Incentivized Sequential Testing

Anonymous authors
Paper under double-blind review

Abstract

We study sequential hypothesis testing when data collection is delegated to a risk-neutral agent who pays for each observation, receives a fixed reward upon rejection, and may quit at any time. The principal chooses an e-process to maximize the probability of rejection before the agent quits. We show that classical growth-optimal e-processes need not be optimal for this objective: tempering the process trades a small loss in evidence growth for reduced fluctuations, which lowers the risk of abandonment and can increase the probability of success. We characterize the abandonment exponent and its unique maximizer within the Gaussian tempered family, and extend the strict improvement to composite nulls under a local stability condition.

1 Introduction

Consider a firm seeking regulatory approval for a product. Before validation begins, the firm and the regulator agree on a class $\mathcal{P}_0$ of null distributions and a particular alternative distribution Q for the validation outcomes. The firm privately knows that Q is the data-generating distribution, but the regulator cannot act on this claim without statistically valid evidence. The regulator therefore commits to a sequential evidence-accumulation rule that controls type-I error under every distribution in $\mathcal{P}_0$, and delegates data collection to the firm. The firm purchases observations one at a time, bears a cost for each, and receives a reward if the accumulated evidence crosses the rejection threshold. After each observation, it may either continue collecting data or abandon the attempt. The regulator's problem is to choose the evidence-accumulation rule that maximizes the probability that the firm produces sufficient evidence to reject the composite null before quitting.

This can be viewed as a principal–agent problem: the regulator is the principal, who wants the null rejected and bears no sampling cost, and the firm is the agent, who bears a per-sample cost and receives a reward upon rejection. Two structural features distinguish the setting from standard sequential analysis. First, the principal and agent have conflicting incentives: the agent maximizes expected net payoff, whereas the principal seeks only rejection. Second, the agent's stopping rule is endogenous. The choice to continue data collection depends on the trajectory of accumulated evidence rather than its average drift alone. The agent quits when the value of collecting one further observation and then behaving optimally is nonpositive. The principal therefore cannot assume complete data collection; the test must be designed to sustain the agent's participation.

We frame this design problem in the language of e-processes: crossing the level $1/\alpha$ yields a level α anytime-valid test (Ramdas & Wang, 2025). The principal's problem becomes: which e-process should it commit to? The classical theory gives a clear answer: the likelihood ratio process maximizes the expected log-growth rate of the evidence (Grünwald et al., 2024). This log-growth optimality remains valid when the agent can quit, but it does not optimize the principal's different objective: the principal cares not only about how fast evidence grows on average, but about the probability that the process reaches the rejection threshold before the agent quits. These two objectives need not agree.

In this paper we show that the growth-optimal e-process can be suboptimal when a cost-sensitive agent may quit. In the Gaussian model, and for composite nulls satisfying a local stability condition, slight tempering loses only second-order log growth while improving the abandonment exponent to first order. We characterize the exponent-optimal Gaussian

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.

parameter. The agent is risk-neutral: the downside sensitivity arises from optimal abandonment, rather than an imposed risk preference. The result concerns the probability of rejection before that endogenous decision, subject to uniform anytime validity.

2 Related Work

Our work sits at the intersection of anytime-valid inference and the economics of statistical testing. On the statistical side, e-values and e-processes have become a standard tool for sequential testing, with admissibility results connecting anytime-valid inference to nonnegative martingales (Ramdas et al., 2020; 2023). Within this theory, the likelihood ratio is the growth-rate optimal e-variable for a simple null against a simple alternative, maximizing the expected log-growth rate of the evidence (Grünwald et al., 2024; Shafer, 2021); recent work extends this log-optimality to composite nulls and general classes of e-processes (Waudby-Smith et al., 2025; Shekhar, 2025). These results optimize evidence accumulation when continuation is controlled by the statistician or specified as part of the testing procedure. Our setting instead separates the test designer from the data collector: continuation is chosen by a cost-bearing agent in response to the realized evidence path.

On the economic side, statistical testing with strategic data producers has been studied through approval thresholds, participation, contracts, and risk preferences (Tetenov, 2016; Bates et al., 2022; Min, 2023; Shi et al., 2024; Chen et al., 2026). Bates et al. (2022) also consider a multi-round setting in which the agent may stop and adapt its data acquisition. Our focus is different: the reward is fixed, and the principal chooses among valid e-processes to maximize the probability of rejection before the agent quits.

Dynamic delegated experimentation already links approval rules to an agent's option to quit. Gerardi & Maestri (2012) study incentive provision with unobservable effort and signals. McClellan (2022) characterizes history-dependent approval mechanisms that maximize a regulator's Bayesian expected payoff, with state-dependent approval payoffs and experimentation costs. Their players learn about project quality from a public signal, with an extension allowing the agent a private initial signal. Recent work also studies subsidies and history-dependent testing thresholds (Artola Velasco et al., 2026; Nguyen et al., 2026). Our benchmark instead fixes the reward and an alternative Q known to the agent, and maximizes rejection probability under Q subject to anytime type-I error control under every null distribution. In the Gaussian model, tempering is exactly equivalent to a time-dependent threshold on the original likelihood ratio (Remark 4.1); the distinction therefore lies in the objective, information structure, and validity constraint.

Trading evidence growth against downward excursions is also related to fractional Kelly betting and risk-constrained Kelly gambling. Busseti et al. (2016) maximize long-run log growth subject to a bound on the probability of wealth falling below a prescribed level. Here the abandonment boundary is determined by a risk-neutral agent's optimal response to each testing rule. Our contribution is to quantify the resulting failure exponent and show that slight tempering improves it over the log-growth-optimal e-process while preserving anytime validity.

3 Problem Setup

We consider a sequential testing setup in which an agent collects independent and identically distributed observations $X_1, X_2, \ldots$ from a distribution F, and the null hypothesis is that $F \in \mathcal{P}_0$ while the alternative is that $F = Q$. The principal wishes to reject the null at level α, and it delegates the data collection to the agent. The principal's testing rule is specified through an e-process.

Definition 3.1. A nonnegative process $E = (E_t)_{t\in\mathbb{N}_0}$, adapted to $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$ with $\mathcal{F}_0$ trivial and $E_0 = 1$, is an e-process if $\mathbb{E}_F[E_\tau] \leq 1$ for every almost surely finite $(\mathcal{F}_t)$-stopping time τ and every $F \in \mathcal{P}_0$.

The connection to sequential testing is the following: the test rejects the null at the first time the e-process crosses the threshold $1/\alpha$, and it can be shown that the probability of

ever rejecting under the null is at most α (Ramdas & Wang, 2025). We study the principal's choice within this class of anytime-valid testing rules.

The agent bears a cost $c > 0$ for each observation it collects and receives a reward $R > 0$ if the null is rejected. For a given e-process E, define the rejection time

$$\tau_E = \inf\{t \in \mathbb{N}_0 : E_t \geq 1/\alpha\},$$

with the convention that $\inf \varnothing = \infty$. We assume that the agent knows $F = Q$.

Remark 3.2. The assumption that the agent knows $F = Q$ is an informed-agent benchmark. The sequential test is not intended to teach the agent whether the null is false. Rather, it allows the agent to produce verifiable evidence for a principal who does not observe the agent's private information. This type of information asymmetry is standard in economic models of statistical testing and approval. Tetenov (2016) studies a perfectly informed proponent who knows the quality of its innovation, whereas the regulator does not. Bates et al. (2022) likewise assume that the agent knows the quality, or type, of its product while the principal lacks this information. In a sequential setting, McClellan (2022) allows the experimenting agent to possess private information about project quality before experimentation begins.

At each time $t < \tau_E$, after observing $X_1, \ldots, X_t$ the agent may either quit or collect one further observation X_{t+1}. A quitting rule is therefore an $(\mathcal{F}_t)$-stopping time σ. Under this rule, the agent's payoff is

$$U_E(\sigma) = R\,\mathbf{1}\{\tau_E \leq \sigma,\ \tau_E < \infty\} - c\,(\tau_E \wedge \sigma).$$

Thus, if the agent quits before rejection, it receives no reward and pays only the costs incurred before quitting. If rejection occurs first, it receives R and pays for the observations collected up to τ_E. We take rejection to have precedence in the event $\tau_E = \sigma$.

Let $\mathcal{T}$ denote the set of $(\mathcal{F}_t)$-stopping times taking values in $\mathbb{N}_0 \cup \{\infty\}$. The agent's ex ante value under the e-process E is

$$V_0^E = \sup_{\sigma \in \mathcal{T}} \mathbb{E}_Q[R\,\mathbf{1}\{\tau_E \leq \sigma,\ \tau_E < \infty\} - c\,(\tau_E \wedge \sigma)]\,.$$

The supremum may be restricted to stopping times satisfying $\mathbb{E}_Q(\tau_E \wedge \sigma) < \infty$, since any other rule has expected payoff equal to $-\infty$. Immediate quitting, corresponding to $\sigma = 0$, is always feasible, and hence $V_0^E \geq 0$.

More generally, let V_t^E denote the agent's maximal expected payoff from time t onward, excluding costs already incurred. On the event $\{t < \tau_E\}$, this value is

$$V_t^E = \operatorname*{ess\,sup}_{\substack{\sigma \in \mathcal{T} \\ \sigma \geq t}} \mathbb{E}_Q[R\,\mathbf{1}\{\tau_E \leq \sigma,\ \tau_E < \infty\} - c\{(\tau_E \wedge \sigma) - t\}\,|\,\mathcal{F}_t]\,,$$

and $V_t^E = R$ on $\{t \geq \tau_E\}$. The value process satisfies the Bellman equation

$$V_t^E = \max\left\{0,\ -c + \mathbb{E}_Q\left(V_{t+1}^E \mid \mathcal{F}_t\right)\right\}, \qquad t < \tau_E.$$

The quantity $C_t^E = -c + \mathbb{E}_Q\left(V_{t+1}^E \mid \mathcal{F}_t\right)$ is the value of collecting one further observation and then behaving optimally. The agent continues when $C_t^E > 0$ and quits when $C_t^E < 0$. To resolve indifference, we adopt the convention that the agent quits when $C_t^E = 0$. The resulting quitting time is

$$\sigma_E^\star = \inf\left\{t \geq 0 : t < \tau_E,\ C_t^E \leq 0\right\}.$$

Here $\inf \varnothing = \infty$: if the agent never quits before rejection, sampling ends at τ_E; in general it ends at $\tau_E \wedge \sigma_E^\star$. The rejection-precedence convention specifies payoffs for arbitrary $\sigma \in \mathcal{T}$. Anticipating the agent's optimal response, the principal's success probability under Q is

$$\pi_Q(E) = \mathbb{P}_Q(\tau_E \leq \sigma_E^\star,\ \tau_E < \infty)\,.$$

The principal does not bear the sampling cost and therefore cares only about whether rejection occurs before the agent quits. Given a class $\mathfrak{E}$ of e-processes, the principal's problem is to find $E \in \mathfrak{E}$ so that $\pi_Q(E)$ is maximized.

4 Suboptimality of the Growth-Optimal E-Process

4.1 The Gaussian case

A central result in the theory of e-processes is that, for testing a simple null $\mathcal{P}_0 = \{P\}$ against a simple alternative Q, the likelihood ratio process

$$E_t := \prod_{i=1}^{t} \frac{dQ}{dP}(X_i) = \prod_{i=1}^{t} L(X_i)$$

is the growth-rate optimal e-process: it maximizes $\mathbb{E}_Q[\log E_t]$ among all e-processes (Shafer, 2021; Grünwald et al., 2024).

A principal who is unconcerned with the agent's quitting behavior, and who simply wants evidence to accumulate as fast as possible, would therefore choose the likelihood ratio e-process and stop there. The question we pose is whether this classical optimum remains the best choice when the principal's objective is not to maximize the rate of evidence accumulation, but to maximize the success probability $\pi_Q(\cdot)$. We show that it need not: the likelihood ratio e-process is suboptimal in the asymptotic regime below.

We first demonstrate this in the Gaussian location model. Let $P = N(0,1)$ and $Q = N(\mu,1)$ with $\mu > 0$, so that the likelihood ratio is $L(X) = \exp(\mu X - \mu^2/2)$. To state the result we restrict our attention to the class of tempered likelihood ratio e-processes $\mathfrak{E} = \{E_t^{(\lambda)}, \lambda \in (0,1]\}$, where

$$E_t^{(\lambda)} = \prod_{i=1}^{t} \frac{L(X_i)^\lambda}{\mathbb{E}_P[L(X)^\lambda]}, \qquad E_0^{(\lambda)} = 1.$$

The parameter λ controls the tempering: $\lambda = 1$ recovers the likelihood ratio e-process, while $\lambda < 1$ yields a process with slower growth but smaller variance.

