José Del Coz. A review on quantification learning. *ACM Computing Surveys (CSUR)*, 50(5):1–40, 2017.

Daniel J Hopkins and Gary King. A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 54(1):229–247, 2010.

Joel L Horowitz and Charles F Manski. Identification and robustness with contaminated and corrupted data. *Econometrica: Journal of the Econometric Society*, pp. 281–302, 1995.

Jingkai Huang, Will Ma, and Zhengyuan Zhou. Optimal bayesian stopping for efficient inference of consistent LLM answers. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=JXN3npSRRD.

Pierre E. Jacob and Alexandre H. Thiery. On nonnegative unbiased estimators. *The Annals of Statistics*, 43(2):769 – 784, 2015. doi: 10.1214/15-AOS1311. URL https://doi.org/10.1214/15-AOS1311.

MS Keane and George L O'Brien. A bernoulli factory. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 4(2):213–219, 1994.

Kevin Kloos. A new generic method to improve machine learning applications in official statistics. *Statistical Journal of the IAOS*, 37(4):1181–1196, 2021.

Gleb Koshevoy and Karl Mosler. Lift zonoids, random convex hulls and the variability of random vectors. *Bernoulli*, 4(3):377 – 399, 1998.

Chungpa Lee, Thomas Zeng, Jongwon Jeong, Jy-yong Sohn, and Kangwook Lee. How to correctly report llm-as-a-judge evaluations. *arXiv preprint arXiv:2511.21140*, 2025.

Jasper CH Lee and Paul Valiant. Uncertainty about uncertainty: Optimal adaptive algorithms for estimating mixtures of unknown coins. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 414–433. SIAM, 2021.

Guokai Li, Alys Liang, Mo Liu, Murray Lei, Stefanus Jasin, Fenghua Yang, and Preet Baxi. Asymptotically optimal sequential testing with heterogeneous llms. *arXiv preprint arXiv:2604.01086*, 2026.

Fridolin Linder, Thomas Leeper, Daniel Haimovich, Niek Tax, Lorenzo Perini, and Milan Vojnovic. Unbiased prevalence estimation with multicalibrated llms. *arXiv preprint arXiv:2604.21549*, 2026.

Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. Large language models: An applied econometric framework. *Annual Review of Economics*, 18(Volume 18, 2026):283–316, 2026. ISSN 1941-1391. doi: https://doi.org/10.1146/annurev-economics-120925-105620.

Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. *Advances in neural information processing systems*, 26, 2013.

Chirag Pabbaraju, Gregory Valiant, and Rishi Verma. A simple geometric proof of the optimality of the sequential probability ratio test for symmetric bernoulli hypotheses. In *2026 SIAM Symposium on Simplicity in Algorithms (SOSA)*, pp. 365–384. SIAM, 2026.

Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)*, pp. 115–124, 2005.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pp. 1631–1642, 2013.

Abraham Wald. Sequential tests of statistical hypotheses. In *Breakthroughs in statistics: Foundations and basic theory*, pp. 256–298. Springer, 1992.

Guangya Wan, Zixin Stephen Xu, Sasa Zorc, Manel Baucells, Mengxuan Hu, Hao Wang, and Sheng Li. Beacon: Bayesian optimal stopping for efficient llm sampling. *arXiv preprint arXiv:2510.15945*, 2025.

Lisu Wang, Yilun Chen, and Jiaqi Lu. Characterizing bias in post-bandit inference under index algorithms. *arXiv preprint arXiv:2608.01069*, 2026.

NOTATION GUIDE

Table 2: Main symbols. The query price c and expected call budget b are different optimization parameters.

Symbol	Meaning
$p, q, s;\ \delta = 1 - 2p$	Truth mass, each noisy type mass, error probability; true treatment effect in the mirrored family.
A_0, α, β, L, d	Initial accuracy; cross history probabilities given B and C; cross history mass; useful third call accuracy gain.
$u, v;\ k_B, k_C$	Executed useful third call masses and their minimum expected call costs per unit mass.
$H(b), \chi$	Budget imbalance and conversion factor for accuracy regret.
$t = 1 - s, \eta = 1 - 2s$	Correct report probability and semantic margin doubled.
$m, h;\ k = \#C - \#B$	Paired stopping boundary, unrevised all C cutoff, and current count difference.
$a_j, z_j, N_m(j), f_{m,j}$	Paired history mass factor, signed posterior, exit time, and cutoff objective term.
c_j, b_m, ϵ	Paired transition price, all C transition price, and relative pure C mass $(p - q)/q$.
γ_c, R, T	Strict Bellman action gap, allowed objective regret, and total call cap.
$s_{\mathrm{ref}}, \xi, \xi_0$	Recorded reference score, calibration band radius, and largest stated unclipped tolerance.
$\widetilde{V}, V;\ J_j, Y$	Control oriented and actual correction; fresh B and C indicators.
$a \in \mathbb{R}^K, \Gamma$	Initially observable structural weight and signed exposure $a(1 - 2G_1)$.
$\sigma, H_T, r_T;\ \mathcal{Z}_T, h_T$	First revision, its occurrence and conditional probability; increment region and support function.
$\rho_C, \beta_B, \mathcal{V}$	Pure C mass, raw type B propensity, and Bellman value in the proofs.

A Related work

Stopping rules for repeated model calls are an active design problem. Likelihood ratio boundaries are classical (Wald, 1992; Ekström & Wang, 2022), and recent work optimizes them for language models, whether to identify the consistent answer (Huang et al., 2026), to test sequentially with heterogeneous models under cost (Li et al., 2026), or to allocate samples by agreement and by reward (Aggarwal et al., 2023; Wan et al., 2025). These are the policies our results are about: each is optimal for the task it names, and none of those tasks is the population contrast a study reports. Chen et al. (2024) show accuracy moving nonmonotonically in the number of calls when queries mix correct and incorrect modes; here every mode is correct and accuracy only increases while a signed quantity reverses.

The separation between per item accuracy and aggregate validity is older than language models. Quantification treats prevalence as its own estimand and shows that an optimal classifier gives a biased one (Forman, 2008; Hopkins & King, 2010; González et al., 2017), and Kloos (2021) states the mechanism we exploit: what matters is how errors distribute, not how many there are, so a more accurate classifier can be the worse quantifier. That literature holds the classifier fixed and estimates a single population quantity. We hold the classifier fixed too, vary only the stopping time, and ask about a signed contrast between two arms.

Closest in spirit among recent results are Linder et al. (2026), who show that high accuracy does not deliver unbiased prevalence while multicalibration does, and Fiedler (2026), who reports model comparisons whose sign reverses with confident intervals, driven by reusing one calibration across the compared systems. In the reversal construction, no shared calibration is used, and the population and response kernel remain fixed as the stopping policy changes. Baumann et al. (2025) document how far configuration choices move annotation based conclusions.

Corrections need information from outside the annotator, whether gold observations (Angelopoulos et al., 2023) or an explicit reference design (Battaglia et al., 2025; Egami et al., 2023; Carlson & Dell, 2025; Ludwig et al., 2026; Lee et al., 2025; Wang et al., 2026). Our catalog is a different input: it fixes the response vectors, leaves the mixing weights unknown, and does not follow from agreement among repeated answers. Egami & Shin (2026) identify the latent label from conditionally independent proxies, which needs the independence (1) deliberately violates at the level of X, since $\mathrm{Cov}(Y_1, Y_2 \mid X, W) = \mathrm{Var}(Q_+(Z) \mid X, W)$ for fresh semantic positive indicators, measured at .213 in Section 7. Partial identification from repeated prompts under arbitrary error dependence is developed by Chen et al. (2026), whose calibrated restrictions condition on the truth where the bands of Section 4 condition on the terminal history.

B Structural substitution and score misspecification

For a criterion $\ell(W, X; \theta)$ defined using the complete data, let

$$L_*(\theta) = \mathbb{E}\ell(W, X; \theta), \qquad L_\tau(\theta) = \mathbb{E}\ell(W, G_\tau; \theta). \tag{15}$$

The scientific target minimizes L_*, whereas the proxy estimator, under the usual consistency conditions, minimizes L_τ. Write $\psi = \nabla_\theta \ell$ and $a_\theta(W) = \psi(W, 1; \theta) - \psi(W, 0; \theta)$. At a fixed θ, suppose the endpoint scores are integrable: $\mathbb{E}\|\psi(W, 0; \theta)\| + \mathbb{E}\|\psi(W, 1; \theta)\| < \infty$. Binary support then gives the exact identity

$$\begin{aligned} B_\tau(\theta) &:= \mathbb{E}[\psi(W, G_\tau; \theta) - \psi(W, G_1; \theta)] \\ &= \mathbb{E}[a_\theta(W)(G_\tau - G_1)]. \end{aligned} \tag{16}$$

If the initial labels satisfy $\mathbb{E}\psi(W, G_1; \theta^*) = \mathbb{E}\psi(W, X; \theta^*) = 0$, then $B_\tau(\theta^*)$ is exactly the score distortion induced by changing the release policy. A nonzero value rules out θ^* as an interior local minimum whenever L_τ is differentiable at θ^* and $\nabla L_\tau(\theta^*) = \mathbb{E}\psi(W, G_\tau; \theta^*)$. For a fixed release policy, a standard sufficient condition for this interchange is an open neighborhood U of θ^* on which both endpoint losses are almost surely continuously differentiable, with $\mathbb{E}|\ell(W, 0; \theta^*)| + \mathbb{E}|\ell(W, 1; \theta^*)| < \infty$ and

$$\sup_{\theta \in U,\ x \in \{0,1\}} \|\psi(W, x; \theta)\| \le M(W), \qquad \mathbb{E}M(W) < \infty.$$

Turning a score bound into a bound on parameter distance additionally requires suitable curvature of the same criterion.

For the treatment contrast, taking $a(W) = 2(2D-1)$ in (16) gives $B_\tau = \Delta_\tau - \Delta_1$. For a saturated linear mean model, Δ_τ is exactly the population proxy coefficient. Misclassification may differ across treatment arms, so repeated querying alone does not imply nondifferential measurement error.

C Explicit price formulas and the full cycle statement

Write $t = 1 - s$, $\eta = t - s$, and, for integers $j \geq 0$, define

$$a_j = t^j + s^j, \qquad z_j = \frac{t^j - s^j}{t^j + s^j}, \qquad c_j = \frac{\eta(z_{j+1} - z_j)}{2\{(j+1)z_{j+1} - jz_j\}}. \tag{17}$$

Within the paired component formed by the two noisy types, z_j is the posterior margin after j agreeing reports and qa_j is the mass of that run; the run C^j also carries the pure positive mass $p - q$. Moreover, $z_0 = 0$, $z_1 = \eta$, and the switching prices c_j decrease strictly to zero. For $m \geq 1$ and $0 \leq j \leq m$, define

$$N_m(j) = \frac{mz_m - jz_j}{\eta}, \quad f_{m,j}(c) = a_j \left\{ \frac{z_m - z_j}{2} - cN_m(j) \right\}, \quad \epsilon = \frac{p-q}{q} > 0. \tag{18}$$

Theorem C.1 (Explicit optimal policies away from paired indifference prices)**.** *For $c > c_1$, immediate release of the first label is uniquely optimal, up to histories that are never reached. If $m \geq 2$ and $c_m < c < c_{m-1}$, choose*

$$h \in \arg\min_{1 \leq j \leq m} \{f_{m,j}(c) + c\epsilon j\}. \tag{19}$$

An optimal policy stops on A. Before the first B, it stops once h consecutive C reports have appeared. After a B has appeared, it continues until the count difference $k = \#C - \#B$ reaches $-m$ or m. Treatment uses the reflected rule. If the minimizing h is unique, this policy is unique up to histories that are never reached, including among randomized policies that may depend on treatment. At a price $c = c_m$, $m \geq 1$, either adjacent paired boundary, together with its minimizing cutoff, is optimal; uniqueness is not asserted at these prices.

Its accuracy, expected additional calls, and proxy effect are

$$\begin{aligned} A_{m,h} &= A_0 + q(z_m - \eta) - \frac{qa_h}{2}(z_m - z_h), \\ C_{m,h} &= \frac{2q(mz_m - \eta) - qa_h(mz_m - hz_h)}{\eta} + (p-q)(h-1), \\ \Delta_{m,h} &= \delta - qa_h \left(1 - \frac{z_h}{z_m}\right). \end{aligned} \tag{20}$$

Every such policy has finite expected cost. In fact, for every integer $\ell \geq 0$,

$$\Pr(\tau > 1 + 2m\ell) \leq (1 - s^{2m})^\ell. \tag{21}$$

All terminal posteriors actually reached by the policy are strict. Neither the policy nor the reported outcomes requires posterior values to be computed during execution.

The boundary m is the classical symmetric Bernoulli sequential test (Girshick, 1946; Pabbaraju et al., 2026), which alone would move the two arms together. The cutoff h breaks this symmetry, and (19) prices the resulting tradeoff. Specifically, $N_m(j)$ is the expected number of further calls of the paired test from $|k| = j$ (Lemma G.1), so it counts the calls saved by imposing a cutoff at j on a run of C reports; $f_{m,j}(c)$ balances the forgone accuracy against the price of those calls, and $c\epsilon j$ charges the pure positive records, which answer C forever and consume computation without improving accuracy. For every finite N, the next result constructs a population depending on N. A cap transfer argument then provides a single finite cap that sustains these sign changes.

Theorem C.2 (Arbitrarily many optimal price sign cycles)**.** *For every $s \in (0, 1/2)$ and integer $N \geq 1$, there exist $0 < q < p < 1/2$ with $\delta = 1 - 2p > 0$ such that, for every $m = 2, \ldots, N+1$, the price*

$$b_m = \frac{\eta a_{m-1}(z_m - z_{m-1})}{2\{a_{m-1}[mz_m - (m-1)z_{m-1}] + \epsilon\eta\}} \tag{22}$$

satisfies $c_m < b_m < c_{m-1}$. The unique optimal regimes under the price objective are

price interval	h	*optimal effect*
(b_m, c_{m-1})	$m-1$	$\delta - qa_{m-1}\left(1 - \frac{z_{m-1}}{z_m}\right) < 0$
(c_m, b_m)	m	$\delta > 0$.

(23)

Immediate release for $c > c_1$ has effect $\delta > 0$. Selecting one price from each displayed interval therefore gives $2N+1$ decreasing prices with alternating effect signs and strictly increasing accuracy at every transition. The same population and the same four response vectors apply at every price. Moreover, every record is semantically correct with probability at least $1 - s$.

Theorem 3.1 follows by selecting one price in each strict interval and an initial price above c_1.

As the price falls below c_{m-1}, the paired boundary rises to m, while the all C cutoff remains at $m-1$ because pure positive records make further calls costly. Below b_m, the cutoff rises to m and restores the effect to δ. Both transitions improve accuracy.

Producing more cycles requires a smaller signal and larger expected call counts. The population is chosen after N is fixed, and (45) together with (46) requires both the true effect and the pure positive mass to decrease as N grows. This dependence is what narrows the intervals below. For fixed s and m, the negative interval has width

$$c_{m-1} - b_m = \frac{c_{m-1}\epsilon\eta}{a_{m-1}[mz_m - (m-1)z_{m-1}] + \epsilon\eta} = O(\epsilon c_{m-1}) \qquad (\epsilon \downarrow 0), \tag{24}$$

where the implicit constant depends on s and m. The displayed adjacent cutoffs also require $\delta < 1/[(N+1)2^N + 1]$ (Appendix O), which places δ below 10^{-4} when $N = 10$. Thus, additional cycles are obtained by shrinking the signal. The reversals characterize the transition as computation becomes cheaper, rather than the limit of arbitrarily cheap computation. For any fixed population, accuracy converges to one and every optimal effect converges to δ as $c \downarrow 0$.

For $(p, q, s) = (.49, .489, .40)$, begin at price .04. For each $m = 2, 3, 4$, select the midpoint of the corresponding negative interval, followed by $(c_m + c_{m-1})/2$ in the positive regime. Figure 2 plots the seven resulting pairs obtained from (20). The true effect is .02, every record is semantically correct with probability at least .60, and the negative intervals have width approximately 2×10^{-5} because $p - q = .001$.

C.1 Geometric tails for the finite cap transfer

For each strict price used in the repeated reversal construction, the optimal rule visits only finitely many relevant states: the paired walk between its two boundaries, the unrevised all C prefix before its cutoff, and the pure A state. Every corresponding action gap is positive. If the walk boundary is m, then a run of $2m$ consecutive C reports forces release from every live state. Its conditional probability is at least s^{2m}, including for the pure C type. Consequently,

$$\Pr(\tau_{\pi^*} > 1 + 2mk) \le (1 - s^{2m})^k, \qquad k = 0, 1, \ldots. \tag{25}$$

To apply (48), choose prices strictly inside the regimes of Theorem C.2, and then make the geometric tail smaller than the finitely many products of action gaps and effect changes. A single sufficiently large T works simultaneously for every selected price. The proof gives an explicit tolerance based on the tail bound. This conclusion does not claim infinitely many changes for a single fixed positive effect.

The stability argument uses standard Bellman regret accounting. Here its role is to show that the phenomenon is not an artifact of the infinite horizon and to cover all capped optima, including nonunique and randomized ones.

D The exact structural region generated by stopping

To place the optimal price path within the full class of possible releases, we characterize every structural increment attainable under one fixed measurement law. The result uses binary outputs and

observable structural weights; it requires neither independent reports, a finite report alphabet, nor Bayesian decoding.

Chung et al. (2016) study consensus recrossings under repeated crowd judgments and adjust aggregate error counts using estimates of positive and negative switches. Their estimation problem differs from characterizing the exact attainable score region over all stopping policies.

The definitions and first revision representation are given in Theorem 6.1. We collect their consequences here.

The set $\mathcal{Z}_T$ in Theorem 6.1 is a zonoid, the standard expectation of random line segments (Koshevoy & Mosler, 1998). What is specific to stopping is that the entire adapted policy class reduces exactly to an initial selection followed by a first revision. Recrossings need not be enumerated. The maximizing rule requires only the initial direction $u^\top a$, the initial label, and observation of subsequent labels; it does not require estimation of r_T or a continuation value.

For a scalar unweighted mean, the result becomes

$$\{\mathbb{E}G_\tau : 1 \le \tau \le T\} = [\Pr(G_t = 1 \ \forall t \le T),\ \Pr(G_t = 1 \text{ for some } t \le T)]\,. \tag{26}$$

The endpoints are attained by releasing at the first zero or the first one, respectively, with release at T if the desired label never appears. For vector scores, the largest Euclidean displacement is

$$\sup_{z \in \mathcal{Z}_T} \|z\|_2 = \frac{1}{2} \sup_{\|u\|_2 = 1} \left\{ \mathbb{E}[|u^\top a| H_T] + \left| \mathbb{E}[(u^\top \Gamma) H_T] \right| \right\}. \tag{27}$$

In particular, when the initial score is valid, every release policy preserves it if and only if $aH_T = 0$ almost surely. This qualitative condition is the binary special case of the usual martingale criterion for invariance under stopping. Formula (14) additionally gives the exact magnitude and the attaining rules.

These are regions under a fixed law, without restrictions on accuracy or computation. The representing first revision rule need not preserve the cost or accuracy of the original policy. If the weight is hidden from the gate, even the support formula can fail: a hidden fair sign a, with $G_1 = 0$ and $G_2 = 1$ deterministically, gives zero mean increment under every admissible stopping rule, although $\mathbb{E}[a_+ H_2] = 1/2$. Such restrictions require projection onto the information actually available to the gate rather than an infeasible initial selection.

D.1 Vanishing terminal error does not shrink the policy region

Return to a binary truth X. Suppose the initial proxy is structurally anchored,

$$\mathbb{E}[a(X - G_1)] = 0, \tag{28}$$

and define weighted prediction error in direction u by $e_{t,u} = \mathbb{E}[|u^\top a| \mathbf{1}\{G_t \neq X\}]$.

Corollary D.1 (Consistency and persistent stopping exposure)**.** *For every finite T,*

$$h_T(\pm u) \ge [e_{1,u}/2 - e_{T,u}]_+. \tag{29}$$

The regions are nested: $\mathcal{Z}_T \subseteq \mathcal{Z}_{T+1}$ for every T, without a consistency assumption. If $G_t \to X$ almost surely, the limiting interval in direction u contains $[-e_{1,u}/2, e_{1,u}/2]$. If G_t is a Bayes decision for the exact posterior $S_t = \mathbb{E}[X \mid \mathcal{F}_t]$, where $\mathcal{H}_t \subseteq \mathcal{F}_t$, then $h_T(\pm u) \le e_{1,u}$ for every finite T, also without requiring consistency. With both Bayesian decoding and almost sure consistency,

$$\frac{e_{1,u}}{2} \le h_\infty(\pm u) \le e_{1,u}. \tag{30}$$

Both limiting constants are sharp over fixed, finite output, conditionally iid interfaces with a correct semantic mode on every record. Sharpness of the upper constant allows the positive mode gap to vary across interfaces.