Remark 4.1. In the Gaussian model, the tempered process has a simple interpretation. Since

$$\frac{L(x)^\lambda}{\mathbb{E}_P\{L(X)^\lambda\}} = \exp\left\{\lambda\mu x - \frac{\lambda^2\mu^2}{2}\right\} = \frac{\mathrm{d}N(\lambda\mu,1)}{\mathrm{d}N(0,1)}(x),$$

$E^{(\lambda)}$ is itself a likelihood-ratio process, but for the smaller design alternative $N(\lambda\mu,1)$. Thus tempering can be viewed as designing the test for a more conservative effect size than the one believed by the agent. The identity $\log E_n^{(\lambda)} = \lambda \log E_n^{(1)} + \lambda(1-\lambda)\mu^2 n/2$ shows that rejection is equivalent to

$$\log E_n^{(1)} \geq \frac{\log(1/\alpha)}{\lambda} - \frac{(1-\lambda)\mu^2 n}{2}.$$

Thus tempering is also a time-dependent threshold on the original likelihood ratio.

Under P, each $E^{(\lambda)}$ is an e-process. Let

$$S_t^{(\lambda)} = \log E_t^{(\lambda)}, \qquad U_\alpha = -\log\alpha,$$

and define $\tau_{\lambda,\alpha} = \inf\left\{t \geq 0 : S_t^{(\lambda)} \geq U_\alpha\right\}$. For a reward–cost ratio $\rho = R/c$, let $\sigma^\star_{\lambda,\alpha,\rho}$ be the minimal Q-optimal quitting time defined by the agent's stopping problem in Section 3. The corresponding success probability is

$$\pi_{\alpha,\rho}(\lambda) = \mathbb{P}_Q\left(\tau_{\lambda,\alpha} \leq \sigma^\star_{\lambda,\alpha,\rho},\ \tau_{\lambda,\alpha} < \infty\right).$$

We study the asymptotic regime in which the significance level tends to zero. Accordingly, we consider a family of decision problems indexed by α, allowing the reward–cost ratio $\rho = \rho_\alpha$ to vary with α. Without quitting, the untempered test has expected sampling cost asymptotic to $2cU_\alpha/\mu^2$; thus $r > 1$ means that the reward asymptotically exceeds this cost.

Theorem 4.2. Fix $\mu > 0$ and suppose $\rho_\alpha\mu^2/(2U_\alpha) \to r > 1$ as $\alpha \downarrow 0$. For every fixed $\lambda \in (0,1]$ with $r(2\lambda - \lambda^2) > 1$, the agent initially continues for all sufficiently small α, and

$$-\frac{\log\{1 - \pi_{\alpha,\rho_\alpha}(\lambda)\}}{U_\alpha} \longrightarrow I_r(\lambda) = \left(\frac{2}{\lambda} - 1\right)\{r(2\lambda - \lambda^2) - 1\}. \tag{1}$$

Since $I_r'(1) = -2(r-1) < 0$, there exists $\varepsilon > 0$ such that every fixed $\lambda \in (1-\varepsilon, 1)$ gives $\pi_{\alpha,\rho_\alpha}(\lambda) > \pi_{\alpha,\rho_\alpha}(1)$ for all sufficiently small α.

Corollary 4.3 (Exponent-optimal Gaussian tempering). The exponent in Theorem 4.2 has a unique maximizer over its participation region. It is the unique $\lambda_r \in (0,1)$ satisfying

$$r\lambda_r^2(2-\lambda_r) = 1.$$

For every other fixed participating λ, $\{1-\pi_{\alpha,\rho_\alpha}(\lambda_r)\}/\{1-\pi_{\alpha,\rho_\alpha}(\lambda)\} \to 0$.

The proof is in Appendix A.1. This optimizes the limiting exponent; the pointwise asymptotics do not establish convergence of finite-α maximizers or guarantee participation at λ_r for a specified finite α. Check that the initial Bellman continuation value is strictly positive for the specified reward and sampling cost.

4.2 Beyond the Gaussian case: tempered oracle Donsker–Varadhan (DV) e-processes

Theorem 4.2 concerns a simple Gaussian null and a simple Gaussian alternative. We now allow the null to be composite while retaining a fixed alternative distribution Q, and show that the suboptimality of the classically optimal e-process extends beyond the Gaussian likelihood-ratio setting.

Suppose that Q and every $P_0 \in \mathcal{P}_0$ admit densities with respect to a common measure, and that the null class $\mathcal{P}_0$ is convex. Define $\gamma^* = \inf_{P_0\in\mathcal{P}_0} d_{\mathrm{KL}}(Q\|P_0)$, and assume that $0 < \gamma^* < \infty$ and that the infimum is attained at some $P_0^* \in \mathcal{P}_0$. Let

$$L^*(x) = \frac{\mathrm{d}Q}{\mathrm{d}P_0^*}(x), \qquad \ell^*(x) = \log L^*(x).$$

For a measurable function f, define $\psi_0(f) = \sup_{P_0\in\mathcal{P}_0} \log \mathbb{E}_{P_0}\{\exp(f(X))\}$.

Proposition 4.4. For $\lambda \in (0,1]$ such that $\psi_0(\lambda\ell^*) < \infty$, the following is an e-process:

$$W_n^{(\lambda)} = \prod_{i=1}^{n} \exp\left\{\lambda\ell^*(X_i) - \psi_0(\lambda\ell^*)\right\}, \qquad W_0^{(\lambda)} = 1.$$

Moreover, $\psi_0(\ell^*) = 0$, and $W^{(1)}$ is log-growth optimal: for every e-process $E = (E_n)_{n\geq 0}$ for $\mathcal{P}_0$ and every $n \geq 1$, $\mathbb{E}_Q\{\log E_n\} \leq \mathbb{E}_Q\{\log W_n^{(1)}\} = n\gamma^*$.

Proposition 4.4 shows that $W^{(1)}$ is optimal under the classical expected log-growth criterion. It does not account for the possibility that the agent quits before the rejection boundary is reached. Following Shekhar (2025), we call $W^{(1)}$ the oracle Donsker–Varadhan (DV) e-process. To study local tempering, define

$$A(t) = \log \mathbb{E}_{P_0^*}\left[\exp\{t\ell^*(X)\}\right].$$

To state our next main theorem we need the following assumption.

Assumption 4.5. We assume that Q and P_0^* are mutually absolutely continuous, that A is finite on an open interval containing $[0,1]$, and that there exists $\eta > 0$ such that

$$\psi_0(\lambda\ell^*) = A(\lambda), \qquad \lambda \in [1-\eta, 1]. \tag{2}$$

Remark 4.6. Condition (2) is equivalent to

$$\mathbb{E}_{P_0}\left[\{L^*(X)\}^\lambda\right] \leq \mathbb{E}_{P_0^*}\left[\{L^*(X)\}^\lambda\right], \qquad P_0 \in \mathcal{P}_0, \quad \lambda \in [1-\eta, 1]. \tag{3}$$

At $\lambda = 1$, this inequality follows from Proposition 4.4. Requiring it for λ in a left neighborhood of one is an additional local stability condition: it rules out the possibility that a different null distribution determines the normalizing constant immediately after the likelihood ratio is tempered. This condition is satisfied in many one-sided models with a monotone likelihood ratio. For example, $\mathcal{P}_0 = \mathrm{conv}\{N(\theta,1) : \theta \leq 0\}$, $Q = N(\mu,1)$, $\mu > 0$. Additional examples are given in Appendix A.4. A three-point counterexample in Appendix A.5 shows that the condition is substantive: without it, the normalizer can change immediately under tempering, and the improvement can reverse.

For $\lambda \in [1-\eta, 1]$, define the rejection time $\tau_\alpha^{(\lambda)} = \inf\left\{n \geq 0 : W_n^{(\lambda)} \geq 1/\alpha\right\}$. Let $\sigma_{\alpha,\rho}^{\star,(\lambda)}$ denote the agent's Q-optimal quitting time when the reward–cost ratio is $\rho = R/c$, and define $\pi_{\alpha,\rho}^{\mathrm{DV}}(\lambda) = \mathbb{P}_Q\left(\tau_\alpha^{(\lambda)} \leq \sigma_{\alpha,\rho}^{\star,(\lambda)},\ \tau_\alpha^{(\lambda)} < \infty\right)$.

Theorem 4.7. Suppose Assumption 4.5 holds, let $U_\alpha = \log(1/\alpha)$, and assume $\rho_\alpha \gamma^* / U_\alpha \to r > 1$. Write $g(\lambda) = \lambda\gamma^* - A(\lambda)$ and $\delta^* = d_{\mathrm{KL}}(P_0^* \| Q)$. For λ sufficiently close to one from the left, let $\kappa(\lambda) > 0$ be the unique positive root of $A(1-\kappa\lambda) + \kappa A(\lambda) = 0$, with $\kappa(1) = 1$. For remaining distance $x = U_\alpha - \log W_n^{(\lambda)} > 0$, let $a_{\lambda,\rho} = \inf\{x > 0 : C_\rho(x) \le 0\}$, where $C_\rho(x)$ is the continuation value with unit sampling cost (Appendix A.3). This stopping boundary satisfies $a_{\lambda,\rho}/\rho \to g(\lambda)$ as $\rho \to \infty$. For each such fixed λ, the agent initially continues for all sufficiently small α, and

$$-\frac{\log\{1 - \pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(\lambda)\}}{U_\alpha} \longrightarrow I_r(\lambda) = \kappa(\lambda)\left\{\frac{r g(\lambda)}{\gamma^*} - 1\right\}. \tag{4}$$

Moreover,

$$g'(1) = 0, \qquad I_r(1) = r - 1, \qquad I_r'(1) = -(r-1)\left(1 + \frac{\gamma^*}{\delta^*}\right) < 0.$$

Here $I_r'(1)$ denotes the left derivative. Consequently, every fixed $\lambda < 1$ sufficiently close to one strictly improves success probability over $W^{(1)}$ for all sufficiently small α.

Under local stability, the loss in expected log growth is second order in $1-\lambda$, while the gain in abandonment exponent is first order. The stopping boundary is endogenous and changes with the test.

The reward scaling maintains participation as the evidence requirement increases. For a fixed ρ and any fixed process covered above, the stopping boundary is finite; eventually U_α exceeds it and the agent quits immediately. Under the stated scaling, both success probabilities tend to one. The improvement is therefore in the decay of failure probability, not a nonvanishing absolute power gap.

5 Numerical Experiments

Theorems 4.2 and 4.7 are asymptotic results. We complement them with numerical experiments to illustrate the finite-sample behavior of the proposed tempering strategy. Reported optimizers maximize over finite parameter grids; plotted curves connect raw computed values without smoothing. We first revisit the Gaussian model and examine how the principal's success probability and the optimal tempering parameter depend on the reward–cost ratio and the significance level. We then consider a non-Gaussian example with a composite null to illustrate the mechanism underlying Theorem 4.7.

5.1 Gaussian model

We first return to the Gaussian setting of Theorem 4.2. Let

$$P = N(0,1), \qquad Q = N(\mu, 1),$$

and consider the tempered likelihood-ratio family

$$E_t^{(\lambda)} = \prod_{i=1}^{t} \frac{L(X_i)^\lambda}{\mathbb{E}_P\{L(X)^\lambda\}}, \qquad \lambda \in (0,1], \quad \text{where } L(x) = \frac{\mathrm{d}Q}{\mathrm{d}P}(x).$$

The case $\lambda = 1$ is the usual likelihood-ratio process, while smaller values of λ reduce both its drift and its fluctuations.

For each λ, the agent's quitting rule is obtained by solving the Bellman equation numerically. Given this optimal stopping rule, we then compute the probability $\pi_{\alpha,\rho}(\lambda)$ that the e-process reaches the rejection threshold before the agent quits. We parameterize the reward–cost ratio through $r = \rho\mu^2/\{2\log(1/\alpha)\}$, which is the normalized quantity appearing in Theorem 4.2. Values $r > 1$ correspond to the asymptotic participation regime of the theorem.

Figure 1 reports the results for $\mu = 1$ and $\alpha = 10^{-3}$. Panel (a) plots $\pi_{\alpha,\rho}(\lambda)$ for three values of r. For sufficiently small λ, evidence accumulates too slowly to justify the sampling

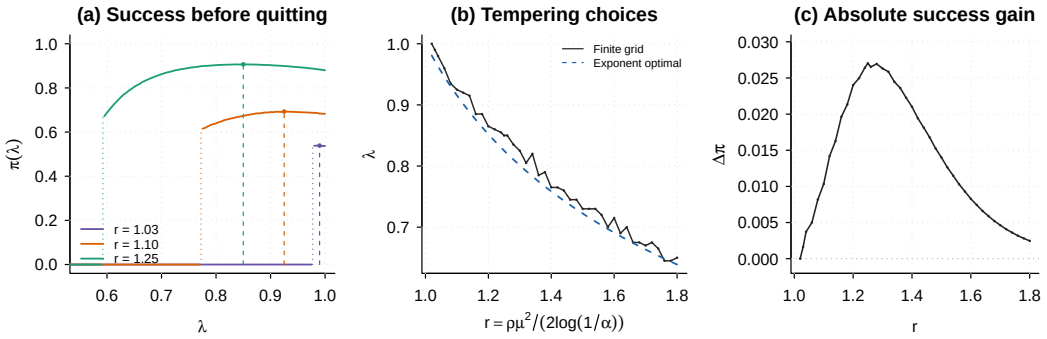

Figure 1: Numerical illustration of Theorem 4.2 for $\mu = 1$ and $\alpha = 10^{-3}$, where $r = \rho\mu^2/\{2\log(1/\alpha)\}$. Panel (a) shows the success probability $\pi(\lambda)$ for selected values of r; dotted and dashed vertical lines mark the participation boundaries and the maximizing values $\widehat{\lambda}$, respectively. Panel (b) compares the finite-grid optimizer $\widehat{\lambda}$ with the exponent-optimal reference λ_r (dashed), and Panel (c) shows the gain $\pi(\widehat{\lambda}) - \pi(1)$.

cost, so the agent quits immediately and the success probability is zero. The dotted vertical segments mark the participation boundaries. Once the agent participates, each curve is maximized at a value $\widehat{\lambda} < 1$, consistent with Theorem 4.2. Thus, although $\lambda = 1$ maximizes expected log-growth, a tempered process can produce a larger probability of rejection before abandonment.

Panel (b) compares the finite-grid tempering parameter $\widehat{\lambda}$ with the exponent-optimal reference λ_r from Corollary 4.3. Both generally decrease as r increases. The reference is asymptotic: near the participation boundary it can induce immediate quitting at the displayed finite significance level. A larger reward makes the agent more tolerant of a long experiment, allowing the principal to place relatively more weight on reducing adverse fluctuations. Panel (c) reports the corresponding gain $\pi_{\alpha,\rho}(\widehat{\lambda}) - \pi_{\alpha,\rho}(1)$. The gain is largest at intermediate values of r. Near the participation boundary, the set of viable tempering parameters is limited, whereas for large r, both the tempered and untempered processes succeed with probability close to one.

Remark 5.1. The irregularity in Panels (b) and (c) reflects numerical discretization: the stopping boundary is represented on a finite log-evidence grid and $\widehat{\lambda}$ is selected from a finite grid of tempering parameters. By contrast, the jumps in Panel (a) at the participation boundaries are genuine consequences of the agent's optimal stopping decision.

5.2 Asymptotic behavior

We fix $\mu = 1$ and $r = 1.25$, compare $\lambda = 0.9$, λ_r, and $\lambda = 1$, vary α, and set $\rho_\alpha = 2r\log(1/\alpha)/\mu^2$. Theorem 4.2 gives $q_{\alpha,\lambda} := 1 - \pi_{\alpha,\rho_\alpha}(\lambda) = \alpha^{I_r(\lambda)+o(1)}$. Here $I_{1.25}(1) = 0.25$, while $I_{1.25}(0.9) \approx 0.290$, so $q_{\alpha,0.9}/q_{\alpha,1} \to 0$. Corollary 4.3 gives $\lambda_{1.25} \approx 0.825$ as the exponent-optimal benchmark.

Figure 2(a) shows the absolute success gains for both tempered choices; they are positive on the displayed range but not monotone. Both success probabilities tend to one, so both gains eventually vanish. Panel (b) compares the computed failure exponents with their limiting values. Finite-threshold gaps reflect boundary and overshoot effects; the theorem gives a pointwise limit, without a quantitative convergence rate. Appendix B.1 supplements these illustrations with reproducible grid-refinement checks and sensitivity to signal strength, comparing the likelihood ratio, the exponent-optimal parameter, and a finite-threshold grid optimizer. It reports both absolute gains and relative reductions in failure probability. At $\mu = 1$, $r = 1.25$, and $\alpha = 10^{-3}$, the finest-grid estimates are 0.8800 for the likelihood ratio, 0.9055 at λ_r, and 0.9063 at the grid optimizer. The exponent-optimal choice thus

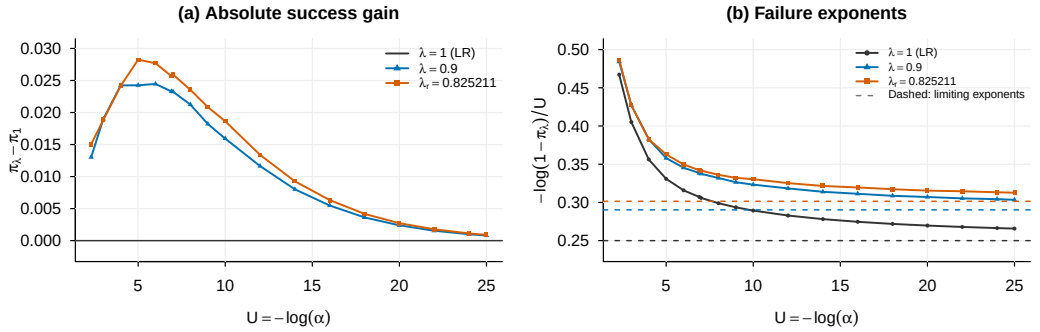

Figure 2: Asymptotic behavior in the Gaussian model with $\mu = 1$ and $r = 1.25$, comparing $\lambda = 0.9$, the exponent-optimal $\lambda_r \approx 0.825211$, and $\lambda = 1$. Panel (a) shows absolute success gains over the likelihood ratio as $U_\alpha = -\log \alpha$ increases; $\lambda = 1$ is the zero-gain baseline. Panel (b) shows failure exponents computed from directly solved failure probabilities; dashed lines mark the respective limiting exponents.

gains about 2.55 percentage points of success probability and reduces failure probability by 21.3%.

5.3 Composite-null Bernoulli model

We next illustrate Theorem 4.7 in a non-Gaussian setting with a composite null. Consider

$$\mathcal{P}_0 = \{\text{Bernoulli}(p) : 0 \le p \le p_0\}, \qquad Q = \text{Bernoulli}(q), \qquad 0 < p_0 < q < 1.$$

The reverse information projection is $P_0^* = \text{Bernoulli}(p_0)$, and hence

$$L^*(x) = \left(\frac{q}{p_0}\right)^x \left(\frac{1-q}{1-p_0}\right)^{1-x}, \qquad x \in \{0, 1\}.$$

For every $\lambda > 0$,

$$\sup_{P_0 \in \mathcal{P}_0} \mathbb{E}_{P_0}\{L^*(X)^\lambda\} = \mathbb{E}_{P_0^*}\{L^*(X)^\lambda\},$$

so Assumption 4.5 holds globally in this example. The tempered oracle e-process therefore has one-step log-increment

$$Y^{(\lambda)} = \lambda \log L^*(X) - \log \mathbb{E}_{P_0^*}\{L^*(X)^\lambda\}.$$

As in the Gaussian experiment, we normalize the reward–cost ratio by the first-order participation scale. Let

$$\gamma^* = d_{\mathrm{KL}}\{\text{Bernoulli}(q) \| \text{Bernoulli}(p_0)\}, \qquad r = \frac{\rho\gamma^*}{\log(1/\alpha)}.$$

For each pair (r, λ), we solve the stopping problem by backward recursion over the exact Bernoulli count states, without spatial interpolation. Lower and upper value functions bound the infinite-horizon response; Appendix B describes the resulting success-probability bounds.

Figure 3 reports the results for $p_0 = 0.3$, $q = 0.5$, and $\alpha = 10^{-3}$. Strong tempering can make participation unattractive, while the best participating choice on the displayed parameter grid is below one. At $r = 1.3$, $\widehat{\lambda} = 0.805$ gives success probability 0.964667, compared with 0.944162 for the likelihood ratio, a gain of about 2.05 percentage points. These are comparisons on finite λ-grids; they do not identify the optimizer over all real-valued tempering parameters.

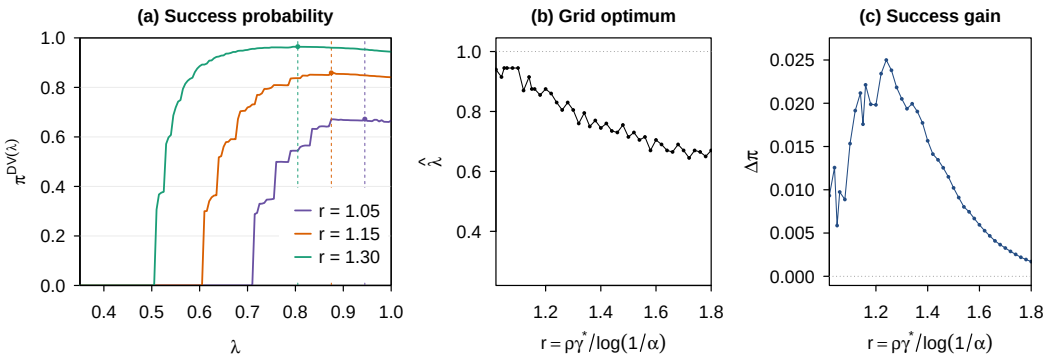

Figure 3: Count-state calculation for the Bernoulli model with $p_0 = 0.3$, $q = 0.5$, and $\alpha = 10^{-3}$, where $r = \rho\gamma^*/\log(1/\alpha)$. Panel (a) shows the success probability $\pi^{\mathrm{DV}}(\lambda)$ for selected values of r; dashed vertical lines mark the maximizers $\widehat{\lambda}$ on $\{0.25, 0.255, \ldots, 1\}$. Panel (b) shows $\widehat{\lambda}$ as a function of r, and Panel (c) shows the gain $\pi^{\mathrm{DV}}(\widehat{\lambda}) - \pi^{\mathrm{DV}}(1)$.

6 Discussion

Our results show that growth rate alone need not be the appropriate design criterion when continued data collection depends on the realized path of the evidence. In our setting, the principal faces a risk–growth trade-off: faster expected evidence accumulation is valuable, but large adverse fluctuations increase the probability that the agent abandons the experiment. Tempering improves performance by sacrificing a small amount of average growth in exchange for greater path stability.

The analysis has a natural connection to two-sided exit and ruin problems (Asmussen, 2010). On the logarithmic scale, the e-process starts between an upper rejection boundary and a lower quitting boundary. The principal succeeds if the upper boundary is reached first. A financial analogy is a trader seeking to reach a profit target before a stop-loss is triggered. The main difference from a conventional ruin or stop-loss model is that the lower boundary is endogenous: it is chosen through the agent's optimal-stopping problem and therefore depends on the reward, the sampling cost, and the law of the evidence increments.

More broadly, tempering is only one way to exploit this trade-off. We characterize the exponent-optimal parameter within the Gaussian tempered family. The e-process that maximizes finite-threshold success probability over all valid processes remains an open design problem. Our objective does not optimize expected duration or the agent's net payoff, and the numerical choices assume specified Q and R/c. Robustness to uncertainty in these inputs, adaptive tempering, and joint reward design remain outside the present analysis. Such objectives may also be relevant in other sequential learning settings with costly or strategic data acquisition, risk-sensitive decision makers, or endogenous stopping.

Supplementary material

Appendix A contains proofs of the main theoretical results, examples satisfying the local-stability condition, and a counterexample without local stability. Appendix B describes the Gaussian and Bernoulli numerical methods, presents additional validation and sensitivity checks, and specifies the computational settings used for the figures.

AI use statement

In this work, we used generative AI tools to assist in the construction and checking of proofs, implement numerical methods and R code, and assist with writing and editing. We did not use generative AI tools to generate synthetic datasets, translate material, clean or reformat datasets, or perform qualitative data analysis; these tasks are not applicable to this work.

All AI-assisted mathematical arguments were reviewed and checked by the authors, and the final theorem statements and proofs were independently verified for logical consistency. AI-assisted numerical code was run, debugged, and tested by the authors, including checks under finer discretization grids to assess numerical stability. Literature-related claims and citations suggested during the writing process were checked against the original sources before inclusion. We take responsibility for the final content of this work, including all text, claims, proofs, code, and other artifacts produced with the aid of generative AI.

Reproducibility Statement

All assumptions used in the theoretical results are stated in the main text, and complete proofs are provided in Appendix A. The numerical solution of the Bellman and hitting-probability equations is described in Appendix B. The additional validation in Appendix B.1 specifies the grids, transition integration, truncation rule, and convergence tolerances; accompanying code and numerical outputs reproduce those checks and all three figures. Figure-specific settings are given in Appendix B.2.

References

Ander Artola Velasco, Stratis Tsirtsis, and Manuel Gomez-Rodriguez. Optimizing social utility in sequential experiments. arXiv preprint arXiv:2605.06520, 2026.

Søren Asmussen. Ruin theory. In Encyclopedia of Quantitative Finance. John Wiley & Sons, Ltd, 2010. doi: 10.1002/9780470061602.eqf21001.

Stephen Bates, Michael I Jordan, Michael Sklar, and Jake A Soloff. Principal-agent hypothesis testing. arXiv preprint arXiv:2205.06812, 2022.

Enzo Busseti, Ernest K. Ryu, and Stephen Boyd. Risk-constrained Kelly gambling. The Journal of Investing, 25(3):118–134, 2016. doi: 10.3905/joi.2016.25.3.118. URL https://doi.org/10.3905/joi.2016.25.3.118.

Yatong Chen, Safwan Hossain, and Yiling Chen. Strategic hypothesis testing. Advances in Neural Information Processing Systems, 38:146285–146312, 2026.