The lower bound follows from a simple obstruction: every initially wrong label must eventually change if the terminal label becomes correct. The anchor balances the positive and negative weighted initial errors, so each side supplies attainable structural drift. The upper bound is a conditional maximal inequality for the posterior martingale. For the randomized treatment contrast, $a = 2(2D-1)$

and $e_{1,1} = 2\Pr(G_1 \neq X)$. Thus each limiting endpoint has magnitude at least the pooled initial error and at most twice that error, even though release at deterministic time T consistently estimates the original contrast as $T \to \infty$.

The consistency premise is compatible with one fixed repeated LLM interface. Suppose X is a function of the raw record and a known semantic decoder $g(W, R_t)$ satisfies

$$\Pr(g(W, R_t) = X \mid Z) \geq \tfrac{1}{2} + \kappa \qquad \text{almost surely,} \qquad \kappa > 0. \tag{31}$$

Conditionally iid calls make semantic majority consistent on every record. Hence X is measurable with respect to $\sigma(W, R_1, R_2, \ldots)$; the exact posterior converges to X, and so does its MAP label. For a deterministic number T of calls, when W is observed before the first report and $a = a(W)$, weighted Bayes optimality also gives

$$e_{T,u} \leq \mathbb{E}|u^\top a(W)|\, e^{-2\kappa^2 T}. \tag{32}$$

This concerns unselected deterministic length sequences. It does not apply the same bound after conditioning on a selected first answer or on an adaptive stopping event. Inserting (32) into (29) gives a finite horizon interval contained in the directional projection $\{u^\top z : z \in \mathcal{Z}_T\}$. Long run prediction guarantees therefore coexist with a nonvanishing range of scientific conclusions obtainable by earlier release. Budget restrictions select a smaller part of this range; the next sections compute that part exactly in one fixed three call experiment.

D.2 The four type interface at every horizon

The four record types in (4) also permit an exact evaluation of the unrestricted stopping region at every horizon. Keep the population, kernel and sticky MAP decoder fixed. Write

$$C_j = \frac{1}{j+1}\binom{2j}{j}, \qquad f_T(s) = \sum_{j=1}^{\lfloor (T-1)/2 \rfloor} C_j \{s(1-s)\}^j, \tag{33}$$

with an empty sum equal to zero. Here C_j is the classical Catalan first passage count, not a computation cost.

Theorem D.2 (Exact effect interval at every horizon)**.** *In the fixed experiment (4), the complete effect interval over all randomized, treatment aware stopping policies $1 \leq \tau \leq T$, without an additional computation constraint, is*

$$\{\Delta_\tau : 1 \leq \tau \leq T\} = \big[\delta - 2q f_T(s),\ \delta + 2q f_T(s)\big]. \tag{34}$$

Every endpoint is attained by an initial report selection followed by the first MAP revision. In particular $2q f_3(s) = L$, and

$$\lim_{T \to \infty} 2q f_T(s) = \frac{2qs}{1-s}. \tag{35}$$

In the same experiment, deterministic time MAP release satisfies $\Pr(G_T \neq X) \to 0$ and $\Delta_T \to \delta$.

After an initial B, the two remaining record types have posterior weights $1 - s, s$. Let the walk count C reports minus B reports, including the initial report. It starts at -1. Equal counts are posterior ties, so the sticky decoder first changes label when this walk reaches $+1$. The first passage after $2j$ additional reports has C_j possible orderings. Averaging their probabilities over the two record types gives $C_j\{s(1-s)\}^j$. An initial C has the reflected ambiguous component with mass q/p, while its pure C component never revises. These facts determine both supports in Theorem 6.1; no optimization over the exponentially many history dependent gates remains.

At $T = 3$, the budget optimal policy at $b_* = k_B L$ attains the full unrestricted negative endpoint $\delta - L$. It takes the least expensive route to that endpoint. The radius then grows to $e_1/(1-s)$, where $e_1 = 2qs$ is this population's initial prediction error. Thus the bounds e_1 and $2e_1$ from Corollary D.1 are approached within this four type family as $s \downarrow 0$ and $s \uparrow 1/2$, respectively. For every fixed $s < 1/2$, every record still has a strictly correct semantic mode. At (10), deterministic time effects converge to .10, while the limiting unrestricted interval is $[-.30, .50]$. These limits concern different release policies applied to the same infinite reference sampler.

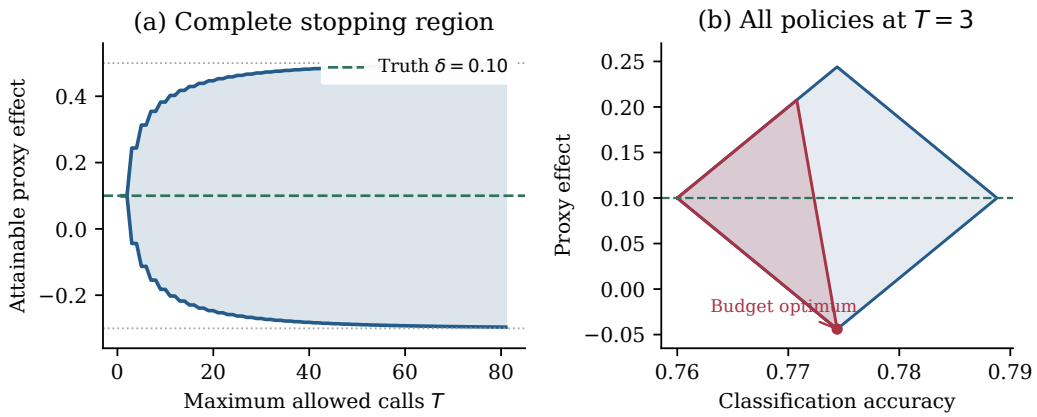


Figure 4: One unchanged sampler, different release policies. At (10), panel (a) evaluates the exact interval in Theorem D.2; its endpoints tend to $-.30$ and $.50$, while deterministic time MAP release is truth consistent. Panel (b) shows the full three call accuracy and effect region (blue) and the subset allowed by budget $b_* = .444$ (red). Its unique accuracy optimum attains the unrestricted minimum effect $-.044$. Both panels evaluate proved formulas without simulation.

E THE COMPLETE BUDGET, ACCURACY, AND EFFECT FRONTIER

The three call experiment permits an exact comparison of every admissible measurement policy within one fixed record population and one fixed sampling kernel. This includes policies chosen to protect the scientific effect, rather than only those that optimize prediction. Throughout this section, a policy may use treatment, the entire observed report history, and fresh independent randomization. It may stop after the first or second call and must stop after the third. Decoding is MAP, with the previous label retained at a posterior tie.

Recall the initial accuracy $A_0 = 1 - 2qs$, the true effect $\delta = 1 - 2p$, and

$$\alpha = 2s(1-s), \qquad \beta = \frac{q\alpha}{p}, \qquad L = q\alpha = p\beta, \qquad d = \frac{1}{2} - s, \qquad k_B = 1 + \frac{1}{\alpha} < k_C = 1 + \frac{1}{\beta}.$$

Let $b \geq 0$ be the population budget for additional calls after the first. Write u for the pooled probability that the policy makes a third call after BC or its barred counterpart, and v for the corresponding probability after CB or its barred counterpart. These probabilities refer to the decision to make the third call; they do not condition on its unseen answer.

Theorem E.1 (Exact frontier within one repeated sampling experiment)**.** *At additional call budget b, the complete set of attainable accuracy and effect pairs is*

$$\left\{ \left(A_0 + d(u+v),\ \delta - u + v\right) : 0 \leq u, v \leq L, \qquad k_B u + k_C v \leq b \right\}. \tag{36}$$

The minimum expected additional call cost for a specified pair (u, v) is exactly $k_B u + k_C v$. It is attained by a common gate that starts a second call with probability u/L after B and v/L after C, and makes a third call exactly when the first two reports disagree. Pure A paths and paths with equal answers stop. The treatment arm uses the reflected rule.

The unique accuracy optimal pair is

$$u_*(b) = \min\{L, b/k_B\}, \qquad v_*(b) = \min\{L, (b - k_B L)_+ / k_C\}. \tag{37}$$

Put $t_*(b) = u_*(b) + v_*(b)$ and $A_*(b) = A_0 + dt_*(b)$. For every exact accuracy regret $0 \leq R \leq dt_*(b)$, let

$$\begin{aligned} t &= t_*(b) - R/d, \\ u_- &= \max\left\{0, t - L, \frac{k_C t - b}{k_C - k_B}\right\}, \qquad u_+ = \min\{t, L\}. \end{aligned} \tag{38}$$

The complete effect interval among policies with regret R is

$$[\,\delta + t - 2u_+,\ \delta + t - 2u_-\,]. \tag{39}$$

Every point and every endpoint is attained within the same fixed experiment.

The coefficients k_B and k_C are effective costs rather than assumed query prices. A unit of useful third call probability after B requires at least $1/\alpha$ units of second call probability to reach it, in addition to the third call itself. The analogous cost after C is larger because $\beta < \alpha$. Each useful third call contributes exactly the same accuracy gain d, but the two routes move the scientific effect in opposite directions. This is where nonmyopic computation and the structural target enter the same feasible set.

Only the optimal pair in Equation (37) is always unique. Partial querying within a first report stratum may admit multiple treatment aware implementations. At $b_* = k_B L$, however, the accuracy optimal stopping policy itself is unique up to null events: query every B report, continue along its BC paths, and stop elsewhere. The budget rules out wasted calls. The region in Equation (36) imposes an upper bound on cost; it does not assert that every policy producing the same pair has minimum cost. At this budget, the prediction optimum also attains $\delta - L$, the smallest effect attainable by any three call stopping policy even without a budget constraint. This follows directly from $u \le L$ and $v \ge 0$.

Theorem E.2 (Exact prediction cost of scientific validity)**.** *Let $R_{\mathrm{valid}}(b)$ be the smallest accuracy regret among policies with effect exactly δ, and let $R_{\mathrm{sign}}(b)$ be the smallest regret among policies with a nonnegative effect. Define*

$$b_{\mathrm{full}} = (k_B + k_C)L, \qquad H(b) = \left[\min\left\{\frac{b}{k_B}, \frac{b_{\mathrm{full}} - b}{k_C}\right\}\right]_+, \qquad \chi = \frac{d(k_C - k_B)}{k_B + k_C}. \tag{40}$$

Then

$$R_{\mathrm{valid}}(b) = \chi H(b), \qquad R_{\mathrm{sign}}(b) = \chi[H(b) - \delta]_+. \tag{41}$$

Both minimum costs are attained by common policies whose continuation probabilities are nondecreasing in posterior uncertainty at each query stage.

When $0 < \delta < L$, the accuracy optimum has the wrong sign exactly for budgets satisfying

$$k_B\delta < b < (k_B + k_C)L - k_C\delta. \tag{42}$$

At any such budget, every policy with regret strictly below $R_{\mathrm{sign}}(b)$ remains wrong signed, while a policy with regret exactly $R_{\mathrm{sign}}(b)$ attains effect zero. Likewise, $R_{\mathrm{valid}}(b)$ is the exact cost of restoring the original effect. The maximum repair costs over budgets occur at $b_* = k_B L$:

$$\max_b R_{\mathrm{valid}}(b) = \chi L, \qquad \max_b R_{\mathrm{sign}}(b) = \chi(L - \delta)_+. \tag{43}$$

For exact validity, choose

$$u = v = \min\{L, b/(k_B + k_C)\}.$$

Inside the interval in Equation (42), the least regret sign repair uses

$$u = v + \delta, \qquad v = \frac{b - k_B\delta}{k_B + k_C}.$$

Outside that interval, the prediction optimum already has the desired sign. In all these cases $u \ge v$. Since the first B is more uncertain than the first C, the initial continuation probabilities u/L and v/L are ordered by uncertainty. At the next stage, the gate continues only at posterior $1/2$ disagreements. Scientific repair therefore does not require abandoning the uncertainty ordering or changing the sampler, decoder, or record population.

For $(p, q, s) = (9/20, 3/10, 2/5)$, we have $A_0 = .76$, $\delta = .10$, $L = .144$, $k_B = 37/12$, and $k_C = 33/8$. At $b_* = .444$, the unique accuracy optimum has accuracy .7744 and effect $-.044$. Its exact regrets for sign repair and full validity are respectively $11/17300$ and $9/4325$. The complete interval of budgets for which the accuracy optimum has the wrong sign is $(37/120, 1251/2000)$. These are evaluations of the complete frontier of one fixed experiment, not comparisons across alternative measurement channels.

F THE KNOWN RESPONSE CATALOG AND THE RANGE OBSTRUCTION

Keep the four control response vectors

$$\begin{array}{c|ccc} X & Q_A & Q_B & Q_C \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1-s & s \\ 1 & 0 & s & 1-s \\ 1 & 0 & 0 & 1 \end{array} \qquad 0 < s < 1/2, \tag{44}$$

together with their reflected treatment versions. The value of s and the six token orientation are known, while the positive mixing weights may be arbitrary and unknown, separately in each treatment arm. In particular, the procedure does not require p, q, their induced prior, initial scientific validity, or any posterior probability. This is a known catalog assumption; it does not assert that repeated unvalidated reports reveal how the response vectors relate to truth.

Proposition F.1 (Why the correction must leave the label range)**.** *Fix a supported control oriented prefix containing no A, and suppose both noisy types remain compatible with all retained information. Let a correction take values in $[0,1]$ and use a number of fresh calls that is finite almost surely under each compatible raw type. Then the correction cannot be unbiased for both noisy types, and hence not for every remaining raw type. No deterministic cap and no bound on expected computation is assumed.*

This is the classical Bernoulli factory endpoint obstruction (Keane & O'Brien, 1994; Jacob & Thiery, 2015): the two noisy types assign positive probability to every finite B/C transcript but have opposite truths. The obstruction concerns recordwise unbiasedness uniformly over compatible types. By contrast, a posterior constructed from a known population prior can preserve a population criterion through the tower property without satisfying this stronger requirement, so the two statements are consistent.

The construction uses finite polynomial interpolation followed by conditional symmetrization, both classical tools. What is specific here is the exact correction and the minimum additional depth after every possible production prefix in the same repeated sampler that generates the dynamic reversal. More general unknown response propensities, constrained only by a correct mode gap, require a different analysis; the known catalog result does not address that broader problem.

G EXACT INFINITE HORIZON POLICIES AND REPEATED PRICE REVERSALS

G.1 THE PAIRED LIKELIHOOD RATIO PROBLEM

After at least one B, neither pure record type remains possible. Let $k = \#C - \#B$ and $Y = 2X - 1$. The posterior satisfies

$$\mathbb{E}[Y \mid k] = z_k, \qquad z_k = \frac{t^k - s^k}{t^k + s^k} = \tanh\left(\frac{k}{2}\log\frac{t}{s}\right),$$

where $z_{-k} = -z_k$. Conditional on Y, each additional count increment has mean ηY. Write $P_j = (1 + z_j)/2$ for $j \geq 0$.

The knots c_j agree with the lower endpoints in Pabbaraju et al. (2026, Theorem 2, Remark 2, equation (14)), after setting their odds ratio to $(1-s)/s$ and their margin to $(1-2s)/2$. The paired test inherits the classical sequential analysis structure (Girshick, 1946); we include a direct verification because the all C cutoff and its consequences for the signed effect depend on the exact values.

Lemma G.1 (Explicit paired stopping boundary)**.** *The sequence c_j in Equation (17) is strictly decreasing to zero. If $c_m < c < c_{m-1}$, the unique optimal paired decision continues for $|k| < m$ and stops for $|k| \geq m$. Starting from $|k| = j \leq m$, its accuracy and expected number of further calls are P_m and $N_m(j)$, respectively. Its value at an interior state is $P_m - cN_m(j)$. If $c \geq c_0$, immediate stopping is optimal at every paired state.*

Proof. Let $\Delta z_j = z_{j+1} - z_j$. These increments are strictly positive and strictly decreasing because $\tanh(x)$ is strictly concave on the positive half line. Moreover,

$$c_j = \frac{\eta}{2\{j + 1 + z_j/\Delta z_j\}}.$$

The denominator is strictly increasing and diverges, which proves the claim about the price sequence.

Let T_m be the first exit from $\{-m+1,\dots,m-1\}$. From any interior state, $2m$ consecutive C reports force an exit, and their conditional probability is at least s^{2m}. Hence T_m has a geometric tail and finite expectation. At exit, $k=\pm m$, so the Bayes accuracy is always P_m. Conditional on Y, the process $Y(k_n-k_0)-\eta n$ is a martingale. Optional sampling is justified by the finite exit expectation and bounded stopped positions, giving

$$\eta\mathbb{E}[T_m \mid k_0=k]=\mathbb{E}[Yk_{T_m}\mid k_0=k]-kz_k=mz_m-kz_k.$$

The last equality conditions on the exit history, where $\mathbb{E}[Y \mid k_{T_m}]=\operatorname{sign}(k_{T_m})z_m$. This proves the formula for $N_m(j)$.

We verify optimality directly, without invoking a boundary theorem. Set $v(k)=P_m-cN_m(|k|)$ for $|k|<m$, and $v(k)=P_{|k|}$ otherwise. At every interior state, the first step decomposition of the exit rule gives the continuation equality. Moreover, for $j<m$,

$$P_m-cN_m(j)-P_j=\sum_{i=j}^{m-1}\left\{\frac{z_{i+1}-z_i}{2}-\frac{c}{\eta}\big[(i+1)z_{i+1}-iz_i\big]\right\}>0,$$

because $c<c_{m-1}\le c_i$. Thus continuation strictly dominates stopping in the interior.

At $k=m$, the one step continuation advantage computed with v is affine and strictly decreasing in c; its slope is $-1-P(k\to m-1)N_m(m-1)<0$. It vanishes at $c=c_m$. To verify this equality, compare the threshold m and threshold $m+1$ values on $|k|\le m$. Their difference is

$$\frac{z_{m+1}-z_m}{2}-\frac{c}{\eta}\big[(m+1)z_{m+1}-mz_m\big],$$

which is zero at c_m; the threshold $m+1$ continuation equality at m therefore gives the claimed equality for v. Hence stopping strictly dominates at m for $c>c_m$, and likewise at $-m$. For $|k|>m$, both successors remain on the same side of the MAP threshold; their expected immediate accuracy equals the current posterior accuracy, so making another call loses exactly c.

The resulting bounded function v satisfies the Bellman inequality everywhere and equality along the exit policy. Applying the inequality to any bounded stopping time and then passing to any stopping time with finite mean proves that its value is an upper bound. Boundedness of v and integrability of the call count justify the limit. Policies with infinite expected cost have value $-\infty$ and cannot improve the bound. The exit rule attains the bound. Strict action inequalities give uniqueness up to histories that are never reached. Finally, if $c\ge c_0=\eta/2$, the policy that stops everywhere is a Bellman upper bound: a single report improves MAP accuracy only at $k=0$, where its gain is $\eta/2$. This proves the remaining assertion. □

G.2 The all C cutoff and the complete price calculation

Proof of Theorem C.1. Let $\rho_C=p-q$ be the pure C mass. At history C^n, the paired mass is qa_n, the pure C mass is ρ_C, and the negative mass is qs^n. Its MAP label is strictly positive. If a B now occurs, its paired state is $n-1$, and its unconditional probability mass is $q(t^ns+s^nt)$. Until that first B, the only possible observed history is another all C history.

No optimal policy continues at C^n for $n\ge m$. Within the paired component, stopping at state $n\ge m$ is already optimal; within the pure C component, no query can improve accuracy and every query costs c. Revealing component membership could only increase the best continuation value. Even with that additional information, continuation has no advantage over stopping in either component. It therefore cannot have an advantage in their mixture, and the positive pure C mass makes continuation strictly inferior.

It remains to consider a cutoff $h\in\{1,\dots,m\}$ before the first B, followed by the optimal paired rule. Every deterministic policy must have such a first cutoff along its all C path. Randomization cannot exceed the maximum over these finitely many deterministic choices. Define $F_n=qf_{m,n}(c)$. Multiplying the paired Bellman equality by its history mass gives, for $1\le n<m$,

$$F_n-F_{n+1}=q(t^ns+s^nt)\{v(n-1)-P_{n-1}\}-cqa_n.$$

Immediate MAP accuracy is unchanged by the next report, including at the posterior tie when $n = 1$. In the full mixture, one step also costs $c\rho_C$ on pure C records. Consequently, continuing along the all C path from $n = 1$ to cutoff h has total unnormalized net advantage

$$\sum_{n=1}^{h-1}\{F_n - F_{n+1} - c\rho_C\} = F_1 - F_h - c\rho_C(h-1).$$

Maximizing this quantity is exactly (19). If the minimizer is unique, every reached earlier all C state strictly continues and the cutoff state strictly stops; otherwise, two tied cutoff choices would have the same initial value. Together with Lemma G.1, this proves the uniqueness statement separately in both arms. For $c > c_1$, the first paired state $|k| = 1$ is optimally stopped, and the pure C component only makes continuation after C more expensive; hence every first report is released immediately.