Dino Gerardi and Lucas Maestri. A principal–agent model of sequential testing. Theoretical Economics, 7(3):425–463, 2012.

Peter Grünwald, Rianne de Heide, and Wouter Koolen. Safe testing. Journal of the Royal Statistical Society Series B: Statistical Methodology, 86(5):1091–1128, 2024.

Jacques Janssen and Raimondo Manca. Applied semi-Markov processes. Springer, 2006.

Gary Lorden. On excess over the boundary. The Annals of Mathematical Statistics, 41(2): 520–527, 1970.

Andrew McClellan. Experimentation and approval mechanisms. Econometrica, 90(5):2215–2247, 2022.

Daehong Min. Screening for experiments. Games and Economic Behavior, 142:73–100, 2023.

Drew T Nguyen, Alireza Fallah, and Michael I Jordan. Incentive design in sequential statistical protocols. In The ICML 2026 Workshop on Hypothesis Testing, 2026.

Aaditya Ramdas and Ruodu Wang. Hypothesis testing with e-values. Foundations and Trends in Statistics, 1(1–2):1–390, 2025. doi: 10.1561/3600000002. URL https://www.emerald.com/ftstat/article/1/1-2/1/1332095/Hypothesis-Testing-with-E-values. Also available as arXiv:2410.23614 [math.ST].

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales. arXiv preprint arXiv:2009.03167, 2020.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. Statistical Science, 38(4):576–601, 2023.

Glenn Shafer. Testing by betting: A strategy for statistical and scientific communication. Journal of the Royal Statistical Society Series A: Statistics in Society, 184(2):407–431, 2021.

Shubhanshu Shekhar. Optimal anytime-valid tests for composite nulls. arXiv preprint arXiv:2512.20039, 2025.

Flora C Shi, Stephen Bates, and Martin J Wainwright. Sharp results for hypothesis testing with risk-sensitive agents. arXiv preprint arXiv:2412.16452, 2024.

Aleksey Tetenov. An economic theory of statistical testing. Technical report, cemmap working paper, 2016.

Ian Waudby-Smith, Ricardo Sandoval, and Michael I Jordan. Universal log-optimality for general classes of e-processes and sequential hypothesis tests. arXiv preprint arXiv:2504.02818, 2025.

A Appendix

A.1 Proof of Theorem 4.2

Proof. Dividing the agent's payoff by c does not change its optimal stopping rule. We may therefore normalize the cost of one observation to one and take the reward upon rejection to be $\rho = \rho_\alpha = R/c$.

For $\lambda \in (0, 1]$, write

$$S_n^{(\lambda)} = \log E_n^{(\lambda)} = \sum_{i=1}^{n} Y_i^{(\lambda)}, \qquad S_0^{(\lambda)} = 0.$$

Since

$$L(x) = \exp\left(\mu x - \frac{\mu^2}{2}\right)$$

and

$$\log \mathbb{E}_P\{L(X)^\lambda\} = \frac{\mu^2}{2}(\lambda^2 - \lambda),$$

the log-increments are

$$Y_i^{(\lambda)} = \lambda\mu X_i - \frac{\lambda^2\mu^2}{2}. \tag{5}$$

Under $Q = N(\mu, 1)$, they are independent and identically distributed with

$$Y_i^{(\lambda)} \sim N\left(g_\lambda, \sigma_\lambda^2\right), \qquad g_\lambda = \frac{\mu^2}{2}(2\lambda - \lambda^2), \qquad \sigma_\lambda^2 = \lambda^2\mu^2. \tag{6}$$

In particular, $g_\lambda > 0$ for every $\lambda \in (0, 1]$.

We first describe the agent's optimal stopping rule and locate its stopping boundary.

The agent's stopping boundary. Fix $\lambda \in (0, 1]$ and temporarily suppress λ from the notation. Thus, let

$$S_n = \sum_{i=1}^{n} Y_i, \qquad Y_i \sim N(g, \sigma^2), \qquad g > 0.$$

If the current log-evidence is $z < U_\alpha$, let

$$x = U_\alpha - z$$

be the remaining distance to rejection. By spatial homogeneity, the agent's decision depends on z and U_α only through x.

For $x > 0$, define

$$\tau_x = \inf\{n \geq 0 : S_n \geq x\}.$$

The agent's value when the remaining distance is x is

$$v_\rho(x) = \sup_\sigma \mathbb{E}_Q \left[\rho\, \mathbf{1}\{\tau_x \leq \sigma\} - (\tau_x \wedge \sigma)\right], \qquad x > 0, \tag{7}$$

where the supremum is over stopping times for the future increments. For convenience, set

$$v_\rho(x) = \rho, \qquad x \leq 0,$$

because rejection has already occurred when the remaining distance is nonpositive.

The value function satisfies

$$v_\rho(x) = \max\left\{0,\, -1 + \mathbb{E}_Q\big[v_\rho(x - Y_1)\big]\right\}, \qquad x > 0. \tag{8}$$

For completeness, this equation can be obtained by first restricting the agent to at most N further observations. The corresponding finite-horizon values satisfy the same recursion and increase with N. They are bounded between zero and ρ, and hence converge to a limit satisfying equation 8. Conversely, every stopping rule with finite expected duration can be truncated at time N, with its expected payoff converging to the untruncated payoff. Thus the limiting value is precisely equation 7.

The function v_ρ is nonincreasing. Indeed, if $0 < x_1 < x_2$, then, along every realization of the increments, $\tau_{x_1} \leq \tau_{x_2}$. Applying the same quitting rule to the two problems therefore gives at least as large a payoff when the remaining distance is x_1. Taking suprema gives

$$v_\rho(x_1) \geq v_\rho(x_2).$$

It follows that the continuation value

$$C_\rho(x) = -1 + \mathbb{E}_Q\big[v_\rho(x - Y_1)\big] \tag{9}$$

is also nonincreasing in x. It is continuous because it is the convolution of the bounded function v_ρ with a Gaussian density.

We next show that $C_\rho(x) < 0$ for all sufficiently large x. Let $N_\rho = \lceil \rho \rceil$. For any quitting rule σ, put $T = \tau_x \wedge \sigma$. Pathwise,

$$\rho\, \mathbf{1}\{\tau_x \leq \sigma\} - T \leq \rho\, \mathbf{1}\{\tau_x \leq N_\rho\}.$$

Indeed, failure gives a nonpositive payoff, and success after time N_ρ also gives a nonpositive payoff. Taking expectations and then the supremum over σ yields

$$0 \leq v_\rho(x) \leq \rho\, \mathbb{P}_Q\left(\max_{0 \leq n \leq N_\rho} S_n \geq x\right). \tag{10}$$

For fixed ρ, the right-hand side tends to zero as $x \to \infty$. Hence $v_\rho(x) \to 0$, and dominated convergence gives

$$C_\rho(x) \longrightarrow -1 \qquad \text{as } x \to \infty.$$

Consequently, under the convention that the agent quits when indifferent, the optimal policy has a finite threshold

$$a_\rho = \sup\big(\{0\} \cup \{x > 0 : C_\rho(x) > 0\}\big). \tag{11}$$

For $x > 0$, the agent continues when $x < a_\rho$ and quits when $x \geq a_\rho$. This threshold rule attains the Bellman value. To see this, let $D_n = x - S_n$ be the remaining distance and stop when $D_n \leq 0$ or $D_n \geq a_\rho$. While $0 < D_n < a_\rho$, equation 8 gives

$$\mathbb{E}_Q\big[v_\rho(D_{n+1}) \mid \mathcal{F}_n\big] = v_\rho(D_n) + 1.$$

Thus

$$v_\rho(D_{n \wedge T}) - (n \wedge T)$$

is a martingale, where T is the first exit time from $(0, a_\rho)$. This exit time is integrable. Indeed, $T \leq \tau_x$, and a positive-drift random walk with finite exponential moments satisfies

$\mathbb{E}_Q(\tau_x) < \infty$. Hence $\mathbb{E}_Q(T) < \infty$. Since $0 \le v_\rho \le \rho$, the stopped martingale is dominated by the integrable random variable $\rho + T$. Optional stopping therefore gives

$$v_\rho(x) = \mathbb{E}_Q\left[\rho\,\mathbf{1}\{D_T \le 0\} - T\right],$$

which is exactly the payoff of the threshold rule.

We now establish the first-order location of a_ρ:

$$\frac{a_\rho}{\rho} \longrightarrow g \qquad \text{as } \rho \to \infty. \tag{12}$$

Fix $\delta \in (0, g)$. First consider

$$x_\rho^- = (g - \delta)\rho.$$

The policy that never quits is feasible. Its expected payoff is

$$\rho - \mathbb{E}_Q(\tau_{x_\rho^-}).$$

Let

$$O_x = S_{\tau_x} - x$$

be the upper overshoot. By Wald's identity (Janssen & Manca, 2006) and Lorden's uniform mean-overshoot bound (Lorden, 1970),

$$\mathbb{E}_Q(\tau_x) = \frac{x + \mathbb{E}_Q(O_x)}{g} = \frac{x}{g} + O(1), \qquad x \to \infty,$$

where the $O(1)$ term is uniform in x. Therefore,

$$\rho - \mathbb{E}_Q(\tau_{x_\rho^-}) = \frac{\delta}{g}\rho + O(1) > 0$$

for all sufficiently large ρ. Hence $v_\rho(x_\rho^-) > 0$, so $C_\rho(x_\rho^-) > 0$, and consequently

$$a_\rho \ge (g - \delta)\rho$$

for all sufficiently large ρ.

For the reverse bound, set

$$x_\rho^+ = (g + \delta)\rho.$$

Applying equation 10 conditionally after one additional increment gives

$$\mathbb{E}_Q\left[v_\rho(x_\rho^+ - Y_1)\right] \le \rho\,\mathbb{P}_Q\left(\max_{1 \le m \le N_\rho + 1} S_m \ge x_\rho^+\right). \tag{13}$$

For $m \le N_\rho + 1 \le \rho + 2$ and all sufficiently large ρ,

$$x_\rho^+ - mg \ge \delta\rho - 2g \ge \frac{\delta\rho}{2}.$$

Since $S_m \sim N(mg, m\sigma^2)$, the Gaussian tail bound and a union bound imply

$$\begin{aligned}\mathbb{P}_Q\left(\max_{1 \le m \le N_\rho + 1} S_m \ge x_\rho^+\right) &\le \sum_{m=1}^{N_\rho + 1} \exp\left\{-\frac{(x_\rho^+ - mg)^2}{2m\sigma^2}\right\} \\ &\le (\rho + 2)\exp\left(-\frac{\delta^2}{16\sigma^2}\rho\right).\end{aligned}$$

The right-hand side of equation 13 therefore converges to zero. It follows that

$$C_\rho(x_\rho^+) = -1 + \mathbb{E}_Q\left[v_\rho(x_\rho^+ - Y_1)\right] < 0$$

for all sufficiently large ρ. Hence

$$a_\rho \le (g + \delta)\rho.$$

Since $\delta > 0$ was arbitrary, equation 12 follows.

Returning to the tempered process, denote its optimal distance boundary by $a_{\lambda,\rho}$. From equation 6 and equation 12,

$$\frac{a_{\lambda,\rho}}{\rho} \longrightarrow g_\lambda = \frac{\mu^2}{2}(2\lambda - \lambda^2) \qquad \text{as } \rho \to \infty. \tag{14}$$

Because $U_\alpha \to \infty$ and

$$\frac{\rho_\alpha \mu^2}{2U_\alpha} \longrightarrow r > 1,$$

we have $\rho_\alpha \to \infty$. At $\lambda = 1$,

$$g_1 = \frac{\mu^2}{2},$$

and hence

$$\frac{a_{1,\rho_\alpha}}{U_\alpha} = \frac{a_{1,\rho_\alpha}}{\rho_\alpha}\frac{\rho_\alpha}{U_\alpha} \longrightarrow r > 1.$$

Thus $a_{1,\rho_\alpha} > U_\alpha$ for all sufficiently small α. Since the initial remaining distance is U_α, the agent continues at time zero under $E^{(1)}$.

The probability of quitting. Fix $\lambda \in (0, 1]$. From equation 14,

$$\frac{a_{\lambda,\rho_\alpha}}{U_\alpha} \longrightarrow r(2\lambda - \lambda^2). \tag{15}$$

Suppose for the moment that

$$r(2\lambda - \lambda^2) > 1.$$

Then the agent initially continues for all sufficiently small α, proving the first claim. For such α, define

$$d_{\alpha,\lambda} = a_{\lambda,\rho_\alpha} - U_\alpha > 0.$$

On the log-evidence scale, the agent's lower stopping boundary is

$$U_\alpha - a_{\lambda,\rho_\alpha} = -d_{\alpha,\lambda}.$$

Therefore, writing

$$\tau_u^+ = \inf\{n \geq 0 : S_n^{(\lambda)} \geq u\}, \qquad \tau_{-d}^- = \inf\{n \geq 0 : S_n^{(\lambda)} \leq -d\},$$

the probability of quitting before rejection is

$$q_{\alpha,\lambda} := 1 - \pi_{\alpha,\rho_\alpha}(\lambda) = \mathbb{P}_Q\left(\tau_{-d_{\alpha,\lambda}}^- < \tau_{U_\alpha}^+\right). \tag{16}$$

The Gaussian increments have a continuous distribution, so the choice of strict or weak inequalities at the two boundaries is immaterial.

Define

$$\kappa_\lambda = \frac{2g_\lambda}{\sigma_\lambda^2} = \frac{2}{\lambda} - 1. \tag{17}$$

Then

$$\mathbb{E}_Q \exp\left\{-\kappa_\lambda Y_1^{(\lambda)}\right\} = 1.$$

Define a probability measure $\widetilde{Q}_\lambda$ on $\mathcal{F}_n$ by

$$\left.\frac{\mathrm{d}\widetilde{Q}_\lambda}{\mathrm{d}Q}\right|_{\mathcal{F}_n} = \exp\left\{-\kappa_\lambda S_n^{(\lambda)}\right\}. \tag{18}$$

Under $\widetilde{Q}_\lambda$, the increments remain Gaussian, but their mean changes from g_λ to

$$g_\lambda - \kappa_\lambda \sigma_\lambda^2 = -g_\lambda.$$

Let

$$T = \tau_{-d_{\alpha,\lambda}}^- \wedge \tau_{U_\alpha}^+$$

and let

$$A = \left\{ \tau^-_{-d_{\alpha,\lambda}} < \tau^+_{U_\alpha} \right\}.$$

On A, write

$$O_{\alpha,\lambda} = -d_{\alpha,\lambda} - S^{(\lambda)}_T \geq 0$$

for the lower overshoot. Applying equation 18 on each event $\{T = n\}$ and summing over n gives the exact identity

$$q_{\alpha,\lambda} = e^{-\kappa_\lambda d_{\alpha,\lambda}} \widetilde{\mathbb{E}}_\lambda \left[e^{-\kappa_\lambda O_{\alpha,\lambda}} \mathbf{1}_A \right]. \tag{19}$$

In particular,

$$q_{\alpha,\lambda} \leq e^{-\kappa_\lambda d_{\alpha,\lambda}}. \tag{20}$$

We also need a matching lower bound. Under $\widetilde{Q}_\lambda$, define the one-sided lower passage time

$$\widehat{\tau}_d = \inf\{n \geq 0 : S^{(\lambda)}_n \leq -d\}$$

and its overshoot

$$\widehat{O}_d = -d - S^{(\lambda)}_{\widehat{\tau}_d}.$$

The random walk $-S^{(\lambda)}_n$ has positive drift g_λ under $\widetilde{Q}_\lambda$, so the strong law gives $\widehat{\tau}_d < \infty$ almost surely. Lorden's inequality therefore gives a constant $K_\lambda < \infty$ such that

$$\sup_{d \geq 0} \widetilde{\mathbb{E}}_\lambda(\widehat{O}_d) \leq K_\lambda. \tag{21}$$

Set $M_\lambda = 2K_\lambda + 1$. By Markov's inequality,

$$\inf_{d \geq 0} \widetilde{\mathbb{P}}_\lambda(\widehat{O}_d \leq M_\lambda) > \frac{1}{2}.$$

Moreover,

$$\exp\left\{ \kappa_\lambda S^{(\lambda)}_n \right\}$$

is a nonnegative mean-one martingale under $\widetilde{Q}_\lambda$. The nonnegative martingale maximal inequality gives

$$\widetilde{\mathbb{P}}_\lambda \left(\tau^+_{U_\alpha} < \infty \right) \leq e^{-\kappa_\lambda U_\alpha}. \tag{22}$$

For all sufficiently small α, the right-hand side is less than $1/4$. Hence

$$\begin{aligned} &\widetilde{\mathbb{P}}_\lambda \left(\widehat{O}_{d_{\alpha,\lambda}} \leq M_\lambda,\, \widehat{\tau}_{d_{\alpha,\lambda}} < \tau^+_{U_\alpha} \right) \\ &\qquad \geq \widetilde{\mathbb{P}}_\lambda \left(\widehat{O}_{d_{\alpha,\lambda}} \leq M_\lambda \right) - \widetilde{\mathbb{P}}_\lambda \left(\tau^+_{U_\alpha} < \infty \right) \geq \frac{1}{4}. \end{aligned}$$

On this event, the overshoot in equation 19 is at most M_λ. Consequently,

$$q_{\alpha,\lambda} \geq \frac{1}{4} \exp\left\{ -\kappa_\lambda (d_{\alpha,\lambda} + M_\lambda) \right\}. \tag{23}$$

Combining equation 20 and equation 23,

$$\kappa_\lambda d_{\alpha,\lambda} \leq -\log q_{\alpha,\lambda} \leq \kappa_\lambda d_{\alpha,\lambda} + \kappa_\lambda M_\lambda + \log 4.$$

Using equation 15, we obtain

$$-\frac{1}{U_\alpha} \log\left\{ 1 - \pi_{\alpha,\rho_\alpha}(\lambda) \right\} \longrightarrow I_r(\lambda), \tag{24}$$

where

$$I_r(\lambda) = \left(\frac{2}{\lambda} - 1 \right) \left\{ r(2\lambda - \lambda^2) - 1 \right\}. \tag{25}$$

At $\lambda = 1$,

$$I_r(1) = r - 1 > 0.$$

Moreover,

$$I_r'(1) = -2(r-1) < 0.$$

Since I_r is continuously differentiable, there exists $\varepsilon > 0$ such that, for every $\lambda \in (1-\varepsilon, 1)$,

$$r(2\lambda - \lambda^2) > 1 \qquad \text{and} \qquad I_r(\lambda) > I_r(1). \tag{26}$$

Fix any such λ, and put

$$\Delta_\lambda = I_r(\lambda) - I_r(1) > 0.$$

By equation 24, for all sufficiently small α,

$$-\frac{1}{U_\alpha} \log\{1 - \pi_{\alpha,\rho_\alpha}(\lambda)\} > I_r(\lambda) - \frac{\Delta_\lambda}{4},$$

whereas

$$-\frac{1}{U_\alpha} \log\{1 - \pi_{\alpha,\rho_\alpha}(1)\} < I_r(1) + \frac{\Delta_\lambda}{4}.$$

Because

$$I_r(\lambda) - \frac{\Delta_\lambda}{4} > I_r(1) + \frac{\Delta_\lambda}{4},$$

it follows that

$$1 - \pi_{\alpha,\rho_\alpha}(\lambda) < 1 - \pi_{\alpha,\rho_\alpha}(1)$$

for all sufficiently small α. Equivalently,

$$\pi_{\alpha,\rho_\alpha}(\lambda) > \pi_{\alpha,\rho_\alpha}(1).$$

Thus every fixed $\lambda < 1$ sufficiently close to one eventually gives the principal a strictly larger success probability than the likelihood-ratio process. Hence $E^{(1)}$ does not maximize the principal's success probability over the tempered likelihood-ratio family. □

Proof of Corollary 4.3. Rewriting the exponent gives

$$I_r(\lambda) = r(2-\lambda)^2 - \frac{2}{\lambda} + 1, \qquad I_r'(\lambda) = \frac{2}{\lambda^2}\{1 - r\lambda^2(2-\lambda)\}.$$

The function $\lambda^2(2-\lambda)$ is strictly increasing from zero to one on $(0,1)$. Thus I_r' changes sign exactly once, from positive to negative, at λ_r. Moreover, $r\lambda_r(2-\lambda_r) = 1/\lambda_r > 1$, so the maximizer lies in the participation region. The ratio conclusion follows by applying the pointwise failure-exponent limit at λ_r and λ. □

A.2 Proof of Proposition 4.4

Proof. Fix $\lambda \in (0,1]$ such that $\psi_0(\lambda\ell^*) < \infty$, and fix $P_0 \in \mathcal{P}_0$. By the definition of ψ_0,

$$\log \mathbb{E}_{P_0}\left[\exp\{\lambda\ell^*(X)\}\right] \le \psi_0(\lambda\ell^*).$$

Equivalently,

$$\mathbb{E}_{P_0}\left[\exp\left\{\lambda\ell^*(X) - \psi_0(\lambda\ell^*)\right\}\right] \le 1. \tag{27}$$

Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Since the observations are independent and identically distributed under P_0, X_n is independent of $\mathcal{F}_{n-1}$. Therefore, for every $n \ge 1$,

$$\begin{aligned}
\mathbb{E}_{P_0}\left[W_n^{(\lambda)} \mid \mathcal{F}_{n-1}\right] &= W_{n-1}^{(\lambda)} \mathbb{E}_{P_0}\left[\exp\left\{\lambda\ell^*(X_n) - \psi_0(\lambda\ell^*)\right\} \mid \mathcal{F}_{n-1}\right] \\
&= W_{n-1}^{(\lambda)} \mathbb{E}_{P_0}\left[\exp\left\{\lambda\ell^*(X) - \psi_0(\lambda\ell^*)\right\}\right] \\
&\le W_{n-1}^{(\lambda)},
\end{aligned}$$

where the inequality follows from equation 27. Hence $W^{(\lambda)}$ is a nonnegative supermartingale under P_0. Since $P_0 \in \mathcal{P}_0$ is arbitrary this shows $W^{(\lambda)}$ is an e-process for $\mathcal{P}_0$.

It remains to establish the growth-optimality claim. Let ν denote the common dominating measure, and let q, p_0^*, and p_0 denote the densities of Q, P_0^*, and an arbitrary $P_0 \in \mathcal{P}_0$, respectively. Since

$$d_{\mathrm{KL}}(Q\|P_0^*) < \infty,$$

we have $Q \ll P_0^*$. Define

$$L^* = \frac{q}{p_0^*}$$

on $\{p_0^* > 0\}$, and set $L^* = 0$ elsewhere.

We first show that L^* is an e-variable for the composite null. Define

$$h_0(x) = \frac{p_0(x)}{p_0^*(x)}$$

on $\{p_0^* > 0\}$, and set $h_0(x) = 0$ elsewhere. For $t \in (0,1)$, let

$$P_t = (1-t)P_0^* + tP_0, \qquad p_t = (1-t)p_0^* + tp_0.$$

By convexity of $\mathcal{P}_0$, $P_t \in \mathcal{P}_0$. Moreover, $p_t \geq (1-t)p_0^*$, so $d_{\mathrm{KL}}(Q\|P_t) < \infty$. Since P_0^* minimizes $d_{\mathrm{KL}}(Q\|P_0)$ over $P_0 \in \mathcal{P}_0$,

$$\begin{aligned} 0 &\leq d_{\mathrm{KL}}(Q\|P_t) - d_{\mathrm{KL}}(Q\|P_0^*) \\ &= -\mathbb{E}_Q\left[\log\left\{\frac{p_t(X)}{p_0^*(X)}\right\}\right] \\ &= -\mathbb{E}_Q\left[\log\{1 + t(h_0(X) - 1)\}\right]. \end{aligned}$$

Consequently,

$$\mathbb{E}_Q\left[\frac{\log\{1 + t(h_0(X) - 1)\}}{t}\right] \leq 0. \tag{28}$$

For $u \geq 0$, define

$$\phi_t(u) = \frac{\log\{1 + t(u-1)\}}{t}.$$

For every $u \geq 0$,

$$\phi_t(u) \longrightarrow u - 1 \qquad \text{as } t \downarrow 0.$$

Moreover, when $0 < t \leq 1/2$,

$$\phi_t(u) \geq \frac{\log(1-t)}{t} \geq -2.$$

Let $t_m = 1/m$, $m \geq 2$. Since $\phi_{t_m}\{h_0(X)\} + 2 \geq 0$, Fatou's lemma and equation 28 give

$$\begin{aligned} \mathbb{E}_Q\{h_0(X) + 1\} &= \mathbb{E}_Q\left[\liminf_{m\to\infty}\{\phi_{t_m}(h_0(X)) + 2\}\right] \\ &\leq \liminf_{m\to\infty} \mathbb{E}_Q\left[\phi_{t_m}(h_0(X)) + 2\right] \\ &\leq 2. \end{aligned}$$

Hence

$$\mathbb{E}_Q\{h_0(X)\} \leq 1.$$

Finally,

$$\begin{aligned} \mathbb{E}_Q\{h_0(X)\} &= \int_{\{p_0^* > 0\}} \frac{p_0(x)}{p_0^*(x)} q(x)\,\mathrm{d}\nu(x) \\ &= \int_{\{p_0^* > 0\}} \frac{q(x)}{p_0^*(x)} p_0(x)\,\mathrm{d}\nu(x) \\ &= \mathbb{E}_{P_0}\{L^*(X)\}. \end{aligned}$$

Therefore,

$$\mathbb{E}_{P_0}\{L^*(X)\} \leq 1, \qquad P_0 \in \mathcal{P}_0.$$

Equality holds at $P_0 = P_0^*$, because

$$\mathbb{E}_{P_0^*}\{L^*(X)\} = \int_{\{p_0^*>0\}} q(x)\,\mathrm{d}\nu(x) = 1.$$