To calculate the outcomes, decompose the raw population; this is only a calculation and does not permit the policy to observe raw type. In the paired component, the symmetric threshold m rule would have accuracy gain $2q(P_m - P_1)$ and further call cost $2qN_m(1)$. The implemented rule instead stops positively at an all C history C^h, whose paired mass is qa_h. Relative to the paired exit calculation, this loses accuracy $qa_h(P_m - P_h)$ and saves $qa_hN_m(h)$ calls. Pure C records remain correct but use $h - 1$ further calls, contributing $\rho_C(h - 1)$. This proves the first two lines of (20).

The symmetric paired exit calculation has positive mean q; including the pure C mass gives the initial mean p. From paired state h, the posterior martingale implies that the probability of eventually exiting at $+m$ is

$$\frac{(1+z_h)/2 - (1-z_m)/2}{z_m} = \frac{1}{2}\left(1 + \frac{z_h}{z_m}\right).$$

Stopping positively at C^h therefore increases the control hard mean by $qa_h(1-z_h/z_m)/2$. Reflection doubles this change with negative sign in the treatment contrast, yielding the effect formula.

The only released paired states are $\pm m$, whose posteriors are strict, and C^h, which is strictly positive; A is certain. From any active state, $2m$ consecutive C reports force either the all C cutoff or the upper paired boundary. Conditional on the raw record, these reports have probability at least s^{2m}, and probability one for the pure C type. Iterating over blocks proves (21). At $c = c_m$, the two adjacent paired values agree for every $|k| \le m$. A cutoff beyond m cannot improve the all C value because the paired continuation advantage at m is zero while the pure C component incurs a strictly positive additional cost. Thus either adjacent boundary, together with its minimizing cutoff, is optimal at the knot. The exact formula also gives $C_{m,h} \le 2qm/\eta + \rho_C(m-1) < \infty$. □

G.3 Simultaneous parameter bounds for any finite number of cycles

Proof of Theorem C.2. Fix s and N. All quantities in this paragraph depend only on them. For $m = 2, \dots, N+1$, define

$$E_m = \min_{1\le j\le m-2} \frac{f_{m,j}(c_{m-1})}{c_{m-1}(m-1-j)}, \qquad B_m = \frac{f_{m,m-1}(c_m)}{c_m},$$

where the empty minimum for $m = 2$ is $+\infty$. At $c = c_{m-1}$, $f_{m,m-1} = 0$, whereas $f_{m,j} > 0$ for $j < m - 1$ by the strict incremental inequalities in Lemma G.1. Also $f_{m,m-1}(c_m) > 0$ because $c_m < c_{m-1}$. Thus all finite bounds are strictly positive. Choose

$$0 < \epsilon < \min_{2\le m\le N+1}\{E_m, B_m\}. \tag{45}$$

We first show that throughout $c_m < c < c_{m-1}$, no cutoff below $m - 1$ can minimize (19). For $j < m - 1$, write $f_{m,j}(c) - f_{m,m-1}(c) = U - cV$. Its intercept is

$$U = \frac{1}{2}\{a_j(z_m - z_j) - a_{m-1}(z_m - z_{m-1})\} > 0,$$

because a_j decreases and z_j increases with $j \ge 1$. Hence $(f_{m,j} - f_{m,m-1})/c = U/c - V$ is decreasing in c. For $c \le c_{m-1}$, it is therefore at least

$$\frac{f_{m,j}(c_{m-1})}{c_{m-1}} > \epsilon(m-1-j).$$

The cutoff objective at j is consequently strictly larger than at $m-1$. Only $h = m-1$ and $h = m$ remain.

Their objectives are equal exactly when $f_{m,m-1}(c) = c\epsilon$. Solving gives (22). This price lies strictly below c_{m-1}, while $\epsilon < B_m$ places it strictly above c_m. Hence, for $b_m < c < c_{m-1}$, the unique cutoff is $m-1$, whereas for $c_m < c < b_m$, it is m. This establishes the policy pattern simultaneously for all requested m.

It remains to choose a single population producing the desired signs. Set

$$D_N = \min_{2 \le m \le N+1} a_{m-1} \left(1 - \frac{z_{m-1}}{z_m}\right) > 0.$$

Choose

$$0 < \delta < \frac{D_N}{2(1+\epsilon) + D_N}, \qquad p = \frac{1-\delta}{2}, \qquad q = \frac{p}{1+\epsilon}. \tag{46}$$

Then $0 < q < p < 1/2$ and $(p-q)/q = \epsilon$, so every raw mass is strictly positive. The displayed bound is exactly $\delta < qD_N$. Equation (20) therefore makes every $h = m-1$ effect strictly negative and every $h = m$ effect equal to $\delta > 0$. Starting from any $c > c_1$ and decreasing through the nonempty intervals in (23) produces at least $2N$ strict sign changes. Choosing prices in the interior of each interval makes every optimal action strict and the policy unique on reached histories.

Accuracy strictly increases at both kinds of transition. When the paired boundary grows from $m-1$ to m while the all C cutoff remains at $m-1$, the increase is

$$q\left(1 - \frac{a_{m-1}}{2}\right)(z_m - z_{m-1}) > 0.$$

When the all C cutoff then moves from $m-1$ to m, the increase is $qa_{m-1}(z_m - z_{m-1})/2 > 0$. The same formulas cover the first transition from immediate release by taking $m - 1 = 1$. Each selected price has finite m and therefore finite expected computation by Theorem C.1. Only the population weights depend on N; the four response vectors remain fixed for the selected s. □

Low price consistency and the order of quantifiers. Fix any one admissible (p, q, s). Let $A(c)$ denote the accuracy of any price optimal policy. A deterministic K call policy is feasible for the infinite horizon price problem, so

$$A(c) \ge \sup_{\tau} J_c(\tau) \ge \Pr(G_K = X) - c(K-1).$$

First let $c \downarrow 0$ and then $K \to \infty$. Correct recordwise modes and the i.i.d. kernel give $\Pr(G_K = X) \to 1$, hence $A(c) \to 1$. The treatment effect error is at most twice the pooled classification error, so every optimal effect converges to $\delta > 0$ and is eventually positive. Thus Theorem 3.1 selects one population for each finite N; it does not claim infinitely many wrong signs for one fixed positive δ.

The widths of the negative phases are explicitly

$$c_{m-1} - b_m = \frac{c_{m-1}\epsilon\eta}{a_{m-1}[mz_m - (m-1)z_{m-1}] + \epsilon\eta}.$$

They may be small when the pure C anchor is small. The strict price intervals establish a population possibility result, not a lower bound on their prevalence in deployed systems.

H Finite cap transfer and stability of scientific effects

Proposition H.1 (Finite cap and regret stability)**.** *Keep the sampler, observable information, and MAP decoder fixed. Suppose a deterministic infinite horizon optimum π^* has finite expected calls. Suppose also that every decision it reaches, including its release decisions in both arms, has a strict Bellman action gap of at least $\gamma_c > 0$. Then every admissible policy π with finite expected calls satisfies*

$$|\Delta_\pi - \Delta_{\pi^*}| \le \frac{2\{J_c^* - J_c(\pi)\}}{\gamma_c}. \tag{47}$$

Let $J^*_{c,T}$ be the optimum under $\tau \le T$. Every policy satisfying $J_c(\pi) \ge J^*_{c,T} - R$, $R \ge 0$, obeys

$$|\Delta_\pi - \Delta_{\pi^*}| \le \frac{2}{\gamma_c} \{R + \Pr(\tau_{\pi^*} > T)\}. \tag{48}$$

The effects are evaluated in the same randomized treatment population, and the result allows randomized policies that depend on treatment. The corresponding absolute accuracy difference is bounded by one half of the right hand side of (47) *or* (48)*.*

Proof of Proposition H.1. Couple π and π^* using one infinite reference report path and independent policy randomization. Let D_π be the first time their stop or continue decisions differ, and set $D_\pi = \infty$ if they make the same release decision. On $\{D_\pi < \infty\}$, the history up to that time is reached by π^*, so choosing the opposite action incurs a conditional Bellman loss of at least γ_c.

To justify this accounting for unbounded stopping, define the nonnegative action gaps

$$g_{\text{stop}}(h) = \mathcal{V}(h) - A(h), \qquad g_{\text{cont}}(h) = \mathcal{V}(h) + c - \mathbb{E}[\mathcal{V}(H_{t+1}) \mid H_t = h].$$

Bayes reward is bounded and $c > 0$, so $1/2 \le \mathcal{V} \le 1$. The Bellman equation makes both gaps nonnegative, with zero gap for the optimal action. Telescoping conditional expectations through the stopping time gives

$$J^*_c - J_c(\pi) = \mathbb{E}\left[\sum_{t=1}^{\tau_\pi - 1} g_{\text{cont}}(H_t) + g_{\text{stop}}(H_{\tau_\pi})\right]. \tag{49}$$

One may first truncate the sum at a deterministic time and then pass to the limit. The bounded value process and finite expected τ_π justify the stopped martingale calculation; equivalently, the absolute expectation of the sum of its bounded increments is finite. The initial value is averaged over the common first report law, whose call cost is excluded from every objective.

Whenever $D_\pi < \infty$, the action at D_π contributes at least γ_c to the right hand side of (49), while all other terms are nonnegative. Hence

$$\Pr(D_\pi < \infty) \le \frac{J^*_c - J_c(\pi)}{\gamma_c}.$$

If no decision differs, the released labels are identical. Since the randomized treatment weight is $a = 2(2D - 1)$, with $|a| = 2$,

$$|\Delta_\pi - \Delta_{\pi^*}| = |\mathbb{E}[a\{G_{\tau_\pi} - G_{\tau_{\pi^*}}\}]| \le 2 \Pr(D_\pi < \infty).$$

This proves (47) without requiring common gates across treatment arms. Here D_π denotes a disagreement time, while D without a subscript continues to denote treatment assignment. The two correctness indicators also agree whenever the policies never differ, so their accuracy difference is bounded by $\Pr(D_\pi < \infty)$, proving the corresponding accuracy statement.

Now truncate π^* at T. Its correctness indicator changes only on $\{\tau_{\pi^*} > T\}$ and by at most one in absolute value, while truncation can only save computation. Therefore

$$0 \le J^*_c - J_c(\pi^* \wedge T) \le \Pr(\tau_{\pi^*} > T).$$

Since the truncated rule is an admissible competitor for the capped problem, $J^*_{c,T} \ge J_c(\pi^* \wedge T)$. Any policy that is within R of the capped optimum therefore has infinite horizon objective regret at most $R + \Pr(\tau_{\pi^*} > T)$. Substituting this bound into (47) proves (48). □

Tail bound and one common finite cap. Under the four type law, every active record emits C on each new call with probability at least s, conditional on its raw type. A block of $2m$ successive C reports either reaches the all C cutoff, which is no larger than m, or sends the paired walk from any interior state to its upper boundary. Conditional on the raw type and the past, this block has probability at least s^{2m}. Conditioning only on the observed past preserves this lower bound. Repeated conditioning over disjoint blocks proves (25); barred treatment reports satisfy the same bound.

For a finite collection of strict prices $\rho_1, \ldots, \rho_K$, write $e_j = |\Delta_{\pi^*_j}| > 0$, and let m_j and γ_j denote the corresponding boundary sizes and minimum positive action gaps. Choose an integer k_j such that

$(1-s^{2m_j})^{k_j} < \frac{\gamma_j e_j}{4}$, and take $T \geq \max_j\{1+2m_jk_j\}$. Then every exactly optimal capped policy at price ρ_j has an effect within $e_j/2$ of its infinite horizon counterpart and therefore has the same sign. More generally, the sign is preserved whenever its capped objective regret is strictly less than $\frac{\gamma_j e_j}{2} - \Pr(\tau_{\pi_j^*} > T)$. Thus one finite cap and one explicit positive tolerance work simultaneously for the entire finite list of prices. No uniform statement is made as the list length tends to infinity with the population held fixed.

For the accuracy assertion in Corollary 3.2, let $d_A > 0$ be the minimum of the finitely many successive accuracy increments in Theorem 3.1. Increase the common T further until $\frac{\Pr(\tau_{\pi_j^*}>T)}{\gamma_j} < \frac{d_A}{3}$ for every selected price. The accuracy part of Proposition H.1 then places each capped optimum within $d_A/3$ of its corresponding infinite horizon accuracy. Hence two successive capped accuracies differ by at least $d_A/3 > 0$.

I Proofs for the stopping region

Proof of Theorem 6.1. Take the reference report path independently of all policy randomization. For an arbitrary randomized admissible τ, put $J_\tau = \mathbf{1}\{G_\tau \neq G_1\}$. Binary support implies

$$G_\tau - G_1 = (1-2G_1)J_\tau, \qquad 0 \leq J_\tau \leq H_T.$$

Consequently, $k_\tau = \mathbb{E}[J_\tau \mid \mathcal{H}_1]$ lies between zero and r_T. Here the conditional expectation integrates both independent policy randomization and the reference continuation. Set $\lambda_\tau = k_\tau/r_T$ on $\{r_T > 0\}$ and zero otherwise. Since Γ is initially measurable,

$$\mathbb{E}[a(G_\tau - G_1)] = \mathbb{E}[\Gamma k_\tau] = \mathbb{E}[\lambda_\tau \Gamma r_T].$$

This proves one inclusion without assuming a Markov property or conditional independence of successive reports.

Conversely, let $\lambda \in [0,1]$ be initially measurable and draw $U \sim \mathrm{Unif}[0,1]$ independently of the entire reference law. Stop at time one on $\{U > \lambda\}$, and otherwise stop at $\sigma \wedge T$. This is an admissible stopping time: whether to continue is determined by current and past information, and the first revision is observable when it occurs. Its final disagreement indicator is $\mathbf{1}\{U \leq \lambda\}H_T$, which has the claimed expected increment. This also covers $T = 1$ and $r_T = 0$.

The order interval $\{\lambda \in L^\infty(\mathcal{H}_1) : 0 \leq \lambda \leq 1\}$ is weak star compact and convex. Since every coordinate of Γr_T is integrable, its image under $\lambda \mapsto \mathbb{E}[\lambda \Gamma r_T]$ is compact and convex in $\mathbb{R}^K$. Maximizing in direction u selects $\lambda = \mathbf{1}\{u^\top \Gamma > 0\}$, giving

$$h_T(u) = \mathbb{E}[(u^\top \Gamma)_+ r_T] = \mathbb{E}[(u^\top \Gamma)_+ H_T].$$

Adding the expressions for u and $-u$ proves the width formula. □

For (26), apply the support formula with $a = 1$. Its lower endpoint is

$$\mathbb{E}G_1 - \Pr(G_1 = 1, H_T = 1) = \Pr(G_t = 1 \ \forall t \leq T),$$

while its upper endpoint is

$$\mathbb{E}G_1 + \Pr(G_1 = 0, H_T = 1) = \Pr(\exists\, t \leq T : G_t = 1).$$

For (27), duality of the Euclidean norm gives

$$\sup_{z \in \mathcal{Z}_T} \|z\|_2 = \sup_{\|u\|_2 = 1} h_T(u), \qquad h_T(u) = \frac{1}{2}\left\{\mathbb{E}[|u^\top a| H_T] + \mathbb{E}[(u^\top \Gamma) H_T]\right\}.$$

The sphere is symmetric, so the second term may be replaced by its absolute value inside the supremum. Finally, $\mathcal{Z}_T = \{0\}$ if and only if its width is zero in every coordinate direction, equivalently $aH_T = 0$ almost surely. Adding the initial mean score gives the stated universal validity criterion.

Proof of Corollary D.1. Fix u and write $c = u^\top a$, $v = c(1-2G_1)$, and $I_t = \mathbf{1}\{G_t \neq X\}$. Since X and G_1 are binary, $c(X - G_1) = vI_1$. The anchor therefore implies

$$\mathbb{E}[v_+ I_1] = \mathbb{E}[(-v)_+ I_1] = e_{1,u}/2. \tag{50}$$

On $\{I_1 = 1, H_T = 0\}$, the terminal label still equals the incorrect initial label, so $I_T = 1$. The exact support formula then gives the slightly sharper finite horizon bounds

$$\begin{aligned} h_T(u) &\geq [e_{1,u}/2 - \mathbb{E}\{|c|\mathbf{1}\{v > 0\}I_T\}]_+ , \\ h_T(-u) &\geq [e_{1,u}/2 - \mathbb{E}\{|c|\mathbf{1}\{v < 0\}I_T\}]_+ . \end{aligned} \tag{51}$$

These imply (29). Integrability of a and $G_t \to X$ almost surely imply $e_{t,u} \to 0$ by dominated convergence. The regions are nested because every stopping policy admissible at an earlier horizon remains admissible at a later one. Equivalently, r_T increases with T, and (14) gives monotone convergence of each support to $h_\infty(u) = \mathbb{E}[v_+ H_\infty]$. A scalar projection of a compact convex set is the interval $[-h_T(-u), h_T(u)]$, proving the limiting interval assertion.

For the upper bound, put $U_1 = \min(S_1, 1 - S_1)$. On $\{G_1 = 0\}$, a first revision requires the nonnegative martingale S_t to reach at least $1/2$; on $\{G_1 = 1\}$, apply the same argument to $1 - S_t$. The conditional maximal inequality gives

$$\Pr(H_T = 1 \mid \mathcal{F}_1) \leq 2U_1. \tag{52}$$

Bayes ties cause no difficulty: every possible revision is contained in the corresponding weak threshold hitting event. Since v_+ is $\mathcal{F}_1$ measurable and $U_1 = \mathbb{E}[I_1 \mid \mathcal{F}_1]$,

$$h_T(u) \leq 2\mathbb{E}[v_+ U_1] = 2\mathbb{E}[v_+ I_1] = e_{1,u}.$$

Applying the same argument to $-v$ gives the other direction. Letting T increase completes the proof. □

Correct semantic mode and the deterministic time bound. Under (31), the conditional strong law implies that semantic majority converges to X for almost every raw record. Hence X is measurable with respect to $\mathcal{F}_\infty = \sigma(\cup_t \mathcal{F}_t)$, and the upward convergence theorem for conditional expectations gives $S_t \to X$ almost surely. Any Bayes tie convention then yields $G_t \to X$. For a deterministic T, Hoeffding's inequality conditional on Z bounds majority error by $e^{-2\kappa^2 T}$, including any majority tie rule. Integrating conditional on the covariate W observed before reporting preserves this bound. Bayes classification minimizes error conditional on the report history, and hence also conditional on W. Multiplying by the nonnegative W measurable weight $|u^\top a(W)|$ and integrating proves (32). This argument does not assert the same bound after conditioning on a first report selection that is itself part of the majority sample.

Sharp lower constant with one fixed iid kernel. Let W be a fair binary covariate and let $\Pr(X \neq W \mid W) = s$, with $0 < s < 1/2$. Fix $\rho \in (1/2, 1)$. Each independent call on record $Z = (W, X)$ emits a verified token V_X with probability ρ and a default token E_W otherwise. There are four possible tokens, whose semantic label is the subscript. A verified token reveals X, whereas a default token contains no additional information given W. Every record has semantic accuracy at least ρ. For the scalar mean, take $a = 1$. Symmetry gives $\mathbb{E}G_1 = \mathbb{E}X = 1/2$ and initial error $e_1 = (1 - \rho)s$. Initially correct labels never change, while initially incorrect labels change exactly at the next verified token. Therefore

$$h_T(1) = h_T(-1) = \frac{e_1}{2}\big[1 - (1 - \rho)^{T-1}\big], \qquad e_T = s(1 - \rho)^T. \tag{53}$$

Thus the lower limiting constant $1/2$ is attained exactly for every fixed $\rho > 1/2$. This example also attains the sharper sector error inequalities in (51).

Sharp upper constant in the main four type family. The all horizon calculation in Theorem D.2 gives a single family sharpness argument. Fix $0 < q < p < 1/2$ and take the ATE weight $a = 2(2D - 1)$ in (4). Its weighted initial error is $e_{1,1} = 4qs$, whereas Equation (35) gives

$$\frac{h_\infty(1)}{e_{1,1}} = \frac{h_\infty(-1)}{e_{1,1}} = \frac{1}{2(1 - s)} \longrightarrow 1 \qquad \text{as } s \uparrow 1/2.$$

Every selected $s < 1/2$ defines one fixed iid kernel with a correct semantic mode on every record. Any smaller upper constant is violated by one such fixed kernel and a sufficiently large finite horizon. Sharpness is not claimed uniformly over a prescribed positive mode gap. The first passage calculation establishing this ratio uses Theorem 6.1, independently of the upper bound being tested.