It follows that

$$\begin{aligned}\psi_0(\ell^*) &= \sup_{P_0\in\mathcal{P}_0} \log \mathbb{E}_{P_0}\{\exp(\ell^*(X))\} \\ &= \sup_{P_0\in\mathcal{P}_0} \log \mathbb{E}_{P_0}\{L^*(X)\} = 0.\end{aligned}$$

Therefore,

$$W_n^{(1)} = \prod_{i=1}^n L^*(X_i),$$

and

$$\mathbb{E}_Q\{\log W_n^{(1)}\} = n\mathbb{E}_Q\{\log L^*(X)\} = n\, d_{\mathrm{KL}}(Q\|P_0^*) = n\gamma^*.$$

Now let $E = (E_n)_{n\geq 0}$ be any e-process for $\mathcal{P}_0$. Write

$$L_n^* = \prod_{i=1}^n L^*(X_i) = \frac{\mathrm{d}Q^n}{\mathrm{d}(P_0^*)^n}(X_1,\ldots,X_n).$$

Because $L_n^* > 0$, Q^n-almost surely, Jensen's inequality gives

$$\begin{aligned}\mathbb{E}_Q\{\log E_n\} &= \mathbb{E}_Q\{\log L_n^*\} + \mathbb{E}_Q\left[\log\left(\frac{E_n}{L_n^*}\right)\right] \\ &\leq n\gamma^* + \log \mathbb{E}_Q\left(\frac{E_n}{L_n^*}\right) \\ &\leq n\gamma^* + \log \mathbb{E}_{P_0^*}(E_n) \\ &\leq n\gamma^*.\end{aligned}$$

The second inequality holds because

$$\mathbb{E}_Q\left(\frac{E_n}{L_n^*}\right) = \int_{\{L_n^*>0\}} E_n\,\mathrm{d}(P_0^*)^n \leq \mathbb{E}_{P_0^*}(E_n).$$

Since $W^{(1)}$ attains the upper bound $n\gamma^*$, it is log-growth optimal.

□

A.3 Proof of Theorem 4.7

Proof. Dividing the agent's payoff by c does not change its optimal stopping rule. We may therefore normalize the cost of one observation to one and take the reward upon rejection to be ρ.

For $\lambda \in [1-\eta, 1]$, local stability gives

$$\log W_n^{(\lambda)} = \sum_{i=1}^n Y_i^{(\lambda)}, \qquad Y_i^{(\lambda)} = \lambda\ell^*(X_i) - A(\lambda).$$

Under Q, these increments are independent and identically distributed, with mean

$$g(\lambda) := \mathbb{E}_Q\{Y_1^{(\lambda)}\} = \lambda\gamma^* - A(\lambda). \tag{29}$$

Since

$$A(1) = 0, \qquad A'(1) = \mathbb{E}_Q\{\ell^*(X)\} = \gamma^*,$$

we have

$$g(1) = \gamma^*, \qquad g'(1) = 0. \tag{30}$$

Since A is convex and

$$A(0) = A(1) = 0,$$

we have $A(\lambda) \le 0$ for every $\lambda \in [0,1]$. Hence

$$g(\lambda) = \lambda\gamma^* - A(\lambda) \ge \lambda\gamma^* > 0, \qquad \lambda \in (0,1].$$

The finiteness of A on an open interval containing $[0,1]$ additionally ensures the exponential-moment conditions used below.

We first describe the agent's optimal stopping boundary.

Step 1: location of the optimal stopping boundary. Fix λ such that $g(\lambda) > 0$, and suppress λ from the notation. Let

$$S_n = \sum_{i=1}^{n} Y_i, \qquad g = \mathbb{E}_Q(Y_1) > 0.$$

If the current log-evidence is $z < U_\alpha$, write

$$x = U_\alpha - z$$

for the remaining distance to rejection. Let $v_\rho(x)$ be the agent's optimal expected payoff when the remaining distance is x, with $v_\rho(x) = \rho$ for $x \le 0$. Standard finite-horizon approximation gives the Bellman equation

$$v_\rho(x) = \max\left\{0,\, -1 + \mathbb{E}_Q\big[v_\rho(x - Y_1)\big]\right\}, \qquad x > 0. \tag{31}$$

The function v_ρ is nonincreasing: reducing the distance to rejection can only make every admissible stopping rule more profitable. Hence the continuation value

$$C_\rho(x) = -1 + \mathbb{E}_Q\big[v_\rho(x - Y_1)\big]$$

is also nonincreasing. As in the proof of Theorem 4.2, let $N_\rho = \lceil \rho \rceil$. For every $x > 0$,

$$0 \le v_\rho(x) \le \rho\, \mathbb{P}_Q\left(\max_{0 \le n \le N_\rho} S_n \ge x\right). \tag{32}$$

For fixed ρ, the probability on the right-hand side tends to zero as $x \to \infty$. Hence $v_\rho(x) \to 0$, and dominated convergence gives

$$C_\rho(x) = -1 + \mathbb{E}_Q\{v_\rho(x - Y_1)\} \longrightarrow -1.$$

Therefore, the stopping set

$$\mathcal{S}_{\lambda,\rho} = \{x > 0 : C_\rho(x) \le 0\}$$

is nonempty. Since C_ρ is nonincreasing, $\mathcal{S}_{\lambda,\rho}$ is an upper interval. Define

$$a_{\lambda,\rho} = \inf \mathcal{S}_{\lambda,\rho}. \tag{33}$$

The agent continues when $x \notin \mathcal{S}_{\lambda,\rho}$ and quits when $x \in \mathcal{S}_{\lambda,\rho}$. In particular, it continues for $x < a_{\lambda,\rho}$ and quits for $x > a_{\lambda,\rho}$. At $x = a_{\lambda,\rho}$, it quits if and only if $a_{\lambda,\rho} \in \mathcal{S}_{\lambda,\rho}$.

We next verify that this stopping rule attains the Bellman value. Let

$$D_n = x - S_n$$

denote the remaining distance to rejection, and define

$$T = \inf\left\{n \ge 0 : D_n \le 0 \text{ or } D_n \in \mathcal{S}_{\lambda,\rho}\right\}.$$

Also let

$$\tau_x = \inf\{n \ge 0 : S_n \ge x\}.$$

Since $T \le \tau_x$, and a positive-drift random walk with finite exponential moments satisfies $\mathbb{E}_Q(\tau_x) < \infty$, we have $\mathbb{E}_Q(T) < \infty$.

For $n < T$, the agent is in the continuation region, and the Bellman equation gives

$$\mathbb{E}_Q\left[v_\rho(D_{n+1}) \mid \mathcal{F}_n\right] = v_\rho(D_n) + 1.$$

Consequently,

$$v_\rho(D_{n\wedge T}) - (n \wedge T)$$

is a martingale. Since $0 \le v_\rho \le \rho$ and $\mathbb{E}_Q(T) < \infty$, optional stopping and dominated convergence give

$$v_\rho(x) = \mathbb{E}_Q\left[\rho\,\mathbf{1}\{D_T \le 0\} - T\right].$$

On $\{D_T \le 0\}$, rejection has occurred and the terminal value is ρ; on $\{D_T \in \mathcal{S}_{\lambda,\rho}\}$, the agent quits and the terminal value is zero. Thus the stopping rule defined by $\mathcal{S}_{\lambda,\rho}$ attains the Bellman value.

We claim that

$$\frac{a_{\lambda,\rho}}{\rho} \longrightarrow g(\lambda) \qquad \text{as } \rho \to \infty. \tag{34}$$

To prove the lower bound, fix $\delta \in (0, g)$ and set

$$x_\rho^- = (g - \delta)\rho.$$

Consider the feasible policy that never quits. Its expected payoff is

$$\rho - \mathbb{E}_Q(\tau_{x_\rho^-}).$$

Because the increments have positive mean and a finite exponential moment, Lorden's inequality (Lorden, 1970) gives

$$\mathbb{E}_Q(\tau_x) = \frac{x}{g} + O(1), \qquad x \to \infty.$$

Consequently,

$$\rho - \mathbb{E}_Q(\tau_{x_\rho^-}) = \frac{\delta}{g}\rho + O(1) > 0$$

for all sufficiently large ρ. Hence the agent continues at x_ρ^-, and

$$a_{\lambda,\rho} \ge (g - \delta)\rho \tag{35}$$

for all sufficiently large ρ.

For the upper bound, set

$$x_\rho^+ = (g + \delta)\rho$$

and let $N_\rho = \lceil \rho \rceil$. For any stopping rule, a success occurring after more than N_ρ observations has nonpositive net payoff. It follows that, for every x,

$$v_\rho(x) \le \rho\,\mathbb{P}_Q\left(\max_{0\le n\le N_\rho} S_n \ge x\right).$$

Conditioning on the first increment therefore gives

$$\mathbb{E}_Q\left[v_\rho(x_\rho^+ - Y_1)\right] \le \rho\,\mathbb{P}_Q\left(\max_{1\le n\le N_\rho+1} S_n \ge x_\rho^+\right). \tag{36}$$

Since Y_1 has a finite positive exponential moment, there exists $\theta > 0$ such that

$$\log \mathbb{E}_Q(e^{\theta Y_1}) \le \theta\left(g + \frac{\delta}{4}\right).$$

Chernoff's inequality and a union bound then yield

$$\begin{aligned}\mathbb{P}_Q\left(\max_{1\le n\le N_\rho+1} S_n \ge x_\rho^+\right) &\le \sum_{n=1}^{N_\rho+1} \exp\left\{-\theta x_\rho^+ + n \log \mathbb{E}_Q(e^{\theta Y_1})\right\} \\ &\le C\rho e^{-c\rho}\end{aligned}$$

for some constants $c, C > 0$ and all sufficiently large ρ. The right-hand side of equation 36 is therefore at most $C\rho^2 e^{-c\rho}$, which tends to zero. Thus

$$C_\rho(x_\rho^+) = -1 + \mathbb{E}_Q\left[v_\rho(x_\rho^+ - Y_1)\right] < 0$$

for all sufficiently large ρ, and hence

$$a_{\lambda,\rho} \le (g+\delta)\rho. \tag{37}$$

Since $\delta > 0$ was arbitrary, equation 34 follows.

Now

$$\frac{\rho_\alpha \gamma^*}{U_\alpha} \longrightarrow r > 1$$

implies $\rho_\alpha \to \infty$. Applying equation 34 at $\lambda = 1$, and using $g(1) = \gamma^*$, gives

$$\frac{a_{1,\rho_\alpha}}{U_\alpha} = \frac{a_{1,\rho_\alpha}}{\rho_\alpha}\frac{\rho_\alpha}{U_\alpha} \longrightarrow r > 1.$$

The initial remaining distance is U_α. Therefore $a_{1,\rho_\alpha} > U_\alpha$ for all sufficiently small α, so the agent continues at time zero under $W^{(1)}$.

More generally, for every fixed λ sufficiently close to one,

$$\frac{a_{\lambda,\rho_\alpha}}{U_\alpha} \longrightarrow \frac{rg(\lambda)}{\gamma^*}. \tag{38}$$

Since g is continuous and $rg(1)/\gamma^* = r > 1$, there exists $\varepsilon_1 \in (0, \eta)$ such that

$$\frac{rg(\lambda)}{\gamma^*} > 1, \qquad \lambda \in [1-\varepsilon_1, 1].$$

Thus the agent also continues at time zero for every fixed λ in this interval when α is sufficiently small.

Step 2: the exponential rate of the quitting probability. We next identify the probability that the log-process reaches the agent's lower boundary before reaching the rejection boundary.

For λ close to one, define $\kappa(\lambda) > 0$ by

$$\mathbb{E}_Q \exp\left\{-\kappa(\lambda) Y_1^{(\lambda)}\right\} = 1. \tag{39}$$

We verify below that such a root exists and is differentiable near $\lambda = 1$. First observe that

$$\mathbb{E}_Q \exp\left\{-\kappa\big(\lambda \ell^*(X) - A(\lambda)\big)\right\} = \exp\left\{A(1-\kappa\lambda) + \kappa A(\lambda)\right\}.$$

Hence $\kappa(\lambda)$ solves

$$F(\kappa, \lambda) := A(1-\kappa\lambda) + \kappa A(\lambda) = 0. \tag{40}$$

Let

$$\delta^* = d_{\mathrm{KL}}(P_0^* \| Q).$$

The assumptions on A and mutual absolute continuity imply that $0 < \delta^* < \infty$, and

$$A'(0) = -\delta^*, \qquad A'(1) = \gamma^*.$$

Since $A(0) = A(1) = 0$,

$$F(1,1) = 0, \qquad \frac{\partial F}{\partial \kappa}(1,1) = -A'(0) = \delta^* > 0.$$

The implicit function theorem therefore gives a continuously differentiable positive root $\kappa(\lambda)$ in a neighborhood of one, with $\kappa(1) = 1$. Strict convexity of $\kappa \mapsto F(\kappa, \lambda)$, with $F(0, \lambda) = 0$ and $\partial_\kappa F(0, \lambda) = -g(\lambda) < 0$, ensures uniqueness of the positive root. After decreasing the neighborhood of one if necessary, we assume throughout the remainder of the proof that $\kappa(\lambda)$ is defined and continuous for every λ under consideration. Differentiating equation 40 at $\lambda = 1$ gives

$$\kappa'(1) = -\left(1 + \frac{\gamma^*}{\delta^*}\right) < 0. \tag{41}$$

Indeed, $u(\lambda) = 1 - \kappa(\lambda)\lambda$ satisfies $u(1) = 0$ and $u'(1) = \gamma^*/\delta^* > 0$, so $u(\lambda) < 0$ just to the left of one. The open domain of A and mutual absolute continuity ensure that these negative power moments are finite.

Fix such a λ, and write

$$d_{\alpha,\lambda} = a_{\lambda,\rho_\alpha} - U_\alpha.$$

By equation 38,

$$\frac{d_{\alpha,\lambda}}{U_\alpha} \longrightarrow \frac{rg(\lambda)}{\gamma^*} - 1 > 0. \tag{42}$$

On the log scale, the rejection boundary is U_α, whereas the agent's quitting boundary is $-d_{\alpha,\lambda}$. Write $q_{\alpha,\lambda} = 1 - \pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(\lambda)$. For $b > 0$, define

$$q^{\mathrm{w}}_{\alpha,\lambda}(b) = \mathbb{P}_Q(\tau^-_{-b} < \tau^+_{U_\alpha}), \quad \tau^-_{-b} = \inf\{n \geq 0 : S^{(\lambda)}_n \leq -b\}, \quad \tau^+_{U_\alpha} = \inf\{n \geq 0 : S^{(\lambda)}_n \geq U_\alpha\}.$$

Whether the agent quits at equality or only after a strict crossing, for every fixed $h > 0$,

$$q^{\mathrm{w}}_{\alpha,\lambda}(d_{\alpha,\lambda} + h) \leq q_{\alpha,\lambda} \leq q^{\mathrm{w}}_{\alpha,\lambda}(d_{\alpha,\lambda}). \tag{43}$$

We bound the weak-exit probability for either $b_\alpha = d_{\alpha,\lambda}$ or $b_\alpha = d_{\alpha,\lambda} + h$.

Define a probability measure $\widetilde{Q}_\lambda$ on $\mathcal{F}_n$ by

$$\left.\frac{\mathrm{d}\widetilde{Q}_\lambda}{\mathrm{d}Q}\right|_{\mathcal{F}_n} = \exp\left\{-\kappa(\lambda) S^{(\lambda)}_n\right\}. \tag{44}$$

equation 39 ensures that this density has mean one. Under $\widetilde{Q}_\lambda$, the random walk has negative drift. Indeed, the log moment-generating function under Q is strictly convex, equals zero at both $-\kappa(\lambda)$ and 0, and has positive derivative $g(\lambda)$ at zero. Its derivative at $-\kappa(\lambda)$, which is the mean under $\widetilde{Q}_\lambda$, is therefore negative. The strong law then gives $S^{(\lambda)}_n \to -\infty$ almost surely under $\widetilde{Q}_\lambda$, so every fixed lower level is reached in finite time almost surely.

Let T be the first exit time from $(-b_\alpha, U_\alpha)$. On the event that the lower boundary is reached first, write

$$O_{\alpha,\lambda} = -b_\alpha - S^{(\lambda)}_T \geq 0$$

for the lower overshoot. Summing the change-of-measure identity over the possible values of T gives

$$q^{\mathrm{w}}_{\alpha,\lambda}(b_\alpha) = e^{-\kappa(\lambda) b_\alpha} \widetilde{\mathbb{E}}_\lambda \left[e^{-\kappa(\lambda) O_{\alpha,\lambda}}; \text{ the lower boundary is reached first}\right]. \tag{45}$$

In particular,

$$q^{\mathrm{w}}_{\alpha,\lambda}(b_\alpha) \leq e^{-\kappa(\lambda) b_\alpha}. \tag{46}$$

A matching lower bound requires control of the lower overshoot under the tilted measure. We first verify the required moment condition. For every s for which the following expectation is finite, the definition of $\widetilde{Q}_\lambda$ gives

$$\begin{aligned}\widetilde{\mathbb{E}}_\lambda \left[\exp\{s Y^{(\lambda)}_1\}\right] &= \mathbb{E}_Q \left[\exp\{(s - \kappa(\lambda)) Y^{(\lambda)}_1\}\right] \\ &= \exp\left[A(1 + \lambda\{s - \kappa(\lambda)\}) - \{s - \kappa(\lambda)\} A(\lambda)\right].\end{aligned} \tag{47}$$

Indeed, the second equality follows from

$$Y^{(\lambda)}_1 = \lambda \ell^*(X_1) - A(\lambda)$$

and

$$\mathbb{E}_Q\left[\exp\{u \ell^*(X)\}\right] = \mathbb{E}_{P^*_0}\left[\exp\{(1+u)\ell^*(X)\}\right] = \exp\{A(1+u)\}.$$

By Assumption 4.5, A is finite on an open interval containing $[0, 1]$. Moreover,

$$\kappa(\lambda) \longrightarrow 1 \qquad \text{as } \lambda \longrightarrow 1.$$

Hence, for every fixed λ sufficiently close to one, there exists $s_\lambda > 0$ such that

$$1 + \lambda\{s - \kappa(\lambda)\}$$

belongs to this interval whenever $|s| < s_\lambda$. It follows from equation 47 that

$$\widetilde{\mathbb{E}}_\lambda \left[\exp\{sY_1^{(\lambda)}\}\right] < \infty, \qquad |s| < s_\lambda.$$

Thus the tilted increment has finite exponential moments in a neighborhood of zero. In particular,

$$\widetilde{\mathbb{E}}_\lambda \left[\left\{\left(-Y_1^{(\lambda)}\right)^+\right\}^2\right] < \infty.$$

For $d \geq 0$, define the one-sided lower passage time and its overshoot by

$$\widehat{\tau}_d = \inf\left\{n \geq 0 : S_n^{(\lambda)} \leq -d\right\}, \qquad \widehat{O}_d = -d - S_{\widehat{\tau}_d}^{(\lambda)}.$$

Under $\widetilde{Q}_\lambda$, the random walk $-S_n^{(\lambda)}$ has positive drift and satisfies the moment condition above. The uniform mean-overshoot bound for positive-drift random walks (Lorden, 1970) therefore gives a constant $K_\lambda < \infty$ such that

$$\sup_{d \geq 0} \widetilde{\mathbb{E}}_\lambda \left(\widehat{O}_d\right) \leq K_\lambda. \tag{48}$$

Let

$$M_\lambda = 2K_\lambda + 1.$$

By Markov's inequality,

$$\inf_{d \geq 0} \widetilde{\mathbb{P}}_\lambda \left(\widehat{O}_d \leq M_\lambda\right) \geq 1 - \frac{K_\lambda}{M_\lambda} > \frac{1}{2}. \tag{49}$$

Moreover,

$$\exp\left\{\kappa(\lambda) S_n^{(\lambda)}\right\}$$

is a mean-one nonnegative martingale under $\widetilde{Q}_\lambda$. Therefore,

$$\widetilde{\mathbb{P}}_\lambda \left(\sup_{n \geq 0} S_n^{(\lambda)} \geq U_\alpha\right) \leq e^{-\kappa(\lambda) U_\alpha}.$$

For sufficiently small α, the right-hand side is less than $1/4$. Combining this bound with equation 49 gives

$$\begin{aligned}
&\widetilde{\mathbb{P}}_\lambda \left(\widehat{O}_{b_\alpha} \leq M_\lambda,\, \widehat{\tau}_{b_\alpha} < \tau_{U_\alpha}^+\right) \\
&\quad \geq \widetilde{\mathbb{P}}_\lambda \left(\widehat{O}_{b_\alpha} \leq M_\lambda\right) - \widetilde{\mathbb{P}}_\lambda \left(\tau_{U_\alpha}^+ < \infty\right) \\
&\quad \geq \frac{1}{4}.
\end{aligned}$$

Substituting this event into equation 45 gives

$$q_{\alpha,\lambda}^{\mathrm{w}}(b_\alpha) \geq \frac{1}{4} \exp\left\{-\kappa(\lambda)\left(b_\alpha + M_\lambda\right)\right\}. \tag{50}$$

The bounds hold for both choices of b_α, with the same M_λ. By equation 43,

$$\frac{1}{4} e^{-\kappa(\lambda)(d_{\alpha,\lambda} + h + M_\lambda)} \leq q_{\alpha,\lambda} \leq e^{-\kappa(\lambda) d_{\alpha,\lambda}}.$$

Since h is fixed, equation 42 gives

$$-\frac{1}{U_\alpha} \log q_{\alpha,\lambda} \longrightarrow I_r(\lambda) := \kappa(\lambda) \left\{\frac{r g(\lambda)}{\gamma^*} - 1\right\}. \tag{51}$$

Step 3: comparison with the untempered process. At $\lambda = 1$,

$$I_r(1) = r - 1.$$

Using equation 30 and equation 41,

$$I_r'(1) = \kappa'(1)(r-1) + \kappa(1)\frac{rg'(1)}{\gamma^*}$$
$$= -(r-1)\left(1 + \frac{\gamma^*}{\delta^*}\right) < 0.$$

By continuity, there exists $\varepsilon \in (0, \varepsilon_1)$ such that

$$I_r(\lambda) > I_r(1), \qquad \lambda \in (1-\varepsilon, 1). \tag{52}$$

Fix any λ in this interval. Applying equation 51 at λ and at 1 gives

$$\frac{1}{U_\alpha} \log\left\{\frac{1-\pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(\lambda)}{1-\pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(1)}\right\} \longrightarrow -\{I_r(\lambda) - I_r(1)\} < 0.$$

Since $U_\alpha \to \infty$, the ratio inside the logarithm converges to zero. Consequently,

$$1 - \pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(\lambda) < 1 - \pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(1)$$

for all sufficiently small α, or equivalently,

$$\pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(\lambda) > \pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(1).$$

Thus $W^{(1)}$ does not maximize the principal's success probability over the tempered oracle family. □

A.4 Examples for Remark 4.6

Consider

$$Q = N(\mu, 1), \qquad \mathcal{P}_0 = \mathrm{conv}\{N(\theta, 1) : \theta \le 0\}, \qquad \mu > 0.$$

Here $P_0^* = N(0,1)$ and

$$\ell^*(x) = \mu x - \frac{\mu^2}{2}.$$

For $X \sim N(\theta, 1)$,

$$\log \mathbb{E}\left[\exp\{\lambda \ell^*(X)\}\right] = \lambda\mu\theta + \frac{\mu^2}{2}(\lambda^2 - \lambda).$$

For every $\lambda > 0$, this expression is maximized over $\theta \le 0$ at $\theta = 0$. By linearity of expectation, the same is true over the convex hull. Hence equation 2 holds for every $\lambda > 0$.

As a second example, let

$$Q = \mathrm{Bernoulli}(q), \qquad \mathcal{P}_0 = \{\mathrm{Bernoulli}(p) : 0 \le p \le p_0\}, \qquad 0 < p_0 < q < 1.$$

Then $P_0^* = \mathrm{Bernoulli}(p_0)$, and

$$\mathbb{E}_{\mathrm{Bernoulli}(p)}\left[\{L^*(X)\}^\lambda\right] = p\left(\frac{q}{p_0}\right)^\lambda + (1-p)\left(\frac{1-q}{1-p_0}\right)^\lambda.$$

Because $q > p_0$, the right-hand side is increasing in p for every $\lambda > 0$. It is therefore maximized at $p = p_0$, and equation 2 again holds for every $\lambda > 0$.

A.5 A counterexample without local stability

On a three-point space, let

$$P^* = (1/4, 1/2, 1/4), \qquad Q = (1/8, 1/2, 3/8), \qquad P_1 = (0, 1, 0), \qquad \mathcal{P}_0 = \mathrm{conv}\{P^*, P_1\}.$$

For $P_t = (1-t)P^* + tP_1$,

$$d_{\mathrm{KL}}(Q\|P_t) = d_{\mathrm{KL}}(Q\|P^*) - \frac{1}{2}\log(1-t^2), \qquad 0 \le t < 1.$$

The divergence is infinite at $t = 1$, so P^* is the reverse information projection. Here $L^* = (1/2, 1, 3/2)$. For $0 < \lambda < 1$, strict concavity gives

$$\mathbb{E}_{P^*}\{(L^*)^\lambda\} < 1, \qquad \mathbb{E}_{P_1}\{(L^*)^\lambda\} = 1.$$

Consequently, $\psi_0(\lambda \ell^*) = 0 > A(\lambda)$, and

$$W_n^{(\lambda)} = (W_n^{(1)})^\lambda.$$

Thus tempering raises the rejection threshold on the original log-likelihood scale to U_α/λ, and its drift is $\lambda\gamma^*$, with a first-order loss at $\lambda = 1$. This illustrates the substantive role of local stability.