J PROOF OF THE ALL HORIZON INTERFACE FORMULA

Proof of Theorem D.2. Consider control. An initial A identifies its pure record and never revises. After an initial B, both pure record types are excluded. For a history with n_B reports B and n_C reports C, the remaining label one versus label zero posterior odds are

$$\left(\frac{1-s}{s}\right)^{n_C-n_B}.$$

The first label is zero, so a sticky decoder first revises precisely when $n_C - n_B$ first reaches $+1$. Starting at -1, this can occur only after $2j$ additional steps, with $j+1$ reports C and $j-1$ reports B. The number of paths that first reach the boundary at that step is

$$\frac{2}{2j}\binom{2j}{j+1} = \frac{1}{j+1}\binom{2j}{j} = C_j.$$

For completeness, reverse such a first passage path and subtract its level at the initial position. Its partial sums are strictly positive, with $j+1$ upward and $j-1$ downward steps. The strict ballot count is $(2/(2j))\binom{2j}{j+1}$, equivalently obtained by the reflection principle.

Conditional on the initial B, the truth one and truth zero weights are s and $1-s$. Thus the probability of first revision at $2j$ additional steps is

$$C_j\left[s(1-s)^{j+1}s^{j-1} + (1-s)s^{j+1}(1-s)^{j-1}\right] = C_j\{s(1-s)\}^j. \tag{54}$$

Summing gives $\Pr(H_T = 1 \mid R_1 = B) = f_T(s)$.

Conditional on the initial C, the ambiguous record weights are qs/p for truth zero and $q(1-s)/p$ for truth one. The remaining mass $(p-q)/p$ is pure C and never revises. Within the ambiguous component, exchange B with C and zero with one in the preceding calculation. Until a first B, a history consisting only of C reports has a strictly positive MAP label. After the first B, the pure type is excluded and the same posterior odds walk applies. Therefore

$$\Pr(H_T = 1 \mid R_1 = C) = \frac{q}{p} f_T(s).$$

The control masses of first label zero and first label one records that ever revise are consequently both $qf_T(s)$. Treatment reflects the entire law and its sticky decoder. For the ATE weight $a = 2(2D-1)$, each sign of $\Gamma = a(1-2G_1)$ has pooled revision mass $qf_T(s)$ and magnitude two. The exact support formula in Equation (14) therefore gives

$$h_T(1) = h_T(-1) = 2qf_T(s).$$

Initial selection followed by the first revision attains each endpoint, and independent initial mixing attains every intermediate effect.

Finally, with $x = s(1-s) < 1/4$, the Catalan generating function gives

$$\sum_{j\geq 1} C_j x^j = \frac{1-\sqrt{1-4x}}{2x} - 1 = \frac{s}{1-s}.$$

This proves (35), including monotone convergence of the finite horizon supports. The semantic correctness bound $1 - s > 1/2$ and the iid record kernel imply exact posterior and MAP consistency as proved after Corollary D.1. Hence both arm means under deterministic time release converge to their true means, giving $\Delta_T \to \delta$. At the stated numerical values, $2qs/(1-s) = 2/5$ and $\delta = 1/10$, so the limiting interval is $[-3/10, 1/2]$. □

K PROOFS FOR THE COMPLETE THREE CALL FRONTIER

Proof of Theorem E.1. We first identify every possible accuracy and label change. Conditional on either cross history BC or CB, the two ambiguous raw types have equal posterior weights, so $S_2 = 1/2$. The third report is B or C with equal probability; its posterior is respectively s or $1-s$.

The conditional accuracy improvement is therefore $d = 1/2 - s$. Sticky MAP keeps the first label when stopping at the cross history. A third query consequently changes the control label upward with probability $1/2$ after BC, and downward with probability $1/2$ after CB.

There are no other accuracy or label gains through time three. Pure A histories reveal label zero. After BB, the possible third posteriors at BBB, BBC are $s^3/(s^3 + (1-s)^3)$ and s, both below $1/2$. After CC, the CCB posterior is $1 - s > 1/2$; at CCC the latent one minus latent zero mass is $q\{(1-s)^3 - s^3\} + p - q > 0$. Thus the label remains positive. The second call itself has zero accuracy value after either initial report and causes no label change under the sticky convention. Reflection preserves accuracy and reverses every label direction.

These facts apply separately in each arm and conditional on the entire available decision history. Independent policy randomization cannot select the unobserved third answer or raw type within that history. Consequently, arbitrary policies that may depend on treatment satisfy exactly

$$\mathrm{Acc}_\pi = A_0 + d(u + v), \qquad \Delta_\pi = \delta - u + v.$$

For the second identity, an armwise executed cross probability produces label change mass equal to half that probability, while the sum of the two armwise probabilities is twice its pooled value. This accounts for the coefficients $-1, +1$ without requiring the gate probabilities in the two arms to coincide.

Let x_B, x_C be the pooled masses of second queries after the corresponding first reports. Their conditional cross probabilities are α, β, unchanged by the permitted policy randomization. Thus

$$u \le \alpha x_B, \qquad x_B \le q, \qquad v \le \beta x_C, \qquad x_C \le p.$$

Every third query counted in u, v also costs a call, whereas any other query can only increase cost. Therefore

$$0 \le u, v \le L, \qquad \mathbb{E}_\pi[\tau - 1] \ge x_B + x_C + u + v \ge k_B u + k_C v.$$

This establishes necessity even for policies that spend calls on A, equal answer paths, or cross paths that later terminate.

Conversely, fix any feasible (u, v). Start second calls with probabilities u/L after B and v/L after C, using fresh independent randomization, and always make the third call following a cross. Stop on all other paths. Its executed cross masses are $q\alpha(u/L) = u$ and $p\beta(v/L) = v$, and its cost is exactly $k_B u + k_C v$. Mirror the same gate in treatment. This proves sufficiency, minimum cost, and attainability of every displayed accuracy and effect pair by a common executable policy.

Accuracy maximizes $u + v$ over this rectangle subject to the single cost constraint. Since $k_B < k_C$, any solution with $u < L$ and $v > 0$ can improve its objective by shifting some of the more costly v allocation to u. Hence it fills the cheaper route first, giving (37). The resulting pair is unique, including at budget boundaries and saturation. At $b = k_B L$, equality in the cost lower bound forces $x_B = q$, $x_C = 0$, every reached BC history to continue, and no other calls. This proves policy uniqueness at that budget, up to null events.

At exact regret R, write $t = u + v = t_* - R/d$. The constraints $v = t - u \in [0, L]$, $u \in [0, L]$, and $k_B u + k_C(t-u) \le b$ are precisely $u_- \le u \le u_+$. Every $t \in [0, t_*]$ is feasible by scaling an optimum, so these bounds never conflict over the stated regret range. Since the effect is $\delta + t - 2u$, taking the two endpoints proves (39); every intermediate value is attained by the same construction. □

Proof of Theorem E.2. Write $K = k_B + k_C$ and $b_{\mathrm{full}} = KL$. Exact validity requires $u = v$. Its maximum attainable accuracy is therefore

$$A_0 + 2d \min\{L, b/K\}.$$

Subtracting this from $A_*(b)$ gives, for $0 \le b \le k_B L$,

$$\frac{d(k_C - k_B)b}{k_B K},$$

and for $k_B L \le b \le KL$,

$$\frac{d(k_C - k_B)(KL - b)}{k_C K},$$

while the regret is zero above KL. These expressions are exactly $\chi H(b)$.

The optimal effect is

$$\Delta_*(b) = \begin{cases} \delta - b/k_B, & 0 \le b \le k_B L, \\ \delta - L + (b - k_B L)/k_C, & k_B L \le b \le KL, \\ \delta, & b \ge KL. \end{cases}$$

It is negative precisely on (42) when $\delta < L$, and is never negative if $\delta \ge L$. Outside this interval, the unconstrained optimum already provides a zero regret sign repair.

Inside the interval, maximize $u + v$ subject additionally to $u - v \le \delta$. The sign constraint binds at an optimum: otherwise, a sufficiently small move toward the unique unconstrained accuracy optimal pair would preserve the strict sign inequality and improve accuracy. With $u = v + \delta$, the constraints give

$$v \le L - \delta, \qquad v \le \frac{b - k_B\delta}{K}.$$

Throughout the bad budget interval, the second bound lies in $(0, L - \delta)$ and is therefore binding. Thus the maximum sign valid total third call mass is

$$t_{\text{sign}}(b) = \delta + \frac{2(b - k_B\delta)}{K}.$$

Subtracting dt_{sign} from dt_* gives

$$R_{\text{sign}}(b) = \begin{cases} \dfrac{d(k_C - k_B)}{k_B K}(b - k_B\delta), & k_B\delta < b \le k_B L, \\ \dfrac{d(k_C - k_B)}{k_C K}(KL - k_C\delta - b), & k_B L \le b < KL - k_C\delta, \\ 0, & \text{otherwise.} \end{cases}$$

This equals $\chi[H(b) - \delta]_+$, including all endpoints. The maximum of H is L at $b = k_B L$, proving (43). Since these regrets minimize over the entire feasible policy class, no policy below the corresponding cost can accomplish the stated repair, while the constructed policies attain both boundaries exactly.

It remains to verify the claimed policy restriction. First report uncertainty is 0 on A, s on B, and $qs/p < s$ on C. The prediction optimal pair always satisfies $u_* \ge v_*$. A validity repair has $u = v$, while a binding sign repair has $u = v + \delta > v$. Their second call probabilities $0, u/L, v/L$ are therefore nondecreasing in posterior uncertainty. At time two, the only continued histories have posterior $1/2$; all other supported posteriors remain strictly on their original sides. Continuing exactly the cross histories is again nondecreasing in uncertainty. These gates are common to both arms because reflection preserves uncertainty and the sticky tie rule preserves complementary labels. □

L SHARP TERMINAL COMPLETION AND THE INTEGRATED CRITERION

Definition L.1 (Terminal band completion class)**.** Let P_{obs} be the joint law of treatment, stopped reports, recorded reference scores, and released labels under the gate. For $\xi \ge 0$, let $\mathcal{C}_\xi(P_{\text{obs}})$ consist of all laws of binary potential truths and potential stopped records that induce P_{obs}, retain independent randomized treatment, obey $(X, H_\tau) = (X(D), H_\tau(D))$, and satisfy, for each arm d and every supported stopped history h,

$$|\mathbb{E}_P[X(d) \mid H_\tau(d) = h] - s_{\text{ref}}(d, h)| \le \xi. \tag{55}$$

Here s_{ref} is the recorded reference score. The identified region is

$$\Theta_{I,\Delta}(\xi) = \{\mathbb{E}_P[X(1) - X(0)] : P \in \mathcal{C}_\xi(P_{\text{obs}})\}.$$

Proof of Proposition 4.1. Under the three call policy at $b_* = k_B L$, the control terminal history law follows directly from (4) and the gate in Theorem E.1:

released history	mass	reference posterior
A	$1 - p - q$	0
C	p	$1 - qs/p$
BB	$q(1 - \alpha)$	$s^2/[s^2 + (1 - s)^2]$
BCB	$L/2$	s
BCC	$L/2$	$1 - s$

(56)

The masses sum to one and their posterior weighted sum is p. Every nonzero score and its distance from one is at least ξ_0: $qs/p < s$, and the BB posterior is below both s and $1/2$. Hence these bands have width 2ξ without clipping, whereas the score zero band is $[0, \xi]$. The nonzero histories have total mass $p + q$. Summing their lower and upper latent one masses gives

$$p_0 \in [p - (p+q)\xi,\ p + \xi].$$

Reflection gives

$$p_1 \in [1 - p - \xi,\ 1 - p + (p+q)\xi].$$

For sharpness, assign any terminal latent one fraction independently within each band. Multiplying by the observed terminal mass gives two latent class masses at every leaf. Sum these masses backward through the deterministic measurement tree. At a queried node, child mass divided by its latent parent mass defines the corresponding conditional response probability, with arbitrary choices at zero mass parents. At a released node, future unobserved responses may be defined arbitrarily. These kernels reproduce all observed terminal masses and respect the known gate.

Draw potential latent trees for the two arms before independently assigning treatment. Thus both armwise intervals, every pair of points in them, and the entire difference interval are attained by coherent causal measurement completions. These completions need not preserve the reference iid kernel or prior.

Subtracting endpoints proves (11). Its upper endpoint is at least δ, so the proxy value $\delta - L$ is excluded exactly when

$$\delta - L < \delta - 2\xi,$$

equivalently $L > 2\xi$. The numerical substitution gives $[.06, .13]$ and $-.044$, as stated. All these completions have well defined binary outcome means, while the reference latent law supplies a nonempty member of the class. □

If the appropriate latent conditional law is known, posterior integration preserves any integrable binary criterion under a bounded stopping policy:

$$\mathbb{E}[S_\tau \ell(W, 1; \theta) + (1 - S_\tau)\ell(W, 0; \theta)] = \mathbb{E}\ell(W, X; \theta). \tag{57}$$

This tower property identity conditions on the full stopped information, including W. If X is a regressor and W contains an outcome Y, integrating squared loss gives

$$(Y - \theta S_\tau)^2 + \theta^2 S_\tau(1 - S_\tau).$$

Simply substituting S_τ and omitting the second term loses the variance correction. This is an oracle identity and does not provide a guarantee for an estimated posterior.

For (57), the stopped sigma field contains W, and bounded stopping preserves

$$S_\tau = \mathbb{E}[X \mid \mathcal{F}_\tau].$$

The conditional expectation of the binary criterion is therefore exactly

$$S_\tau \ell(W, 1; \theta) + (1 - S_\tau)\ell(W, 0; \theta).$$

Taking expectations proves the identity. In the squared loss example, expansion gives

$$Y^2 - 2\theta Y S_\tau + \theta^2 S_\tau,$$

which equals

$$(Y - \theta S_\tau)^2 + \theta^2 S_\tau(1 - S_\tau).$$

The covariate or outcome Y is part of W in this example; a posterior that omits available Y need not satisfy this conditional identity.

M PROOFS FOR POSTERIOR FREE REPAIR

M.1 FRESH CALLS AFTER A STOPPED PRODUCTION HISTORY

Conditional on a fixed raw record Z, the entire report sequence is iid from its response vector, and the policy coins are independent of that sequence. For every deterministic t, the future reports are independent of the first t reports and the production coins, conditional on Z. A production stopping event $\{\tau = t\}$ is measurable with respect to those past variables. Summing the deterministic time factorization over t shows that, conditional on Z and the stopped history, any fixed number of subsequent reports has the product law of the same response vector. This is the discrete stopped sequence analogue of the usual fresh increments property. Almost sure finiteness of τ is sufficient for this argument; $\mathbb{E}\tau < \infty$ additionally ensures finite expected total computation after the repair calls are added.

Independent policy randomization does not reveal a raw type beyond the retained report history. Conditional on any fixed finite report sequence, the probability of taking its specified production actions is the same across types that can emit that sequence. Since the mixing weights are positive, every such type remains possible at each stopped report history of positive probability. With continuous policy coins, the analogous statement holds almost surely by the same likelihood factorization. This observation justifies the remaining type classes used in the minimum depth claims below.

Proof of Theorem 5.1. Work in control orientation and condition on a fixed supported production history. An occurrence of A identifies the pure negative type. Such a supported history contains only A reports, so the zero call output is exactly its true label.

If the history contains B, neither pure type is possible. For the negative ambiguous type, the next semantic positive probability is s; for the positive ambiguous type, it is $1 - s$. The fresh call property above therefore gives

$$\mathbb{E}[\widetilde{V} \mid Z, H_\tau] = \frac{s - s}{1 - 2s} = 0 \qquad \text{or} \qquad \frac{(1 - s) - s}{1 - 2s} = 1,$$

as required. This covers prefixes consisting only of B reports as well as mixed B/C prefixes of every finite length; their posterior weights need not be known.

For a history consisting only of C reports, the pure positive type and both ambiguous types remain possible. Write $\beta_B = Q_B(Z)$. The three possibilities, together with their true labels, are

$$(\beta_B, X) = (0, 1), \qquad (s, 1), \qquad (1 - s, 0).$$

The two fresh J_j are independent Bernoulli variables with parameter β_B, even after conditioning on the production history. Thus

$$\mathbb{E}[\widetilde{V} \mid Z, H_\tau] = 1 - \frac{\beta_B^2 - s\beta_B}{(1 - s)(1 - 2s)} = \begin{cases} 1, & \beta_B = 0 \text{ or } s, \\ 0, & \beta_B = 1 - s. \end{cases} \tag{58}$$

This proves conditional unbiasedness in all three cases.

The one call values are $-s/(1 - 2s)$ and $(1 - s)/(1 - 2s)$. For the two call formula, the values at $K = J_1 + J_2 = 0, 1, 2$ are respectively

$$1, \qquad 1 + \frac{s}{2(1 - s)(1 - 2s)}, \qquad -\frac{2s}{1 - 2s}. \tag{59}$$

For $0 < s < 1/2$, these values and their complements lie in the interval displayed in (12). Treatment reflection complements both the scientific label and the oriented estimator, establishing the same claims in that arm. Fixed s gives a uniform finite bound, so every conditional second moment exists. The prescribed call counts are deterministic given the production history and give the stated expected additional cost.

We next prove minimum depth. Zero further calls are possible after an A history, and no negative number of calls is permitted. After a history containing B, the two compatible types have different truths. A zero call correction, even with fresh auxiliary randomization, has the same conditional distribution under both types because the production history is fixed. It therefore cannot have

expectation zero under one type and one under the other. Thus at least one possible additional call is necessary, and $\widetilde{V} = \frac{Y-s}{1-2s}$ attains this depth.

After a history consisting only of C reports, suppose a correction uses at most one fresh call. Pad it with an unused fresh call if it stops without making one. Integrating over its auxiliary randomization gives two possible conditional means v_0, v_1, corresponding to fresh $J = 0, 1$. Its expectation as a function of the remaining response parameter is therefore the affine function $v_0(1-\beta_B) + v_1\beta_B$. It must equal one at both $\beta_B = 0$ and $\beta_B = s$, which forces $v_0 = v_1 = 1$. Its expectation at $\beta_B = 1-s$ is then also one, contradicting the required value zero. Hence the worst case depth is at least two, and the displayed construction attains it. These lower bounds concern the stated remaining catalog: if additional information excluded a type, or its mixing weight were known to be zero, the minimum depth could change.

Finally, consider variance within the minimum depth cap. After a prefix containing B, pad every candidate to one fresh call and average over its auxiliary randomization conditional on the fresh outcome. Conditional expectation cannot increase variance under either compatible type. The two unbiasedness equations at the distinct probabilities s and $1-s$ uniquely determine the conditional output values, giving $\widetilde{V} = \frac{Y-s}{1-2s}$.

After a prefix consisting only of C reports, pad every candidate to two fresh calls, including when its second call decision was adaptive. Given $K = J_1 + J_2$, the ordered pair is uniform over pairs with that count for every compatible type whenever the count has positive probability. This conditional distribution, as well as the law of the auxiliary coins, does not depend on β_B. Averaging the candidate conditional on K therefore produces the same statistic under every type and weakly decreases every conditional variance. Write its three values as w_0, w_1, w_2. Their mean is the polynomial of degree at most two

$$w_0(1-\beta_B)^2 + 2w_1\beta_B(1-\beta_B) + w_2\beta_B^2.$$

Its values at the three distinct points $0, s, 1-s$ must be $1, 1, 0$. Polynomial interpolation is unique, and (58) already gives that polynomial. Its Bernstein coefficients are exactly (59). Hence the conditional mean of every unbiased candidate given K is the proposed estimator, proving the variance comparison by conditional Jensen. Candidates with infinite variance cannot improve it. The pure A correction has zero variance. Reflection preserves variances and minimum depths in treatment. □

Proof of Proposition F.1. Fix the production prefix and its two compatible noisy types. Every finite B/C continuation has positive probability under both types, and auxiliary coins have the same law. Their finite transcript measures are therefore mutually absolutely continuous. If a correction stops after a finite number N of fresh calls almost surely under each type, the same remains true on the stopped sigma field: decompose any null event into its intersections with $\{N = n\}$ and apply the finite transcript statement for each n.

A $[0, 1]$ valued output with expectation zero under the negative type must equal zero almost surely under that type. Absolute continuity then makes it zero almost surely under the positive type as well, contradicting the required expectation one. This argument conditions on a fixed prefix and does not compare posterior expectations under different population priors. □

Corollary M.1 (Preservation of the binary scientific criterion)**.** *Suppose the scientific covariates W are part of the raw state and are observed before repair. For every criterion satisfying $\mathbb{E}|\ell(W, 0; \theta)| + \mathbb{E}|\ell(W, 1; \theta)| < \infty$, define*

$$\ell^{\mathrm{corr}}(W, V; \theta) = V\ell(W, 1; \theta) + (1-V)\ell(W, 0; \theta). \tag{60}$$

For every fixed θ and every production policy above,

$$\mathbb{E}\ell^{\mathrm{corr}}(W, V; \theta) = \mathbb{E}\ell(W, X; \theta). \tag{61}$$

The same conclusion holds for integrable binary score functions.

Proof of Corollary M.1. The raw state fixes both X and the observed W. Conditional unbiasedness yields, for each fixed θ,

$$\mathbb{E}[\ell^{\mathrm{corr}}(W, V; \theta) \mid Z, H_\tau] = X\ell(W, 1; \theta) + (1-X)\ell(W, 0; \theta) = \ell(W, X; \theta).$$

The uniform finite bound on V at fixed s, together with integrability of the two endpoint criteria, permits iterated expectation and proves (61). Replacing the criterion by each coordinate of an integrable binary score gives the score assertion. The same reasoning does not justify replacing the linear extension by $\ell(W, V; \theta)$: in general, expectation does not commute with a nonlinear function of V. □

N NUMERICAL CERTIFICATES, STATISTICAL COST, AND SENSITIVITY

All calculations in this section use the declared finite response catalog. They evaluate the preceding results rather than experiments with a deployed language model. The supplementary code retains exact rational outputs, including unsuccessful sensitivity checks. Independent implementations start from the raw type probabilities rather than only substituting parameters into the final formulas.

N.1 A COMMON CAP AND A POSITIVE REGRET ALLOWANCE

For $(p, q, s) = (.49, .489, .40)$, select the seven prices

$$.04, \quad \rho_2^-, \quad \rho_2^+, \quad \rho_3^-, \quad \rho_3^+, \quad \rho_4^-, \quad \rho_4^+, \quad \rho_m^- = (b_m + c_{m-1})/2, \quad \rho_m^+ = (c_{m-1} + c_m)/2.$$

Here b_m, c_m are the exact prices in (22) and (17). The negative prices are the centers of their respective negative effect intervals. The positive prices lie in their asserted positive effect intervals for this population; this last statement is an evaluation of the specified parameters, not a claim about arbitrary populations. The negative interval widths for $m = 2, 3, 4$ are approximately 2.3286×10^{-5}, 1.6839×10^{-5}, and 1.6777×10^{-5}. Their centers are approximately .0324207893, .0184059029, and .0120890634.

Table 3: Certificate for the seven selected prices, in decreasing order. Every tail bound is a terminating decimal rational upper bound on the exact probability. C^j denotes an unrevised prefix; k denotes a paired count. The gaps include both continuation and release decisions.

Regime (m, h)	Minimum gap γ_j	Minimizing decision	$\Pr(\tau_j^* > 168)$ upper bound
Immediate	$7/625$	Release at $k = -1$	0
$(2, 1)$	$3/90650$	Release at C^1	8.196×10^{-28}
$(2, 2)$	$3/185$	Release at $\lvert k\rvert = 2$	1.640×10^{-27}
$(3, 2)$	$45/1247681$	Release at C^2	1.175×10^{-12}
$(3, 3)$	$18/1955$	Release at $\lvert k\rvert = 3$	1.410×10^{-12}
$(4, 3)$	$135/3078202$	Release at C^3	6.307×10^{-8}
$(4, 4)$	$108/17855$	Release at $\lvert k\rvert = 4$	6.714×10^{-8}

The minimum successive reference accuracy increase is $d_A = 13203/1515625 \simeq .00871126$. Write π_j^* for the infinite horizon optimum at selected price j and $\Delta_j^* := \Delta_{\pi_j^*}$ for its released proxy effect; the true effect is the separate constant $\Delta^* = \delta = .02$. For every row, exact arithmetic verifies, with $R = 10^{-9}$,

$$R + \Pr(\tau_j^* > 168) < \gamma_j \min\{|\Delta_j^*|/4, d_A/3\}. \tag{62}$$

Proposition H.1 therefore places every capped policy within objective regret R of its capped optimum within $|\Delta_j^*|/2$ of the reference effect and within $d_A/3$ of its reference accuracy. All seven signs and all six strict accuracy increases follow. This gives a *sufficient* common cap of 168 total calls, including the first call. It does not identify the minimum cap needed for reversals. With $R = 10^{-9}$ included, the smallest certified absolute effect margin is approximately .00307833 and the smallest successive accuracy margin is approximately .00723928. The small regret allowance also makes the limited quantitative size of this fixed model guarantee explicit.

For reproducibility, the gap calculation enumerates the finitely many states reached by each reference policy and evaluates both Bellman actions using the globally optimal value function. In particular, a hypothetical extra call after a release state is followed by optimal actions rather than by an artificially forced continuation of the already ended policy. The tail calculation advances the live state distribution under that reference policy. One implementation uses observable posterior transitions; another carries the four raw type masses and their iid transition probabilities. Their live state distributions agree at every cap up to 168. The short bounds in Table 3 are checked against the retained exact fractions.

N.2 BUDGET OPTIMA AT FIVE AND TEN CALLS

At (10), let $A_T^*(b)$ be the maximum accuracy among all randomized treatment aware policies with $\tau \leq T$ and $\mathbb{E}(\tau - 1) \leq b$. Write $[\underline{\Delta}_T^*(b), \overline{\Delta}_T^*(b)]$ for the complete effect interval among those accuracy optima. Exact rational calculations give

$$\{b : \overline{\Delta}_T^*(b) < 0\} = \begin{cases} (37/120, 1251/2000), & T = 3, \\ (1147/2280, 45187/50000), & T = 5, \\ (19329799/32925000, 34418803/31250000), & T = 10. \end{cases} \tag{63}$$

At both endpoints in every row, $\overline{\Delta}_T^* = 0$. These statements concern the specified finite problems and are supported by the exact certificates below; they do not extrapolate to other horizons.

Table 4: Exact finite budget optimization; decimals are rounded from rational outputs. Effect extrema range over every accuracy optimum.

T	b	$A_T^*(b)$	$\underline{\Delta}_T^*(b)$	$\overline{\Delta}_T^*(b)$
3	.350	.771351	-.013514	-.013514
3	.444	.774400	-.044000	-.044000
3	.550	.776970	-.018303	-.018303
3	.700	.780606	.018061	.018061
3	1.00	.787879	.090788	.090788
5	.350	.771351	-.013514	.030427
5	.444	.774400	-.044000	.011742
5	.550	.777838	-.078378	-.009329
5	.700	.782444	-.036350	-.036350
5	1.00	.790366	.017174	.017174
10	.350	.771351	-.013514	.040399
10	.444	.774400	-.044000	.024391
10	.550	.777838	-.078378	.006341
10	.700	.782703	-.068177	-.019823
10	1.00	.791461	-.015234	-.015234

Every policy and every optimal tie. The implementation in `verify_budget_horizons.py` first enumerates all supported ordered histories and their raw type masses; it does not compress histories with identical counts but different retained labels. For each arm and each nonterminal history h, an occupation variable is the population mass that continues at h. It lies between zero and the mass reaching h; child reach equals parent continuation mass times the conditional report probability. Every feasible vector is implemented by continuing with probability equal to continuation mass divided by reach mass. Cost sums these variables, while accuracy and effect telescope as linear sums of one call increments. Thus the two arm occupation linear program represents exactly the admissible randomized policies.

A separate rational Bellman recursion maximizes $A - cC$. Recursive support queries enumerate every edge of the upper accuracy cost frontier. At its supporting price c, the Bellman value $v(c)$ gives the dual bound $A \leq v(c) + cb$, attained by the stored policies. Retain all Bellman optimal actions, including ties. Stopping contributes a point $(C, \mathbb{E}G)$; continuation sums the child polygons and adds its call cost; a tied action takes the convex hull of both choices. The resulting polygon contains every cost and label pair on that optimal face. For $c > 0$, its vertical slice at $C = b$ gives the complete effect interval. Reflection and convexity ensure that allowing separate treatment gates does not enlarge the pooled polygon. At $c = 0$, use the slice $C \leq b$; a maximum accuracy policy with effect .10 rules out another all negative interval on the final flat frontier.

Each polygon edge is additionally certified by scalar Bellman optimization of label mass minus its supporting slope times cost on the retained optimal action graph. The stored primary and secondary dual coefficients are rational. All 15 rows also agree with a separate explicit two arm floating point occupation program; its maximum feasibility residual is below 1.7×10^{-15}. The interval endpoints come from the exact faces rather than from a grid or an accuracy equality tolerance.

Perturbations that preserve and remove reversal. With masses $(1-p-q_0, q_0, q_1, p-q_1)$ and noisy error rates s_0, s_1, the truth remains .10 at $p = .45$, but the initial anchor need not hold. At $T = 10, b = .7$, the maximal optimal effects for $(q_0, q_1) = (.301, .299), (.299, .301), (.31, .29)$ and $s_0 = s_1 = .4$ are respectively $-.021243, -.066687, +.263913$. The third case has no negative component on the increasing accuracy frontier at any tested cap $3, 5, 10$. An anchored asymmetric case instead sets $q_0 = .301, q_1 = .299, s_0 = .12/q_0, s_1 = .12/q_1$; equal first call error masses preserve the anchor, and the $T = 10, b = .7$ maximal effect is $-.023848$. These finite cases retain both successes and failures.

Two additional checks keep the base true population and use a frozen decoder with either $(q_0, q_1) = (.301, .299)$ or $(.299, .301)$, leaving its error rates .4. The tested accuracy frontiers and effect extrema match the base case. Stopping is still optimized under the true evaluation law for each fixed decoder; this is neither a fitted decoder experiment nor a theorem about posterior estimation. Exact outputs for 11 alternative specifications and 165 budget cases are retained, with separate occupation program discrepancies below 4.4×10^{-14}.

N.3 Finite cap perturbations: preserved and failed cases

For a direct check outside the symmetric construction, keep all raw type masses and the noisy positive type's $Q_C = .6$ fixed. Set the noisy negative type's Q_C to .39999, .4, or .40001, retaining $Q_B = 1 - Q_C$ and reflecting treatment. The true effect remains .02. The initial proxy effect is now .02000978, .02, or .01999022, respectively, so the perturbed laws no longer satisfy the exact initial anchor of the symmetric construction.

At a cap of 65 total calls, exact rational Bellman optimization and a separate forward occupation calculation give Table 5. The posterior decoder and gate are reoptimized separately under each declared true kernel. There are no reached nonterminal stop or continue ties in these 21 problems. At each of the three kernels, accuracy strictly increases down the table. For example, at $Q_C = .40001$ the seven accuracies are approximately .60879511, .65393119, .69906824, .75627833, .77644121, .82241332, and .83822542.

Table 5: Optimal effects at cap 65 for the interval center price list. These are three specified kernels, not a uniform perturbation theorem.

Price	$Q_C = .39999$	.40000	.40001
.04	.02000978	.02000000	.01999022
ρ_2^-	$-.21470545$	$-.21472000$	$-.21473455$
ρ_2^+	.02001736	.02000000	.01998264
ρ_3^-	$-.05418255$	$-.05412211$	$-.05406164$
ρ_3^+	.02002156	.02000000	.01997844
ρ_4^-	$-.00597134$	$-.00599442$	$-.00326187$
ρ_4^+	.02002254	.02000000	.01997746

Price selection matters. An earlier illustration used the negative prices $c_{m-1} - 10^{-6}$, close to the upper interval boundaries. At those prices and cap 65, the unperturbed sampler has all six sign changes, but replacing the noisy negative $Q_C = .4$ by .40001 makes all seven effects positive. Choosing interval centers avoids that failure in the three cases just checked. Neither calculation establishes a uniform kernel neighborhood, robustness to estimated posteriors, or whether cycles move to other prices under a perturbed model. The cap 168 regret certificate above applies to the unperturbed kernel.

The finite cap implementation uses unnormalized raw type history masses $w_z(h) = \pi_z \prod_{r \in h} Q_r(z)$. If $U_t(h)$ is the unnormalized value at a supported history of length t, then

$$U_T(h) = \max_{x \in \{0,1\}} \sum_{z : X(z) = x} w_z(h),$$

$$U_t(h) = \max \left\{ \max_x \sum_{z : X(z) = x} w_z(h),\ \sum_r U_{t+1}(hr) - c \sum_z w_z(h) \right\}.$$

This recursion optimizes over all history dependent stopping rules; linearity makes randomization unnecessary except at action ties. Exported labels use the specified sticky MAP convention, tracked separately from the value. A forward pass verifies released mass, accuracy, effect, and expected calls. The supplementary checks also reconstruct the three call frontier by history occupation linear programming and verify the general first revision support formula on small binary history trees, including a hidden weight negative control.

N.4 What the known catalog repair costs statistically

Consider $(p, q, s) = (.45, .30, .40)$ and the production optimum at $b_* = .444$. It stops at the first A or C, stops at BB, and makes a third call after BC. At a prefix containing B, the corrected output is -2 or 3, with conditional variance 6 under either noisy type. At a prefix consisting only of C, the outputs for zero, one, or two fresh B reports are $1, 8/3, -4$. The corresponding conditional variances are $28/3$ for the noisy negative type, $16/3$ for the noisy positive type, and zero for pure C. More generally, the one call variance is $s(1-s)/(1-2s)^2$, which diverges as $s \uparrow 1/2$.

Averaging over production histories gives the following raw type moments. The variance column conditions on the type, rather than only on the terminal history. Since $\mathbb{E}[V \mid \text{type}, H_\tau] = X$, total

Table 6: Repair and total sampler cost in the three call example. Treatment has reflected outputs and the same variances and costs.

Control type	Mass	$\mathrm{Var}(V \mid \text{type})$	Production calls	Fresh calls
Pure A	$1/4$	0	1	0
Noisy negative	$3/10$	$22/3$	$46/25$	$7/5$
Noisy positive	$3/10$	$28/5$	$41/25$	$8/5$
Pure C	$3/20$	0	1	2

variance gives

$$\mathrm{Var}(V \mid D = d) = \frac{99}{400} + \frac{3}{10}\frac{22}{3} + \frac{3}{10}\frac{28}{5} = \frac{1651}{400}, \qquad d = 0, 1.$$

The expected production and fresh call counts are 1.444 and 1.2, respectively, giving 2.644 total calls per record. The maximum total count for this particular procedure is four. If only the corrected scientific object is needed, releasing after the first report and then applying the same repair gives the same typewise distribution of V at 2.2 expected total calls. Indeed, the additional production calls never change which of the three repair classes contains the prefix. This cheaper route does not also provide the production optimum's improved hard label accuracy. No expected call optimality claim follows.

For a fixed production policy and fixed known $s < 1/2$, draw n_d iid superpopulation records in each arm, independently across arms, with independent production seeds, repair seeds, and policy coins across records. Let $\bar{V}_d$ and $\widehat{\sigma_d^2}$ be the arm sample mean and usual sample variance. The estimator and its standard error are

$$\widehat{\Delta}_V = \bar{V}_1 - \bar{V}_0, \qquad \widehat{\mathrm{se}}^{\,2} = \widehat{\sigma_1^2}/n_1 + \widehat{\sigma_0^2}/n_0. \tag{64}$$

The estimator is unbiased. With $n_0 = n_1 = n/2$, its exact variance in the example is $16.51/n$. Observing true labels would give $.99/n$, a ratio $1651/99 \simeq 16.68$; this is an information benchmark, with no cost assigned to obtaining those true labels. The stopped hard label contrast instead has mean $-.044$, variance $.998064/n$, and bias $-.144$ for the true effect. Its lower variance therefore does not make it a valid measurement of that effect.

For fixed positive arm variances, $n_d \to \infty$ with $n_1/(n_0 + n_1)$ bounded away from zero and one, the boundedness of V gives the ordinary independent sample central limit theorem and consistency of the sample variances. Consequently, $\widehat{\Delta}_V \pm z_{1-\alpha/2}\widehat{\mathrm{se}}$ has asymptotic coverage $1 - \alpha$ under the maintained known catalog. Here $z_{1-\alpha/2}$ is the standard normal quantile. This is a direct textbook inference calculation, not an efficiency result or a guarantee with an estimated catalog. A shared, data adaptive gate or repair parameter requires its own dependence analysis. The population identity for an arbitrary signed nonlinear criterion does not by itself guarantee existence or uniform convergence of an empirical minimizer.

N.5 MISSPECIFYING THE REPAIR PARAMETER

Keep the true population, kernel, production gate, and decoder of the preceding example fixed. Replace only s in the repair weights by $u \in (0, 1/2)$. For a noisy type with $\beta_B = Q_B$, the conditional means at a prefix containing B and a prefix consisting only of C are, respectively,

$$m_B(\beta_B; u) = \frac{1 - \beta_B - u}{1 - 2u}, \qquad m_C(\beta_B; u) = 1 - \frac{\beta_B(\beta_B - u)}{(1 - u)(1 - 2u)}.$$

The two formulas follow by taking expectations of one Bernoulli report and of two independent Bernoulli reports. This production policy assigns the noisy type to the two repair classes with probabilities β_B and $1 - \beta_B$, so its mean is $\beta_B m_B(\beta_B; u) + (1 - \beta_B)m_C(\beta_B; u)$. Substituting $\beta_B = 1 - s$, s and the equal noisy masses q gives

$$\Delta_V(u) - \delta = \frac{2q(s - u)(s + u - 1)}{(1 - u)(1 - 2u)}. \tag{65}$$

This identity concerns the specified production policy and equal mass population. The aggregate cancellation is not a bound for arbitrary mixtures or arbitrary production policies. At $u = s = .4$, the derivative of the bias with respect to u is $2q/(1 - s) = 1$.

Table 7: Misspecified repair weights in the fixed three call example. The variance is about each displayed population mean; it omits uncertainty in an estimated u. Calls are unchanged.

Supplied u	Repaired effect	Effect bias	$n \operatorname{Var}(\widehat{\Delta}_V)$
.38	.082258	−.017742	11.376177
.39	.090611	−.009389	13.584568
.40	.100000	.000000	16.510000
.41	.110734	.010734	20.494005
.42	.123276	.023276	26.106615
.45	.181818	.081818	68.582975

These calculations quantify probability error while preserving the catalog's support and orientation. They do not address additional response types, incorrect semantic orientation, or confidence interval coverage when u is estimated. Even without a truth prior, repair still requires validated information linking response probabilities to the latent label.

O ADVERSARIAL BOUNDARIES OF THE CONSTRUCTION AND REPAIR

How large can the true effect be? The three call optimum can have negative effect exactly when $\delta < 2qs(1 - s)$. Since $q < p = (1 - \delta)/2$, some admissible q exists exactly when

$$0 < \delta < \frac{s(1 - s)}{1 + s(1 - s)} < \frac{1}{5}.$$

Sufficiency follows by choosing q strictly between $\delta/[2s(1 - s)]$ and p. At $s = 2/5$, the sharp supremum is $6/31$. For the displayed adjacent cutoff cycle with $h = m - 1$, simplify its negative displacement to qK_m, where

$$K_m = a_{m-1}\left(1 - \frac{z_{m-1}}{z_m}\right) = \frac{2(1 - 2s)[s(1 - s)]^{m-1}}{(1 - s)^m - s^m}.$$

The denominator divided by $1 - 2s$ is $\sum_{j=0}^{m-1}(1 - s)^{m-1-j}s^j$, at least $m[s(1 - s)]^{(m-1)/2}$ by AM–GM. Hence $K_m \le 2^{2-m}/m$, and negativity requires $\delta < K_m/(2 + K_m) \le 1/[m2^{m-1} + 1]$. Taking $m = N + 1$ proves the necessary bound stated after Theorem C.2. It concerns these adjacent cutoff cycles, not every conceivable pattern in a different measurement family.

What changes with the tie rule? For any fixed Bayes tie convention, every useful third call has the same strict terminal posterior, gain d, and minimum route cost. Before useful saturation, an accuracy optimum uses the minimum cost allocation: stopping at a reached cross history tie would waste a second call and cannot attain that allocation. At saturation, every cross history receives its third call. Thus every three call accuracy optimum releases at strict posteriors and has the same effect under these conventions. The strict price policies also release at strict posteriors, and their Bellman values depend only on Bayes accuracy, so their effects are unchanged. The finite cap comparison applies with each convention fixed.

The full nonoptimal frontier does depend on sticky ties. At (10), query once more after initial B in each oriented arm, then stop. Cost is .3 and accuracy is .76 under every Bayes tie convention. Sticky ties give effect .10; independent fair tie coins give $-.044$. Therefore the complete sticky frontier cannot be extended unchanged to arbitrary tie conventions. This negative control is also checked directly from primitive histories in the supplementary code.

O.1 Fixed depth robustness and support contamination

Apply the same production policy, bounded by T, and the nominal known catalog correction to a perturbed kernel on the same alphabet. Keep truths, type masses, initial information, and orientation fixed. Suppose each type's report row is within total variation $\epsilon \in [0, 1]$ of its nominal row. On any previously unsupported history, define the nominal repair by the same branches: A present first, otherwise B present, otherwise all C.

Conditional on a raw type, couple each nominal and actual report with disagreement probability at most ϵ, independently over calls, and use the same policy coins. The two outputs agree if the first $T+2$ reports agree. Control oriented outputs lie between $-2s/(1-2s)$ and $(1-s)/(1-2s)$, whose span is $(1+s)/(1-2s)$. Since the nominal output has conditional mean X,

$$|\mathbb{E}[V \mid Z] - X| \le \frac{1+s}{1-2s}\left\{1-(1-\epsilon)^{T+2}\right\} \le \frac{1+s}{1-2s}(T+2)\epsilon. \tag{66}$$

The treatment contrast bias is at most twice this bound. The result applies to the unchanged policy and nominal correction; it does not bound a reoptimized effect or provide calibration conditional on a selected history. It quantifies why a finite production cap matters.

To see why depth dependence cannot be discarded, change only the pure positive row from $(0, 0, 1)$ to $(0, \epsilon, 1-\epsilon)$, with $0 < \epsilon < s$, and reflect treatment. Stop production at the first B or at T. If A occurs, stop immediately, which affects only the unchanged pure negative type. For the perturbed pure positive type, let $a'_T = 1-(1-\epsilon)^T$. The nominal one call and two call corrections have conditional biases $(s-\epsilon)/(1-2s)$ and $\epsilon(s-\epsilon)/[(1-s)(1-2s)]$, respectively. Its averaged bias is

$$B(\epsilon, T) = a'_T \frac{s-\epsilon}{1-2s} + (1-a'_T)\frac{\epsilon(s-\epsilon)}{(1-s)(1-2s)}, \qquad \Delta_V - \delta = -2(p-q)B(\epsilon, T). \tag{67}$$

The other types retain exact unbiasedness after this production stop. With $T = \lceil 1/\epsilon \rceil$ and $\epsilon \downarrow 0$,

$$B(\epsilon, T) \to (1-\exp(-1))\frac{s}{1-2s} > 0,$$

even as every report row converges to its nominal value. At (10), $\epsilon = .001$ and $T = 1000$ give $\Delta_V = -.2788010645$ while $\delta = .10$. Every record still has semantic correctness at least .60. This is a failure of uniform approximate catalog robustness over production depths and does not contradict Theorem 5.1 under its exact catalog.

P The measured response kernel

This appendix documents Section 7: how the kernel was measured, how treatment was constructed on it, and the full grid of results. No result in the main theory depends on it.

P.1 Annotation protocol and its validation

The pool is 3000 sentences drawn without replacement from the sentence polarity corpus of Pang & Lee (2005), 1500 carrying each polarity label, truncated to 300 characters. Each was sent once to `gpt-4o-mini` with the prompt

```
Classify the sentiment of the following movie-review
sentence.  Reply with exactly one word:  positive, negative,
or mixed.
Sentence:  "{text}"
Answer:
```

at temperature one, with a two token limit and the twenty most probable first tokens returned with their log probabilities. Mass was collected onto the three answer words after stripping case, leading spaces and punctuation, then renormalized, giving one row $Q(z)$ of a 3000×3 catalog with *negative* and *mixed* mapped to semantic label zero. Total cost was about four cents.

Simulations use independent draws from the stored $Q(z)$ by construction. For actual endpoint calls, (1) is an assumption. A marginal frequency check queried 100 records ten more times at temperature one; no answer fell outside the three words, and the mean total variation distance between empirical frequencies and $Q(z)$ was .0154. A sampling temperature θ rescales the log probabilities before renormalization, so the family $Q_r^{(\theta)}(z) \propto Q_r(z)^{1/\theta}$ comes from the same annotations. The transform acts on category probabilities that already aggregate token variants from a truncated token list, so it need not coincide with the endpoint's own distribution at that temperature, which tempers every token. Two limits on that device are worth stating. The chat interface accepts $\theta \in [0, 2]$, so larger values describe no request that can be made; and tempering acts on the whole vocabulary, whereas $Q^{(\theta)}$ here is the conditional law given an answer in the three words, an approximation that degrades as θ grows. Every number in Section 7 and in Table 21 therefore uses $\theta \in \{1, 2\}$.

P.2 Design and decoder

Treatment is a design. Arm d places mass π_d on the positive records and $1 - \pi_d$ on the negative ones, so $\Delta^* = \pi_1 - \pi_0$ holds exactly. Within a label class the weights are proportional to $\exp(\pm\lambda Q_+(z))$, the treatment arm tilted by $+\lambda$ and the control arm by $-\lambda$. The *natural* design takes $\lambda = 0$ and gives an attenuated first call contrast; the *anchored* design solves for the λ at which the decoded first call contrast equals Δ^*, the empirical form of (6). On the ambiguous subpool at temperature one and $\Delta^* = .04$ this gives $\lambda = .51$. Bisection can fail when the decoded contrast jumps across Δ^* at a label flip; those cells are marked in Table 21 and excluded from the figure.

The decoder uses one rule built from the pooled population and does not observe treatment. The stopping rule has a separate continuation probability for each arm at each history, coupled only by the pooled budget, as $W = D$ permits in Section 2. Decoding with each arm's own posterior, which $W \supseteq \{D\}$ permits, splits the *mixed* answer, which carries probability .28 in the pooled first answer distribution, in opposite directions in the two arms; at $\Delta^* = .10$ that alone moves the first call contrast from $+.068$ to $+.350$ on the whole pool and from $+.042$ to $+.535$ on the ambiguous subpool, before any stopping occurs. The displacement studied in this paper is a property of when the gate stops, so both arms use the same decoder while their stopping rules may differ. The common rule comparison below also shares the stopping probabilities.

The decoder estimation check splits the pool in half at random, stratified on the label, and fits the sticky MAP rule using one half. On the test half, the tilt λ is solved again using the true labels to anchor the fitted decoder's first contrast, and stopping is optimized and evaluated with the test population's true response and label law. Thus only the decoder is fitted on separate data; the design and stopping policy use oracle information on the test population. Over five splits, negative optima occur on the budget grid in four at temperature two, with 253 test records, and in two at temperature one, with 120. Appendix P.7 fixes the tilt and the policy on the training half as well.

P.3 One stopping rule shared by both arms

A common rule uses the same continuation probability at each report history in both arms. The linear program uses one variable for the product of continuation probabilities along each path, shared by the two arms, with child products bounded by their parent. The objective is the same pooled accuracy and the constraint the same pooled call budget. Table 8 compares the two policy classes at 20 budgets from .05 to 1 in steps of .05, on seven pools at temperature one. These comparison budgets differ from the 41 point grid in Tables 20 and 21.

In the natural design the distribution of report histories conditional on X is the same in both arms. Under a common rule, let t and f be the probabilities of releasing label one given $X = 1$ and $X = 0$. The arm means are $\pi_d t + (1 - \pi_d) f$, giving $\Delta_\tau = \Delta^*(t - f)$ and $|\Delta_\tau| \leq \Delta^*$. With pooled accuracy $A_\tau = \Pr(G_\tau = X)$ and prevalence one half, $\Delta_\tau = \Delta^*(2A_\tau - 1)$. Anchoring tilts the conditional record mixtures differently across arms, so that identity no longer applies. No common optimum in the comparison grid reverses, but the anchored polarity effect under `gpt-4o` reaches .083 from a truth of .04.

Table 8: Effect ranges on seven pools at temperature one, $\Delta^* = .04$, horizon five, and the 20 budgets $.05, .10, \ldots, 1$. The decoder is shared in both policy classes. Parenthesized pairs give the rounded first and last tested budgets where every optimum with stopping chosen separately by arm is negative. No common rule optimum is negative. Ranges include the numerical extrema across optima.

pool	model	n	design	separate stopping	common stopping	negative budgets
polarity, ambiguous	`gpt-4o-mini`	239	anchored	$[-.037, .056]$	$[.041, .061]$	$(.15, .35)$
			natural	$[.022, .107]$	$[.017, .024]$	none
polarity, ambiguous	`gpt-4.1-mini`	284	anchored	$[-.036, .061]$	$[.042, .067]$	$(.15, .30)$
			natural	$[.019, .105]$	$[.015, .021]$	none
polarity, ambiguous	`gpt-4o`	458	anchored	$[-.291, .115]$	$[.044, .083]$	$(.05, .40)$
			natural	$[.023, .388]$	$[.016, .023]$	none
polarity, premise	`gpt-4o-mini`	159	anchored	$[.036, .111]$	$[.040, .043]$	none
			natural	$[.039, .103]$	$[.032, .039]$	none
tweets, ambiguous	`gpt-4o-mini`	380	anchored	$[.050, .104]$	$[.041, .048]$	none
			natural	$[.019, .076]$	$[.013, .017]$	none
tweets, premise	`gpt-4o-mini`	267	anchored	$[.042, .106]$	$[.041, .045]$	none
			natural	$[.037, .098]$	$[.032, .037]$	none
treebank, ambiguous	`gpt-4.1-mini`	318	anchored	$[.051, .117]$	$[.041, .049]$	none
			natural	$[.020, .089]$	$[.013, .018]$	none

Table 9 executes the common optimum with the smallest effect using the same sample sizes and budgets as Table 18. Anchored coverage ranges from .877 to .983. Natural coverage ranges from .727 to .950, with attenuation leaving the ambiguous pools below the nominal rate. Thus sharing the rule removes the measured reversals but does not make every contrast unbiased. The simulation holds the catalog and policy fixed, draws records with replacement and generates fresh responses from their stored vectors.

Table 9: Common stopping rule: coverage of a nominal 95 percent Wald interval and population effect at the eight displayed budgets, $\Delta^* = .04$, horizon five, 300 replications of 1000 records per arm. The selected accuracy optimum has the smallest effect within the common rule class; both arms use identical continuation probabilities.

pool	design	budget b							
		.05	.10	.15	.20	.25	.30	.40	.60
polarity, premise, 159	anchored	.93	.96	.94	.95	.96	.94	.96	.92
	effect	$+.040$	$+.041$	$+.041$	$+.042$	$+.042$	$+.042$	$+.043$	$+.042$
	natural	.93	.95	.93	.92	.94	.94	.95	.93
	effect	$+.032$	$+.033$	$+.034$	$+.035$	$+.036$	$+.036$	$+.038$	$+.039$
polarity, ambiguous, 239	anchored	.95	.97	.94	.93	.94	.95	.94	.88
	effect	$+.041$	$+.043$	$+.044$	$+.045$	$+.046$	$+.048$	$+.050$	$+.054$
	natural	.79	.81	.80	.87	.86	.84	.87	.88
	effect	$+.017$	$+.017$	$+.018$	$+.018$	$+.019$	$+.019$	$+.020$	$+.022$
tweets, ambiguous, 380	anchored	.93	.97	.98	.92	.92	.94	.95	.92
	effect	$+.041$	$+.041$	$+.042$	$+.043$	$+.043$	$+.044$	$+.045$	$+.047$
	natural	.75	.73	.78	.74	.75	.79	.80	.84
	effect	$+.013$	$+.013$	$+.013$	$+.014$	$+.014$	$+.014$	$+.015$	$+.016$

P.4 Two further corpora and two further models

The same protocol was run on two corpora whose ground truth has a middle class, and with two further models. The tweets are the sentiment test split of TweetEval (Barbieri et al., 2020), labelled negative, neutral or positive; the treebank sentences are the five class Stanford Sentiment Treebank (Socher et al., 2013), all splits, labelled from very negative to very positive. For both, the truth is $X = 1$ for a positive text and $X = 0$ otherwise, so the neutral answer has a home in the truth: a neutral text answered *neutral* is correct, and no coding decision assigns the middle answer a side. Each pool

holds 1500 positive texts and 1500 drawn from the rest in proportion, tweets 572 negative and 928 neutral, treebank 319 very negative, 698 negative and 483 neutral against 940 positive and 560 very positive, all between 20 and 300 characters. The prompt replaces *mixed* by *neutral*, and "movie-review sentence" by "tweet" for the tweets. The models are `gpt-4o-mini`, `gpt-4.1-mini` and `gpt-4o`; the `gpt-5` family does not return log probabilities and was excluded. All nine corpus and model pairs were annotated; the treebank under `gpt-4o-mini` ran last, across a daily request cap. Each model name is a served endpoint rather than fixed weights, so the archived catalog, not a rerun, is what a reproduction must match. Off vocabulary mass averaged below 10^{-3} on eight kernels and .008 on the treebank under `gpt-4o-mini`, which answered *mixed* on 18 sentences where the prompt offered *neutral*; that mass is dropped as off vocabulary and the remaining mass renormalized over the three words.

The ambiguous pool check uses the same 100 records at temperatures one and two, with ten completions in one request per record. Table 10 reports the archived frequencies and the exact expected total variation under independent draws from the stored vector, conditioning on each record's number of answers within the three categories. The observed distances exceed this reference. Section P.5 locates the excess: every call returns its own first token distribution, the catalog is one draw of it, and separate requests are consistent with independent draws from its mean, whereas the completions of one request are not.

Table 10: Marginal frequency fit on the same 100 ambiguous polarity records, with 1000 requested completions per temperature. Total variation and modal frequencies are averaged over records after excluding answers outside the three categories. The independent reference uses each record's effective completion count and stored response vector.

temperature	observed TV	expected TV	outside	modal observed	modal predicted
1	.158	.098	2	.747	.785
2	.145	.121	11	.673	.667

For each record and category, the independent reference sums $\Pr(K = k)|k/n - q|$ over $k = 0, \ldots, n$ for $K \sim \text{Binomial}(n, q)$, then sums over categories and divides by two. Here n excludes answers outside the categories. The standard deviation $\sqrt{q(1-q)/n}$ bounds the mean absolute deviation from above and is not its expectation.

P.5 Recorded repeated answers

The catalog fixes one response vector per record and the simulations draw from it independently. To test both steps against the endpoint, every record of the ambiguous polarity pools under the three models, and of the ambiguous tweets under `gpt-4o-mini`, was queried again on 24 September 2026: twenty separate requests and one request for twenty completions, each at temperature one with a two token limit, the catalog prompt, and the twenty most probable first tokens returned for every completion. A record's requests were issued together, in the order an adaptive pipeline would issue them, and records in random order. This took 28,581 requests and about two dollars. Answers outside the three words are dropped, as the catalog renormalizes over them; they number at most 35 per pool and sampling protocol, and 18 of each kind come from one polarity sentence whose catalog already placed .945 of its mass outside the alphabet, which keeps its catalog vector. The diagnostics therefore cover 238 of the 239 records of that pool; the replays keep that record with its catalog vector and leave the design weights unchanged, and the cross fitted comparison splits each record's twenty separate requests into the first and the last ten and drops invalid answers within each group, so the two groups never share an answer; a group with fewer than five valid answers keeps the record's catalog vector, which happens for 3 of the 478 groups of that pool and for none elsewhere.

Table 11 compares the recorded answers with independent draws. Each call's vector is its first token mass on the three words, renormalized, and a record's mean vector averages the vectors of the calls whose sampled answer lies in the alphabet. The first token distribution returned by a call is not fixed: across separate requests it lies at total variation .035 to .060 from the record's mean vector, and the catalog vector, one such draw, lies .045 to .087 from that mean. Given the mean, separate requests are consistent with independent draws: the share of agreeing pairs of answers matches its expectation within .01, inside 1.5 standard errors on every pool, and the total variation between frequencies and

the mean vector matches its binomial expectation. These are checks of pairwise agreement and marginal frequencies, not a test of the conditional independence of a whole answer sequence. This resolves the excess in Table 10, which measured frequencies against the catalog draw rather than against the mean. The twenty completions of a single request are a different matter: on average .95 of a request's completions carry its most common first token vector, all of them do in only 6 to 11 percent of requests, and they agree .020 to .031 more often than separate requests, from two to three and a half standard errors.

To replay a policy on recorded answers, a record's history law is the chance that t of its answers, drawn without replacement in random order, spell the history; averaged over the draw this is the probability that the real process produces the history whenever the answers are exchangeable, whatever their dependence. The tree is indexed as in the catalog, so a catalog policy applies to it index by index, with the catalog decoder and design weights. Standard errors resample each record's separate answers with replacement, 500 times, holding weights, decoder and policy fixed; the completions of one request are never resampled, since they are not independent. Table 12 evaluates the catalog optimum with the smallest effect on the catalog, on the separate answers and on the completions of the single request. On separate answers no released effect moves by more than .004, or .010 under `gpt-4o`; on the completions of one request by at most .013. Repeating the coverage simulation of Table 18 with fresh draws taken from the recorded separate answers gives minima of .08, .09, .05 and .11 on the premise, the two smaller polarity and the tweets pools under the anchored design and .17, .12, .18 and .50 without tilt, against .11, .03, .05 and .13 anchored and .16, .11, .11 and .61 without tilt from the catalog with the same seeds; under `gpt-4o` both read .00 between $b = .10$ and .30.

Table 13 asks whether the sign survives when the optimization itself uses the recorded answers. Stopping optimized on the recorded law keeps the reversal on the two smaller polarity pools, in sample and when fitted on one half of each record's requests and evaluated on the other. Under `gpt-4o` it does not: the recorded optimum releases $+.44$ at $b = .25$, where the replayed catalog optimum releases $-.29$. That policy spends .256 calls on the recorded law rather than .25, and at that cost the recorded optimum, $+.43$, is .0011 more accurate, so the replayed policy lies just outside the set of policies within .001 of optimal accuracy, which spans $[-.21, +.45]$ there. The two are not far apart in the sense that matters. Policies within .001 of optimal accuracy span both signs on every ambiguous polarity pool, on the catalog as on the recorded answers, and Proposition 2.1 offers the analogous explanation in the four type model: the accuracy given up to preserve the effect is $\chi H(b)$ with $\chi \propto k_C - k_B$, so when the two routes cost nearly the same, accuracy no longer chooses between them. The decoder sits on a similar margin. On the pooled first answer *mixed*, which carries .46 of the first answer mass under `gpt-4o`, the catalog posterior is .504 and the recorded one .498; refitting the decoder on the recorded answers flips that label, and after the tilt is solved again the arm specific optimum inflates .04 to .111 instead of reversing. The same flip occurs under `gpt-4.1-mini`, from .520 to .496, with inflation to .418. What the recorded answers fix is therefore the size of the displacement, which Theorem 6.1 bounds, and not its direction.

Table 11: Recorded answers against independent draws. Distances are total variation, averaged over records: between one call's first token vector and the mean of the record's call vectors, and between the catalog vector and that mean. Agreement is the share of pairs of answers that coincide; the first difference compares separate requests with independent draws from the mean vector, the second compares completions of one request with separate requests. Standard errors in parentheses. The last column is the share of a request's completions that carry its most common first token vector.

pool	model	n	calls	call to mean	catalog to mean	separate − independent	within − separate	shared
polarity, ambiguous	`gpt-4o-mini`	238	20	0.048	0.048	+0.002 (0.006)	+0.025 (0.012)	0.96
polarity, ambiguous	`gpt-4.1-mini`	284	20	0.035	0.060	+0.004 (0.006)	+0.024 (0.012)	0.95
polarity, ambiguous	`gpt-4o`	458	20	0.060	0.087	−0.002 (0.004)	+0.020 (0.007)	0.95
tweets, ambiguous	`gpt-4o-mini`	380	20	0.055	0.045	−0.003 (0.005)	+0.031 (0.009)	0.95

P.6 Stopping rules blind to treatment

The comparison of Section P.3 optimizes a common rule. Rules used in practice are simpler: release the first answer, release after three or five calls, stop as soon as two consecutive answers coincide, or stop as soon as one semantic label leads the other by two, the symmetric boundary with $m = 2$. None observes treatment. In the natural design any such rule satisfies $\Delta_\tau = \Delta^*(t - f)$ and, since pooled

Table 12: Released population effect of the catalog optimum with the smallest effect, $\Delta^* = .04$, horizon five, evaluated on the catalog, on the recorded separate answers and on the completions of the single request. Decoder, design weights and policy are those of the catalog. Bootstrap standard errors for the replays on separate requests are at most 0.004.

		$b = .15$			$b = .25$		
pool	design, stopping	catalog	separate	within	catalog	separate	within
polarity, premise, `gpt-4o-mini`	anchored, by arm	$+0.111$	$+0.112$	$+0.108$	$+0.083$	$+0.083$	$+0.080$
	natural, by arm	$+0.103$	$+0.105$	$+0.101$	$+0.077$	$+0.078$	$+0.074$
	anchored, common	$+0.041$	$+0.040$	$+0.037$	$+0.042$	$+0.041$	$+0.039$
polarity, ambiguous, `gpt-4o-mini`	anchored, by arm	-0.012	-0.010	-0.002	-0.037	-0.034	-0.026
	natural, by arm	$+0.072$	$+0.070$	$+0.071$	$+0.107$	$+0.105$	$+0.103$
	anchored, common	$+0.044$	$+0.045$	$+0.050$	$+0.046$	$+0.047$	$+0.052$
polarity, ambiguous, `gpt-4.1-mini`	anchored, by arm	-0.012	-0.011	-0.002	-0.036	-0.034	-0.023
	natural, by arm	$+0.069$	$+0.067$	$+0.063$	$+0.105$	$+0.102$	$+0.094$
	anchored, common	$+0.045$	$+0.045$	$+0.048$	$+0.048$	$+0.048$	$+0.050$
polarity, ambiguous, `gpt-4o`	anchored, by arm	-0.208	-0.215	-0.205	-0.280	-0.289	-0.277
	natural, by arm	$+0.271$	$+0.276$	$+0.267$	$+0.388$	$+0.396$	$+0.384$
	anchored, common	$+0.053$	$+0.052$	$+0.053$	$+0.062$	$+0.060$	$+0.061$
tweets, ambiguous, `gpt-4o-mini`	anchored, by arm	$+0.086$	$+0.088$	$+0.082$	$+0.104$	$+0.107$	$+0.099$
	natural, by arm	$+0.057$	$+0.060$	$+0.054$	$+0.076$	$+0.080$	$+0.073$
	anchored, common	$+0.042$	$+0.041$	$+0.039$	$+0.043$	$+0.043$	$+0.040$

Table 13: The sign at $b = .25$, anchored design, stopping chosen separately by arm. Columns: the catalog optimum on the catalog and on the recorded answers; the optimum computed on the recorded answers, in sample and fitted on one half of each record's calls and evaluated on the other; the range of effects over policies within $.001$ of optimal accuracy, on the catalog and on the recorded answers; and the optimum after the decoder is also derived from the recorded answers and the tilt λ solved again to anchor the first call. Cross fitting splits each record's twenty separate requests into the first and the last ten and keeps the valid answers of each group; a group with fewer than five valid answers keeps the record's catalog vector.

	catalog optimum		recorded optimum		within .001 of optimal accuracy		decoder	refitted
pool	catalog	recorded	in sample	cross fitted	catalog	recorded	λ	effect
polarity, premise, `gpt-4o-mini`	$+0.083$	$+0.083$	$+0.083$	$+0.083$	$[+0.026, +0.089]$	$[+0.031, +0.089]$	0.59	$+0.084$
polarity, ambiguous, `gpt-4o-mini`	-0.037	-0.034	-0.034	-0.033	$[-0.046, +0.107]$	$[-0.043, +0.132]$	0.48	-0.035
polarity, ambiguous, `gpt-4.1-mini`	-0.036	-0.034	-0.032	-0.032	$[-0.045, +0.133]$	$[-0.044, +0.133]$	0.18	$+0.418$
polarity, ambiguous, `gpt-4o`	-0.280	-0.289	$+0.440$	$+0.440$	$[-0.322, +0.447]$	$[-0.199, +0.455]$	0.19	$+0.111$
tweets, ambiguous, `gpt-4o-mini`	$+0.104$	$+0.107$	$+0.110$	$+0.110$	$[+0.065, +0.115]$	$[+0.070, +0.116]$	0.14	$+0.111$

Table 14: Stopping rules that never observe treatment, with the pooled catalog decoder, $\Delta^* = .04$, horizon five. The tilt λ weights records within each class by $\exp(\mp\lambda Q_+)$ in control and treatment; $\lambda = 0$ is the natural design and the middle row is the anchored design. The last column repeats the five call rule on the recorded separate answers.

pool	λ	first answer	three calls	five calls	two agree	lead of two	five calls, recorded
polarity, ambiguous, `gpt-4o-mini`	0	0.017	0.023	0.025	0.021	0.021	0.025
	0.51	0.040	0.059	0.066	0.052	0.052	0.067
	1	0.063	0.093	0.105	0.083	0.083	0.107
polarity, ambiguous, `gpt-4.1-mini`	0	0.014	0.020	0.022	0.018	0.018	0.021
	0.53	0.040	0.058	0.071	0.054	0.055	0.069
	1	0.063	0.091	0.113	0.082	0.082	0.112
polarity, ambiguous, `gpt-4o`	0	0.016	0.022	0.024	0.020	0.020	0.025
	0.39	0.040	0.083	0.083	0.081	0.081	0.080
	1	0.078	0.179	0.174	0.177	0.177	0.167
tweets, ambiguous, `gpt-4o-mini`	0	0.012	0.017	0.018	0.016	0.015	0.020
	0.14	0.040	0.048	0.049	0.047	0.047	0.049
	1	0.213	0.244	0.239	0.242	0.242	0.233

prevalence is one half, $\Delta_\tau = \Delta^*(2A_\tau - 1)$ exactly; the released effect then rises with the rule's accuracy and never exceeds the truth. Table 14 evaluates the rules on the catalog and on the recorded separate answers along the within class tilt λ that defines the anchored design. At the positive tilts tested, where the two arms express a class with different clarity, every rule releases an effect above the truth, although the effect need not grow with the number of calls, and none reverses. The accuracy optima along the same tilt behave differently: the common optimum is never negative and stays above the first call contrast, whereas the optimum chosen separately by arm on the polarity pools inflates for $\lambda \leq .35$ and reverses for λ from .5 to 1, the switch occurring between .35 and .40 under `gpt-4o-mini`, and between .45 and .50 under `gpt-4.1-mini`, with no decoder label changing: the cheaper revisions move to the other arm.

P.7 Stopping fixed in advance

The comparisons above optimize stopping on the population they report, as the decoder estimation check still does for the tilt and the policy. Here the whole pipeline is fixed on training records. The premise pool and the nine ambiguous pools of the nine kernels are each split at random into halves stratified by label, fifty times. On the training half, on the catalog, we fit the pooled sticky MAP decoder, solve the anchored tilt, and choose the policy: for stopping by arm, the price optimal Bellman policies of Section 3 for the two arms at one price, ties stopping; for a common rule, the same on the pooled training law. A deterministic price policy spends calls in jumps, so each budget of .15, .25 and .40 extra calls is met exactly by randomizing, record by record, between the two Bellman policies on either side of the critical price, which is the accuracy optimum at that budget chosen without reference to the effect. The test half builds its arms from its own labels at prevalences .48 and .52 with the training tilt, without solving the anchor again, so its first call contrast need not equal Δ^*; its labels enter only there and in scoring. The frozen pipeline is evaluated on the test catalog and, where recorded answers exist, on the test records' recorded separate answers. Two label free rules stand beside these: a middle first answer is asked again with a probability tuned on each arm's training answers so that the arm spends the budget, after which the record stops at a semantic lead of two or at the horizon; and the same rule with one probability tuned on the pooled answers. Neither tunes its stopping with a label, although the released label still comes from the decoder fitted with training labels; the first is the per batch tuning of practice. Two references complete the design: stopping optimized on the held out half itself with the training decoder and tilt, at the calls the fitted by arm policy actually spends there, which shares that policy's population and cost and so measures the cost of learning the policy, and a full refit of decoder, tilt and policy on the held out half, which changes that population and is reported for completeness.

Table 15 reports $b = .25$, and Table 16 the held out first call contrast, accuracy gain and calls. Fitted in advance, arm specific stopping has mean absolute held out contrast displacement of .06–.12 on the nine ambiguous pools, within .017 of the corresponding mean for stopping optimized on the held out half at the same cost and within about .04 of the corresponding mean under the full refit, and leaves coverage at .03 to .22 under the anchored design and .17 to .69 without tilt; on the premise pool the displacement is .04 to .05. The budget is spent on the held out half as intended, .24 to .26 extra calls on average, every held out first call contrast is positive, and the mean accuracy gains of the fitted, refit and common policies agree within .003. The sign fitted on the training half is the sign observed on the test half in 96 to 100 percent of halvings at $b \geq .25$, and it is the sign of the held out half's own optimum at the same cost in 96 percent of them. Negative effects occur in 74, 54 and 18 percent of anchored halvings of the three polarity pools, in 2 percent of the treebank halvings under `gpt-4o-mini` and in none of the others; stopping refit on the held out half reverses in up to 84 percent of anchored polarity halvings and in 8 percent of those treebank halvings; no halving reverses without the tilt. The common rule fitted the same way, and both label free rules, have mean absolute contrast displacement at most .008 and no observed reversals, with mean coverage on the ambiguous pools of .87–.95 under the anchored design and .68–.83 without tilt, the latter reflecting attenuation. The fitted arm specific rule's mean accuracy advantage over the common rule is .0005–.003, much smaller than its mean absolute contrast displacement. Coverage here is the normal approximation for 1000 records per arm; on the first ten halvings of the three polarity pools, 300 simulated studies per cell agree with it to a mean absolute difference of .014 over 240 cells, 95 percent of them within two Monte Carlo standard errors, and the simulated effects agree within .004.

Table 15: Stopping fixed on half of each pool and evaluated on the other half, on the recorded answers of the held out records, $b = .25$, $\Delta^* = .04$, horizon five, 50 random halvings stratified by label. Each block gives the mean absolute displacement $|\Delta_\tau - \Delta_1|$, the share of halvings with a negative effect, and the mean coverage of a nominal 95 percent interval with 1000 records per arm. The first block optimizes stopping on the held out half itself, with the decoder and tilt of the training half and at the calls the fitted by arm policy actually spends there, so it shares that policy's population and cost; by arm and common policies are accuracy optima on the price path fitted on the training half; the label free rule asks a middle first answer again with a probability tuned on each arm's training answers to meet the budget, stops at a lead of two, and releases the training decoder's label. A dagger marks pools without recorded answers, evaluated on the catalog.

		refit at matched cost			fitted, by arm			fitted, common			label free, by arm		
pool	design	|disp|	< 0	cov.	|disp|	< 0	cov.	|disp|	< 0	cov.	|disp|	< 0	cov.
polarity, premise, `gpt-4o-mini`	anchored	0.043	0.00	0.53	0.043	0.00	0.53	0.003	0.00	0.95	0.003	0.00	0.95
	natural	0.046	0.00	0.62	0.046	0.00	0.62	0.005	0.00	0.95	0.003	0.00	0.94
polarity, ambiguous, `gpt-4o-mini`	anchored	0.097	0.84	0.13	0.099	0.74	0.13	0.007	0.00	0.87	0.004	0.00	0.89
	natural	0.098	0.00	0.17	0.107	0.00	0.17	0.002	0.00	0.83	0.002	0.00	0.82
polarity, ambiguous, `gpt-4.1-mini`	anchored	0.105	0.62	0.07	0.104	0.54	0.08	0.008	0.00	0.90	0.004	0.00	0.91
	natural	0.111	0.00	0.17	0.109	0.00	0.19	0.002	0.00	0.78	0.002	0.00	0.77
polarity, ambiguous, `gpt-4o`	anchored	0.106	0.20	0.06	0.102	0.18	0.06	0.007	0.00	0.91	0.005	0.00	0.93
	natural	0.132	0.00	0.25	0.118	0.00	0.27	0.002	0.00	0.80	0.002	0.00	0.80
tweets, ambiguous, `gpt-4o-mini`	anchored	0.070	0.00	0.10	0.075	0.00	0.10	0.004	0.00	0.94	0.003	0.00	0.95
	natural	0.069	0.00	0.51	0.074	0.00	0.51	0.002	0.00	0.76	0.001	0.00	0.74
tweets, ambiguous, `gpt-4.1-mini`†	anchored	0.128	0.00	0.02	0.111	0.00	0.03	0.006	0.00	0.93	0.004	0.00	0.93
	natural	0.126	0.00	0.30	0.119	0.00	0.33	0.002	0.00	0.74	0.001	0.00	0.73
tweets, ambiguous, `gpt-4o`†	anchored	0.087	0.02	0.02	0.093	0.00	0.03	0.004	0.00	0.94	0.003	0.00	0.94
	natural	0.097	0.00	0.30	0.104	0.00	0.29	0.002	0.00	0.75	0.001	0.00	0.74
treebank, ambiguous, `gpt-4o-mini`†	anchored	0.115	0.08	0.08	0.115	0.02	0.07	0.006	0.00	0.92	0.005	0.00	0.93
	natural	0.118	0.00	0.41	0.118	0.00	0.42	0.001	0.00	0.69	0.001	0.00	0.68
treebank, ambiguous, `gpt-4.1-mini`†	anchored	0.100	0.00	0.04	0.102	0.00	0.04	0.005	0.00	0.94	0.003	0.00	0.94
	natural	0.102	0.00	0.32	0.104	0.00	0.33	0.002	0.00	0.76	0.001	0.00	0.74
treebank, ambiguous, `gpt-4o`†	anchored	0.058	0.00	0.22	0.058	0.00	0.22	0.003	0.00	0.95	0.003	0.00	0.95
	natural	0.057	0.00	0.69	0.057	0.00	0.69	0.001	0.00	0.78	0.001	0.00	0.77

Table 16: Companion to Table 15: on the held out half, the first call contrast Δ_1 under the training tilt, the accuracy gained over the first call by each policy, and the extra calls per record actually spent, all averaged over the 50 halvings at the training budget $b = .25$. Every held out first call contrast is positive, so a negative effect in Table 15 is a reversal produced by stopping.

			accuracy gain				extra calls		
pool	design	Δ_1	refit, matched	by arm	common	label free	by arm	common	label free
polarity, premise, `gpt-4o-mini`	anchored	+0.039	0.058	0.058	0.057	0.037	0.26	0.26	0.26
	natural	+0.031	0.058	0.058	0.057	0.037	0.26	0.26	0.26
polarity, ambiguous, `gpt-4o-mini`	anchored	+0.045	0.028	0.028	0.027	0.021	0.25	0.24	0.25
	natural	+0.016	0.029	0.028	0.027	0.021	0.25	0.24	0.25
polarity, ambiguous, `gpt-4.1-mini`	anchored	+0.042	0.026	0.026	0.025	0.019	0.24	0.24	0.24
	natural	+0.013	0.027	0.027	0.025	0.019	0.24	0.24	0.24
polarity, ambiguous, `gpt-4o`	anchored	+0.039	0.030	0.030	0.030	0.022	0.26	0.26	0.26
	natural	+0.015	0.032	0.032	0.029	0.021	0.26	0.26	0.26
tweets, ambiguous, `gpt-4o-mini`	anchored	+0.039	0.022	0.022	0.020	0.015	0.26	0.26	0.26
	natural	+0.012	0.022	0.022	0.020	0.015	0.26	0.26	0.26
tweets, ambiguous, `gpt-4.1-mini`†	anchored	+0.041	0.026	0.025	0.024	0.017	0.25	0.24	0.24
	natural	+0.011	0.027	0.027	0.024	0.017	0.25	0.24	0.24
tweets, ambiguous, `gpt-4o`†	anchored	+0.040	0.023	0.023	0.021	0.015	0.25	0.25	0.25
	natural	+0.012	0.023	0.022	0.021	0.015	0.25	0.25	0.25
treebank, ambiguous, `gpt-4o-mini`†	anchored	+0.040	0.018	0.018	0.016	0.012	0.25	0.25	0.25
	natural	+0.009	0.018	0.018	0.016	0.012	0.25	0.25	0.25
treebank, ambiguous, `gpt-4.1-mini`†	anchored	+0.040	0.024	0.024	0.022	0.016	0.25	0.25	0.25
	natural	+0.012	0.024	0.024	0.022	0.016	0.25	0.25	0.25
treebank, ambiguous, `gpt-4o`†	anchored	+0.040	0.020	0.020	0.018	0.013	0.25	0.25	0.25
	natural	+0.014	0.019	0.019	0.018	0.013	0.25	0.25	0.25

P.8 Coding of the middle answer

On the polarity corpus the semantic map sends *mixed* to label zero, the counterpart of the B report in Section 2. The map governs three things: the semantic positive probability $Q_+(z)$ along which the anchored design tilts, the partition of the ambiguous pool into records that answer correctly more often than not and records that do not, and the descriptive accuracy of the modal answer. It does not govern the released label. The gate is a sticky MAP rule for the truth X built from the pooled population, and on four of the five polarity pools it sends a first answer of *mixed* to the positive side, because positive sentences are called mixed more often than negative ones are.

The map does more than it seems. Under `gpt-4o-mini`, 522 of the 3000 records have a semantically wrong modal answer, and 445 of them are answered *mixed*; only 77 have the opposite polarity, and with *mixed* set aside the modal answer is correct for .964 of records. Of the 80 ambiguous records that answer incorrectly more often than not, 66 are positive sentences the model calls mixed. Sending *mixed* to the positive side instead moves both the tilt and the partition. The natural design, which has no tilt, is unchanged by construction. The anchored design, tilted along the other definition of Q_+, loses every wrong sign interval: on the 239 ambiguous records its optimum then lifts .04 to .130, a displacement of .090 against the .077 of the reversal. The reversal is therefore conditional on the coding and on the tilt, and the displacement is not, which is why the two further corpora were chosen to carry a middle class in the truth: there every kernel inflates and none reverses.

P.9 The threshold in the true effect

Section 2 says a reversal needs the true effect to be smaller than the mass of cheap revisions, and at the budget $k_B L$ the released effect is $\delta - L$ whatever δ is: the testable content is a displacement that does not follow the truth, and a flip where the truth overtakes it. Horizon five puts these measurements outside Equation (9), which is proved for three calls, but Theorem 6.1 makes the reachable displacement a functional of the first revision exposure at any horizon, with the truth entering only through the arm weights. On the 239 ambiguous polarity records at temperature one the displacement of the lowest accuracy optimum is nearly constant in Δ^*, and the sign flips where Δ^* crosses it (Table 17). Bisection places the flip at $\Delta^* \in (.0686, .0698)$ under `gpt-4o-mini`, $(.0663, .0674)$ under `gpt-4.1-mini` and $(.0885, .0897)$ under `gpt-4o`, between .28 and .39 of the nonmodal answer probability of each pool; on the 169 records below cutoff .90 it is $(.0838, .0850)$, and on the whole polarity corpus under `gpt-4o`, where the nonmodal answer probability is .038, effects below .011 still flip.

Table 17: Extreme effect among accuracy optima on the 239 ambiguous polarity records under `gpt-4o-mini`, temperature one, horizon five, as the true effect varies: the lowest optimum under the anchored design and the highest under the natural design, whose lowest is the attenuated first call. The bias column is $\Delta_\tau - \Delta^*$; under the anchored design $\Delta_1 = \Delta^*$, so it is also the displacement. Extrema use the 41 budgets in $[0, 2]$; parenthesized budget pairs give rounded grid ranges, including the endpoints.

Δ^*	design	extreme optimum	bias $\Delta_\tau - \Delta^*$	wrong sign budgets
0.02	anchored	-0.063	-0.083	$(0.10, 0.70)$
0.02	natural	$+0.098$	$+0.078$	none
0.04	anchored	-0.037	-0.077	$(0.15, 0.35)$
0.04	natural	$+0.107$	$+0.067$	none
0.08	anchored	$+0.012$	-0.068	none
0.08	natural	$+0.125$	$+0.045$	none
0.16	anchored	$+0.103$	-0.057	none
0.16	natural	$+0.208$	$+0.048$	none

P.10 Coverage against the budget

Table 18 executes the accuracy optimum with the smallest effect at each displayed budget, with stopping optimized separately by arm, using 1000 records per arm and 300 fresh replications, and reports both the released population effect and how often a nominal 95 percent interval covers the truth. Three pools are shown: the 159 polarity records that satisfy the premise, the 239 ambiguous

polarity records, and the 380 ambiguous tweets under `gpt-4o-mini`. Each curve reaches its minimum between .15 and .25 extra calls per record, at .04 on the ambiguous polarity records under the anchored design and at .57 on the tweets without the tilt. The premise pool is back above .85 by .30; the ambiguous polarity pool recovers past .4; the ambiguous tweets under the anchored design do not, reading .79 at .40 and .63 at .60 as the released effect drifts upward again. On these grids, reversal appears only when the pool includes records that violate the premise, whereas poor coverage occurs either way. The $b = .30$ cell for the ambiguous polarity records also appears in Table 22, from an earlier set of 300 replications, at .20 against .27 here, two Monte Carlo standard errors apart.

Table 18: Coverage of the nominal 95 percent Wald interval and the released population effect at the eight displayed budgets, $\Delta^* = .04$, horizon five, 300 replications of 1000 records per arm. The optimum with the smallest effect uses stopping optimized separately by arm. The catalog and policy are fixed across replications.

pool	design	budget b							
		.05	.10	.15	.20	.25	.30	.40	.60
polarity, premise, 159	anchored	0.83	0.41	0.14	0.14	0.49	0.86	0.91	0.94
	effect	+0.064	+0.087	+0.111	+0.104	+0.083	+0.061	+0.056	+0.036
polarity, premise, 159	natural	0.90	0.60	0.16	0.25	0.57	0.90	0.93	0.93
	effect	+0.055	+0.079	+0.103	+0.098	+0.077	+0.055	+0.051	+0.046
polarity, ambiguous, 239	anchored	0.82	0.63	0.34	0.07	0.04	0.27	0.85	0.91
	effect	+0.023	+0.005	−0.012	−0.030	−0.037	−0.019	+0.018	+0.025
polarity, ambiguous, 239	natural	0.95	0.89	0.67	0.38	0.13	0.33	0.92	0.93
	effect	+0.035	+0.053	+0.072	+0.090	+0.107	+0.089	+0.055	+0.048
tweets, ambiguous, 380	anchored	0.87	0.63	0.45	0.13	0.14	0.32	0.79	0.63
	effect	+0.055	+0.071	+0.086	+0.102	+0.104	+0.091	+0.064	+0.073
tweets, ambiguous, 380	natural	0.91	0.97	0.88	0.63	0.57	0.81	0.96	0.95
	effect	+0.027	+0.042	+0.057	+0.072	+0.076	+0.062	+0.033	+0.042

Table 19: The measured kernel, whole pool and by ambiguity cutoff. A record is kept when its most probable answer falls below the cutoff. The Youden index is $\Pr(\text{positive} \mid X = 1) - \Pr(\text{positive} \mid X = 0)$ on the first call, and $\mathrm{Var}(Q_+ \mid X = 1)$ is the covariance of two fresh semantic positive indicators given the truth.

temp.	cutoff	n	$\mathbb{E}[1 - \max_r Q_r]$	mode correct	Youden	$\mathrm{Var}(Q_+ \mid X{=}1)$
1	all	3000	0.019	0.826	0.652	0.213
1	0.99	404	0.137	0.634	0.309	0.148
1	0.95	239	0.213	0.669	0.359	0.119
1	0.90	169	0.268	0.651	0.338	0.089
1	0.80	106	0.339	0.708	0.384	0.071
2	all	3000	0.039	0.826	0.651	0.202
2	0.99	751	0.154	0.672	0.363	0.151
2	0.95	505	0.216	0.655	0.346	0.119
2	0.90	389	0.258	0.638	0.318	0.094
2	0.80	240	0.330	0.667	0.356	0.071

P.11 FULL GRID

Table 21 gives, for three effect sizes, both admissible temperatures and five pools, the range of the released effect over accuracy optima at horizon five on a grid of 41 budgets in $[0, 2]$, together with the tested budgets at which every optimum is negative. Parenthesized endpoints are rounded grid ranges, not certified continuous intervals. The implementation optimizes effects subject to an accuracy tolerance of 10^{-11} and, if that secondary linear program fails, multiplies the tolerance by 100 and retries, up to six times. Recomputing the grids behind Tables 20, 21 and 17 with the tolerance logged, none of the 10,004 secondary solves at 10^{-11}, nor of the 10,004 at the near optimal slack of .001, needed a wider tolerance, and every recomputed extremum matched the archived value. The reported ranges therefore concern a numerical tolerance around optimal accuracy; they do not certify extrema over exact optimizers or behavior between grid points. Table 20 extends it to every corpus and model at temperature one and $\Delta^* = .04$.

Three patterns hold across both. The natural design has no negative optimum on the grid. The anchored design has negative optima on the ambiguous polarity pools under every model; on tweets

and treebank texts the largest effects instead exceed the truth, with no negative optimum on the grid. The size of the displacement tracks the nonmodal answer probability rather than the accuracy of the annotator, typically at a third of it, and on nine cells, the largest on the tweets under `gpt-4.1-mini` below cutoff .90, at more than the nonmodal answer probability itself.

Table 20: Every measured kernel at temperature one, $\Delta^* = .04$, horizon five: the whole pool and the records below two cutoffs, both designs. Lowest and highest effects range over numerical accuracy optima and the 41 budgets in $[0, 2]$. Parenthesized budget pairs are rounded grid ranges including endpoints; stopping is optimized separately by arm.

task	model	pool	n	$\mathbb{E}[1 - \max_r Q_r]$	design	Δ_1	lowest	highest	wrong sign budgets
polarity	`gpt-4.1-mini`	all	3000	0.024	anchored	+0.040	+0.032	+0.042	none
polarity	`gpt-4.1-mini`	all	3000	0.024	natural	+0.029	+0.029	+0.037	none
polarity	`gpt-4.1-mini`	< 0.95	284	0.238	anchored	+0.040	−0.036	+0.071	(0.15, 0.30)
polarity	`gpt-4.1-mini`	< 0.95	284	0.238	natural	+0.014	+0.014	+0.105	none
polarity	`gpt-4.1-mini`	< 0.90	214	0.290	anchored	+0.040	+0.040	+0.151	none
polarity	`gpt-4.1-mini`	< 0.90	214	0.290	natural	+0.013	+0.013	+0.118	none
polarity	`gpt-4o`	all	3000	0.038	anchored	+0.040	+0.031	+0.043	none
polarity	`gpt-4o`	all	3000	0.038	natural	+0.031	+0.031	+0.042	none
polarity	`gpt-4o`	< 0.95	458	0.230	anchored	+0.040	−0.291	+0.115	(0.05, 0.40)
polarity	`gpt-4o`	< 0.95	458	0.230	natural	+0.016	+0.016	+0.388	none
polarity	`gpt-4o`	< 0.90	351	0.278	anchored	+0.040	−0.172	+0.140	(0.50, 0.65)
polarity	`gpt-4o`	< 0.90	351	0.278	natural	+0.016	+0.016	+0.349	none
polarity	`gpt-4o-mini`	all	3000	0.019	anchored	+0.040	+0.034	+0.043	none
polarity	`gpt-4o-mini`	all	3000	0.019	natural	+0.027	+0.027	+0.035	none
polarity	`gpt-4o-mini`	< 0.95	239	0.213	anchored	+0.040	−0.037	+0.067	(0.15, 0.35)
polarity	`gpt-4o-mini`	< 0.95	239	0.213	natural	+0.017	+0.017	+0.107	none
polarity	`gpt-4o-mini`	< 0.90	169	0.268	anchored	+0.040	−0.058	+0.071	(0.10, 0.35)
polarity	`gpt-4o-mini`	< 0.90	169	0.268	natural	+0.017	+0.017	+0.128	none
tweets	`gpt-4.1-mini`	all	3000	0.039	anchored	+0.040	+0.040	+0.052	none
tweets	`gpt-4.1-mini`	all	3000	0.039	natural	+0.026	+0.026	+0.038	none
tweets	`gpt-4.1-mini`	< 0.95	460	0.236	anchored	+0.040	+0.040	+0.118	none
tweets	`gpt-4.1-mini`	< 0.95	460	0.236	natural	+0.012	+0.012	+0.083	none
tweets	`gpt-4.1-mini`	< 0.90	358	0.280	anchored	+0.040	+0.040	+0.436	none
tweets	`gpt-4.1-mini`	< 0.90	358	0.280	natural	+0.012	+0.012	+0.419	none
tweets	`gpt-4o`	all	3000	0.049	anchored	+0.040	+0.040	+0.058	none
tweets	`gpt-4o`	all	3000	0.049	natural	+0.020	+0.020	+0.038	none
tweets	`gpt-4o`	< 0.95	604	0.228	anchored	+0.040	+0.040	+0.120	none
tweets	`gpt-4o`	< 0.95	604	0.228	natural	+0.013	+0.013	+0.092	none
tweets	`gpt-4o`	< 0.90	464	0.274	anchored	+0.040	+0.040	+0.380	none
tweets	`gpt-4o`	< 0.90	464	0.274	natural	+0.012	+0.012	+0.358	none
tweets	`gpt-4o-mini`	all	3000	0.030	anchored	+0.040	+0.040	+0.048	none
tweets	`gpt-4o-mini`	all	3000	0.030	natural	+0.025	+0.025	+0.032	none
tweets	`gpt-4o-mini`	< 0.95	380	0.220	anchored	+0.040	+0.040	+0.104	none
tweets	`gpt-4o-mini`	< 0.95	380	0.220	natural	+0.012	+0.012	+0.076	none
tweets	`gpt-4o-mini`	< 0.90	273	0.274	anchored	+0.040	+0.040	+0.116	none
tweets	`gpt-4o-mini`	< 0.90	273	0.274	natural	+0.011	+0.011	+0.085	none
treebank	`gpt-4.1-mini`	all	3000	0.026	anchored	+0.040	+0.040	+0.047	none
treebank	`gpt-4.1-mini`	all	3000	0.026	natural	+0.031	+0.031	+0.038	none
treebank	`gpt-4.1-mini`	< 0.95	318	0.224	anchored	+0.040	+0.040	+0.117	none
treebank	`gpt-4.1-mini`	< 0.95	318	0.224	natural	+0.013	+0.013	+0.089	none
treebank	`gpt-4.1-mini`	< 0.90	241	0.271	anchored	+0.040	+0.040	+0.131	none
treebank	`gpt-4.1-mini`	< 0.90	241	0.271	natural	+0.012	+0.012	+0.101	none
treebank	`gpt-4o`	all	3000	0.031	anchored	+0.040	+0.040	+0.048	none
treebank	`gpt-4o`	all	3000	0.031	natural	+0.032	+0.032	+0.040	none
treebank	`gpt-4o`	< 0.95	374	0.232	anchored	+0.040	+0.040	+0.111	none
treebank	`gpt-4o`	< 0.95	374	0.232	natural	+0.014	+0.014	+0.082	none
treebank	`gpt-4o`	< 0.90	292	0.278	anchored	+0.040	+0.040	+0.120	none
treebank	`gpt-4o`	< 0.90	292	0.278	natural	+0.012	+0.012	+0.088	none
treebank	`gpt-4o-mini`	all	3000	0.022	anchored	+0.040	+0.040	+0.044	none
treebank	`gpt-4o-mini`	all	3000	0.022	natural	+0.032	+0.032	+0.036	none
treebank	`gpt-4o-mini`	< 0.95	273	0.229	anchored	+0.040	+0.039	+0.391	none
treebank	`gpt-4o-mini`	< 0.95	273	0.229	natural	+0.010	+0.010	+0.368	none
treebank	`gpt-4o-mini`	< 0.90	199	0.286	anchored	+0.040	+0.040	+0.112	none
treebank	`gpt-4o-mini`	< 0.90	199	0.286	natural	+0.010	+0.010	+0.080	none

P.12 FINITE SAMPLES

Table 22 executes the stored policies with fresh draws from each record's response vector, at 1000 records per arm and 300 replications, on the ambiguous subpool at temperature one. The estimator concentrates on the population value of the policy that produced it, and a conventional interval covers that value at its nominal rate while covering Δ^* far less often. Nothing about the sampling is unusual; the target has moved.

Table 21: Released effect over numerical accuracy optima, horizon five, at the 41 tested budgets in $[0, 2]$. Stopping is optimized separately by arm. Parenthesized budget pairs are rounded grid ranges including endpoints. A dagger marks an anchor not reached by bisection.

Δ^*	temp.	pool	design	Δ_1	effect range over optima	wrong sign budgets
0.10	1	all	anchored	$+0.1000$	$[+0.0937, +0.1067]$	none
0.10	1	all	natural	$+0.0676$	$[+0.0676, +0.0755]$	none
0.10	1	< 0.99	anchored	$+0.1000$	$[+0.0540, +0.1463]$	none
0.10	1	< 0.99	natural	$+0.0416$	$[+0.0416, +0.1030]$	none
0.10	1	< 0.95	anchored	$+0.1000$	$[+0.0329, +0.1658]$	none
0.10	1	< 0.95	natural	$+0.0415$	$[+0.0415, +0.1347]$	none
0.10	1	< 0.90	anchored	$+0.1000$	$[+0.0213, +0.1759]$	none
0.10	1	< 0.90	natural	$+0.0425$	$[+0.0425, +0.2117]$	none
0.10	1	< 0.80	anchored	$+0.1000$	$[+0.1000, +0.2787]$	none
0.10	1	< 0.80	natural	$+0.0384$	$[+0.0384, +0.1807]$	none
0.10	2	all	anchored	$+0.1000$	$[+0.0871, +0.1141]$	none
0.10	2	all	natural	$+0.0671$	$[+0.0671, +0.0837]$	none
0.10	2	< 0.99	anchored	$+0.2204^{\dagger}$	$[-0.2475, +0.2404]$	$(0.15, 1.10)$
0.10	2	< 0.99	natural	$+0.0392$	$[+0.0392, +0.1093]$	none
0.10	2	< 0.95	anchored	$+0.1000$	$[+0.0303, +0.1703]$	none
0.10	2	< 0.95	natural	$+0.0395$	$[+0.0395, +0.1366]$	none
0.10	2	< 0.90	anchored	$+0.1000$	$[+0.0211, +0.1858]$	none
0.10	2	< 0.90	natural	$+0.0386$	$[+0.0386, +0.1496]$	none
0.10	2	< 0.80	anchored	$+0.1000$	$[+0.1000, +0.2441]$	none
0.10	2	< 0.80	natural	$+0.0392$	$[+0.0392, +0.2342]$	none
0.04	1	all	anchored	$+0.0400$	$[+0.0340, +0.0427]$	none
0.04	1	all	natural	$+0.0270$	$[+0.0270, +0.0347]$	none
0.04	1	< 0.99	anchored	$+0.0400$	$[-0.0140, +0.0590]$	$(0.20, 0.30)$
0.04	1	< 0.99	natural	$+0.0166$	$[+0.0166, +0.0756]$	none
0.04	1	< 0.95	anchored	$+0.0400$	$[-0.0374, +0.0668]$	$(0.15, 0.35)$
0.04	1	< 0.95	natural	$+0.0166$	$[+0.0166, +0.1068]$	none
0.04	1	< 0.90	anchored	$+0.0400$	$[-0.0577, +0.0708]$	$(0.10, 0.35)$
0.04	1	< 0.90	natural	$+0.0170$	$[+0.0170, +0.1275]$	none
0.04	1	< 0.80	anchored	$+0.0400$	$[+0.0400, +0.1348]$	none
0.04	1	< 0.80	natural	$+0.0154$	$[+0.0154, +0.1076]$	none
0.04	2	all	anchored	$+0.0400$	$[+0.0273, +0.0457]$	none
0.04	2	all	natural	$+0.0269$	$[+0.0269, +0.0429]$	none
0.04	2	< 0.99	anchored	$+0.0400$	$[-0.0173, +0.0618]$	$(0.20, 0.30)$
0.04	2	< 0.99	natural	$+0.0157$	$[+0.0157, +0.0837]$	none
0.04	2	< 0.95	anchored	$+0.0400$	$[-0.0415, +0.0688]$	$(0.15, 0.35)$
0.04	2	< 0.95	natural	$+0.0158$	$[+0.0158, +0.1100]$	none
0.04	2	< 0.90	anchored	$+0.0400$	$[-0.0598, +0.0751]$	$(0.10, 0.35)$
0.04	2	< 0.90	natural	$+0.0154$	$[+0.0154, +0.1225]$	none
0.04	2	< 0.80	anchored	$+0.0400$	$[+0.0070, +0.1741]$	none
0.04	2	< 0.80	natural	$+0.0157$	$[+0.0157, +0.1441]$	none
0.02	1	all	anchored	$+0.0200$	$[+0.0133, +0.0214]$	none
0.02	1	all	natural	$+0.0135$	$[+0.0135, +0.0211]$	none
0.02	1	< 0.99	anchored	$+0.0200$	$[-0.0372, +0.0297]$	$(0.10, 0.75)$
0.02	1	< 0.99	natural	$+0.0083$	$[+0.0083, +0.0664]$	none
0.02	1	< 0.95	anchored	$+0.0200$	$[-0.0630, +0.0335]$	$(0.10, 0.70)$
0.02	1	< 0.95	natural	$+0.0083$	$[+0.0083, +0.0975]$	none
0.02	1	< 0.90	anchored	$-0.3732^{\dagger}$	$[-0.3732, +0.1355]$	$(0.00, 1.35)$
0.02	1	< 0.90	natural	$+0.0085$	$[+0.0085, +0.1175]$	none
0.02	1	< 0.80	anchored	$+0.0200$	$[+0.0200, +0.1120]$	none
0.02	1	< 0.80	natural	$+0.0077$	$[+0.0077, +0.0984]$	none
0.02	2	all	anchored	$+0.0200$	$[+0.0058, +0.0228]$	none
0.02	2	all	natural	$+0.0134$	$[+0.0134, +0.0293]$	none
0.02	2	< 0.99	anchored	$+0.0200$	$[-0.0421, +0.0314]$	$(0.10, 0.70)$
0.02	2	< 0.99	natural	$+0.0078$	$[+0.0078, +0.0752]$	none
0.02	2	< 0.95	anchored	$+0.0200$	$[-0.0670, +0.0352]$	$(0.10, 0.70)$
0.02	2	< 0.95	natural	$+0.0079$	$[+0.0079, +0.1011]$	none
0.02	2	< 0.90	anchored	$+0.0200$	$[-0.0823, +0.0386]$	$(0.05, 0.75)$
0.02	2	< 0.90	natural	$+0.0077$	$[+0.0077, +0.1134]$	none
0.02	2	< 0.80	anchored	$+0.0200$	$[-0.0255, +0.1508]$	$(0.60, 0.70)$
0.02	2	< 0.80	natural	$+0.0078$	$[+0.0078, +0.1348]$	none

Table 22: Monte Carlo on the 239 ambiguous records at temperature one, anchored design, $\Delta^* = .04$, $n = 1000$ per arm, 300 replications. Accuracy is the realized share of released labels equal to the truth.

policy	population mean	est.	s.d.	$\Pr(\text{est.} < 0)$	covers Δ^*	accuracy
release the first answer	$+0.0400$	$+0.0404$	0.020	0.03	0.95	0.708
budget $b = 0.30$	-0.0190	-0.0198	0.022	0.81	0.20	0.741
budget $b = 0.60$	$+0.0254$	$+0.0275$	0.022	0.09	0.93	0.771

P.13 Reproduction

The supplementary archive contains the annotation files, response catalogs, recorded answers, and code that reproduces the reported numbers and figures in Section 7 from these archived data without further language model calls. It includes the scripts for data collection and the repeated sampling check, policy replay code, and code for the seeded training/test splits and pipeline evaluations of Section P.7. Rebuilding one catalog requires 3000 calls per corpus and model. The same code path computes the four type results of Sections 2 and 3, and a unit test checks it against their exact rational values, including $A^* = .7744$ at (10) and the first revision masses .072.

P.14 Limitations

The formal information set includes treatment, whereas an anonymous measurement supplier may not observe it. The measured shared alphabets allow a common stopping rule, which differs from optimizing each arm separately. The comparison above therefore identifies an information requirement of the measured mechanism.

Independent draws from a fixed catalog satisfy the simulation model. The recorded separate requests of Section P.5 are consistent with independent draws from a record's mean first token distribution, of which the catalog is one draw, while the completions of a single request are not independent; the recorded answers cover four ambiguous pools at temperature one, and replays resample them without replacement.

Section P.7 fixes decoder, tilt, price and policy on training records, but the held out labels still define the arms and score the release, both halves come from one pool, and the sign of a fitted arm specific policy varies with the halving.

Finite budget grids report numerical optima at their sampled points. They do not certify continuous intervals or exact endpoints between points, and the accuracy tolerance is part of the numerical calculation.

The distortion concerns released hard labels. An interface that exports the true posterior conditional on treatment and all information available at release preserves the population contrast by the tower property, even when release is adaptive. A posterior from the pooled population alone need not.

Catalog repair requires the specified response vectors. Its sensitivity to misspecification grows with the production horizon, and the support contamination example in Section O.1 gives a nonvanishing bias under vanishing rowwise perturbations.

Empirical coverage conditions on the stored catalog and the selected policy. Records are sampled with replacement from the design weights; the Wald interval uses the two arm sample variances. Catalog estimation, policy learning and calibration uncertainty are not included.