Under the same scaling $\rho_\alpha \gamma^*/U_\alpha \to r > 1$, tempering is detrimental in this example. The increments are bounded, and their positive adjustment coefficient is $\kappa(\lambda) = 1/\lambda$, since $\mathbb{E}_Q e^{-\ell^*} = 1$. The boundary and change-of-measure arguments in Appendix A.3 therefore give, for every fixed $\lambda \in (1/r, 1]$,

$$-\frac{1}{U_\alpha} \log\{1 - \pi^{\mathrm{DV}}_{\alpha,\rho_\alpha}(\lambda)\} \longrightarrow r - \frac{1}{\lambda}.$$

This is strictly smaller than $r - 1$ when $\lambda < 1$, so such tempering eventually decreases the principal's success probability.

B Numerical solution of the agent's stopping problem

We briefly describe the numerical procedure used in Section 5. For a fixed e-process and reward–cost ratio, let $Z_t = \log E_t$, let $U = \log(1/\alpha)$, and let Y denote one log-increment under the alternative Q. After normalizing the sampling cost to one, the agent's value function satisfies

$$v(z) = \begin{cases} \rho, & z \geq U, \\ \max\{0,\, -1 + \mathbb{E}_Q[v(z+Y)]\}, & z < U. \end{cases} \tag{53}$$

Gaussian spatial recursion. For the Gaussian calculations, we discretize the log-evidence state z on a uniform grid $z_1 < \cdots < z_m < U$. Starting from $v^{(0)}(z_j) = 0$, we apply value iteration,

$$v^{(k+1)}(z_j) = \max\left\{0,\, -1 + \mathbb{E}_Q\left[v^{(k)}(z_j + Y)\right]\right\}, \tag{54}$$

where values between grid points are obtained by linear interpolation and $v(z) = \rho$ for $z \geq U$. Iteration continues until the supremum norm difference between successive iterates falls below a prespecified tolerance.

The optimal continuation region is then

$$\mathcal{C} = \{z < U : -1 + \mathbb{E}_Q[v(z+Y)] > 0\}.$$

Given this region, the success probability $h(z)$ is obtained from

$$h(z) = \begin{cases} 1, & z \geq U, \\ 0, & z \notin \mathcal{C},\ z < U \\ \mathbb{E}_Q[h(z+Y)], & z \in \mathcal{C}. \end{cases} \tag{55}$$

We solve equation 55 by fixed-point iteration on the same grid. The reported success probability is $h(0)$.

The expectation in equation 54 is evaluated using the Gaussian transition law. The selected tempering parameter $\widehat{\lambda}$ maximizes the computed success probability on the stated finite λ-grid.

Bernoulli count-state bounds. For Bernoulli observations, write the two increments as y_+ and y_-. At time n, after k successes, the exact log-evidence is

$$z_{n,k} = ky_+ + (n-k)y_-.$$

The two child states are $(n+1, k+1)$ and $(n+1, k)$, with probabilities q and $1-q$. We use these transitions directly, with the same reward ρ, unit sampling cost, and rejection condition $z_{n,k} \geq U$ as in equation 53.

At a computational horizon H, unrejected states receive remaining-value bounds $v_H^- = 0$, $v_H^+ = \rho$, and success bounds $p_H^- = 0$, $p_H^+ = 1$. At each preceding unrejected state, set

$$C_{n,k}^{\pm} = -1 + qv_{n+1,k+1}^{\pm} + (1-q)v_{n+1,k}^{\pm}, \qquad v_{n,k}^{\pm} = \max\{0, C_{n,k}^{\pm}\},$$

and propagate success bounds by

$$p_{n,k}^{\pm} = \mathbf{1}\{C_{n,k}^{\pm} > 0\}\{qp_{n+1,k+1}^{\pm} + (1-q)p_{n+1,k}^{\pm}\}.$$

Every rejected state instead has $v^- = v^+ = \rho$ and $p^- = p^+ = 1$. A positive lower continuation value guarantees continuation; a nonpositive upper continuation value guarantees quitting. For an unresolved decision, zero is a lower success bound and continuing with the child upper bounds gives an upper bound. Thus the recursion bounds the infinite-horizon problem in exact arithmetic, including quitting at indifference. The horizon is a computational device, not a mandatory deadline for the agent. Absorption at rejected states prevents paths that have already succeeded from contributing further sampling costs.

We increase H until the initial success-bound width meets the reported tolerance and retain both endpoints. Calculations use ordinary double precision, not directed rounding or interval arithmetic; the computed widths are horizon diagnostics rather than numerical certificates, and a zero displayed width does not certify zero error. No spatial interpolation or lower spatial truncation is used for the Bernoulli results. For each parameter pair, the Gaussian method uses $O(Km \log m)$ time and $O(m)$ working memory for m grid nodes and K total iterations. Layerwise Bernoulli recursion uses $O(H^2)$ time and $O(H)$ working memory; horizon doubling preserves these orders at the final horizon.

B.1 Additional numerical validation

The following calculations provide additional accuracy and sensitivity checks. All use $\alpha = 10^{-3}$ and $U = \log(1/\alpha) = 6.907755$; figure-specific settings appear in Appendix B.2.

The Gaussian validation in revision_numerics.R uses remaining-distance nodes $x_j = jU/N$, so $x = U$ is represented exactly. Its base domain is $0 \leq x \leq 4U$, equivalently $-3U \leq z \leq U$, with lower endpoint -20.723266. We set $v = \rho$ and $h = 1$ for $z \geq U$, and $v = h = 0$ below the lower endpoint. Gaussian transition expectations integrate the piecewise-linear nodal interpolant analytically using the normal CDF and density; zero-padded FFT convolution evaluates all interior node offsets without a quadrature cutoff. Writing $\bar{v}$ for the interpolant on $[-3U, U]$, including its terminal node $\bar{v}(U) = \rho$, the transition expectation is

$$\int_{-3U}^{U} \bar{v}(y)\, \frac{\phi((y-z-g_\lambda)/\sigma_\lambda)}{\sigma_\lambda}\, dy + \rho\left[1 - \Phi\left(\frac{U-z-g_\lambda}{\sigma_\lambda}\right)\right],$$

where $\sigma_\lambda = \lambda\mu$ and ϕ, Φ are the standard normal density and CDF. The tail and the terminal-node interpolation contribution on the last grid interval are added analytically to the interior convolution. The lower extension is zero. Iterations start at zero, with successive supremum-norm tolerances 10^{-10} for values and 10^{-12} for probabilities, and a limit of 20,000 iterations. Recomputed residuals measure satisfaction of the discrete equations; without a stability bound they are not bounds on value or probability error, and they do not measure spatial discretization error. In the Gaussian rows of Table 1, $r = 1.25$, $\mu = 1$, and $\lambda_r = 0.825211$; the state spacings U/N are 0.0134917, 0.00674585, and 0.00337293.

The Bernoulli rows use count-state bounds with $(p_0, q) = (0.3, 0.5)$ and $r = 1.3$, computed by Bernoulli_summaries_countstate.R. For both models, $\widehat{\lambda}$ is selected from $\{0.5, 0.5+$

$\Delta\lambda, \ldots, 1\}$. The Table 1 scan at $r = 1.3$ and $H = 1000$ has maximum computed probability-bound width 6.76×10^{-12}, and distinguishes each reported grid maximizer from its runner-up. The baseline and selected probabilities are reevaluated at the listed horizons H; checks through $H = 4000$ agree at the displayed precision. All per-parameter results and bound endpoints are saved.

Model	N or H	$\Delta\lambda$	$\widehat{\lambda}$	$\pi(1)$	$\pi(\lambda_r)$	$\pi(\widehat{\lambda})$
Gaussian	512	0.0100	0.830	0.880168	0.905288	0.907025
Gaussian	1024	0.0050	0.850	0.879768	0.905741	0.906795
Gaussian	2048	0.0025	0.845	0.879971	0.905518	0.906348
Bernoulli	1000	0.0100	0.8000	0.944162	—	0.964149
Bernoulli	2000	0.0050	0.8050	0.944162	—	0.964667
Bernoulli	4000	0.0025	0.8025	0.944162	—	0.964699

Table 1: Gaussian spatial-grid resolution N and Bernoulli horizon H for the reported baseline and selected probabilities, with finite tempering-grid refinement. The separately evaluated λ_r is the Gaussian reference from Corollary 4.3; dashes denote that this reference is not used for the Bernoulli rows. Each probability incorporates the agent's recomputed response; Bernoulli bounds agree to the displayed precision.

Between the last two Gaussian grids, the largest displayed probability change is 4.48×10^{-4}, and the selected parameter changes from 0.850 to 0.845. The finest-grid gain is 0.026377, changing by 0.000650 in the last refinement, and reduces failure probability by 21.98%. At $N \geq 1024$, extending the lower endpoint to $-5U = -34.538776$ at fixed spacing changes the checked Gaussian probabilities by less than 3×10^{-14}. Continuation ends inside both domains. These checks assess spatial and domain sensitivity; they do not certify a continuous-parameter optimum. Because Table 1 changes both N and $\Delta\lambda$, shifts in $\widehat{\lambda}$ cannot be attributed to either refinement alone. The fixed-parameter checks in Appendix B.2 assess spatial sensitivity separately.

For Bernoulli observations, the finest parameter grid selects 0.8025, with gain 0.020537 and relative failure reduction 36.78%; refining from $\Delta\lambda = 0.005$ increases the selected success probability by 3.23×10^{-5}. For the fixed choice $\lambda = 0.8$, the gain is 0.019987 and the relative failure reduction is 35.80%. An independent spatial-interpolation calculation at $N = 2048$ gives 0.964567 for this fixed choice, versus 0.964149 from count states, a difference of 4.18×10^{-4}. Small changes between nearby spatial grids can therefore understate the remaining Bernoulli error; all primary Bernoulli probabilities and selections use count states. None of the finite-grid selections identifies the optimum over every real λ.

μ	$\widehat{\lambda}$	$\pi(1)$	$\pi(\lambda_r)$	$\pi(\widehat{\lambda})$	Δ_r	F_r
0.5	0.840	0.912253	0.935270	0.935770	0.023017	26.23%
1.0	0.850	0.879768	0.905741	0.906795	0.025973	21.60%
2.0	0.835	0.829255	0.863769	0.864390	0.034515	20.21%

Table 2: Gaussian signal-strength sensitivity at $\alpha = 10^{-3}$, $r = 1.25$, $N = 1024$, and $\Delta\lambda = 0.005$. Here $\lambda_r = 0.825211$ for all three signals, $\Delta_r = \pi(\lambda_r) - \pi(1)$, and $F_r = \Delta_r / \{1 - \pi(1)\}$ is the relative failure reduction. Differences and ratios are computed before rounding.

Table 2 compares the same exponent-optimal rule across signal strengths, with rewards adjusted to hold r fixed. Its absolute gains are 2.30–3.45 percentage points and its relative failure reductions are 20.21–26.23%. The computed finite-grid choices improve success over λ_r by at most 0.00106 in these cases. Thus the exponent supplies a useful finite-threshold reference here, without identifying the exact finite-threshold optimizer.

B.2 Reproducible settings for the updated figures

Figure 1 uses $\alpha = 10^{-3}$, $N = 1024$, and the Gaussian domain and transition interpolation above. Its tempering grid is $\{0.25, 0.255, \ldots, 1\}$, and its reward grid is $\{1.02, 1.04, \ldots, 1.80\}$,

supplemented with the highlighted rewards. It recomputes λ_r for each r separately from the grid optimizer. Figure 2 uses $U \in \{\log 10, 3, 4, \ldots, 10, 12, 14, \ldots, 24, 25\}$, with $\log 1000$ added, and $N = \lceil U/0.01 \rceil$. Its choices are 1, 0.9, and $\lambda_{1.25}$; failure probabilities are solved directly without a positive numerical floor before taking logarithms. These Gaussian runs use value and probability tolerances 10^{-10} and 10^{-12} and check the raw continuation policy for monotonicity. Figure 1's warm starts across increasing rewards are checked against cold starts. After doubling N, the largest fixed-parameter success changes in the scripted checks are 1.50×10^{-3} and 3.30×10^{-4} for Figures 1 and 2, respectively; checked domain-extension changes are below 3×10^{-12}. Across Figure 2 and its refinement checks, the value, success, and direct-failure iterations required at most 267, 272, and 152 steps, respectively. The direct-failure linear-equation residual was at most 8.40×10^{-13}, with the residual qualification above.

Figure 3 uses the count-state method with $\alpha = 10^{-3}$, $p_0 = 0.3$, $q = 0.5$, and the same λ-grid $\{0.25, 0.255, \ldots, 1\}$. Its reward grid is $\{1.02, 1.04, \ldots, 1.80\}$, supplemented with 1.05, 1.15, and 1.30. Each parameter pair (r, λ) begins at horizon 1000, with doubling up to 8000 if needed to reach initial probability-bound width at most 10^{-10}; unresolved parameter pairs cause the script to stop. The script exports both probability endpoints, horizon checks, and the separation of the selected grid maximum from competing choices. The plotted probabilities are the midpoints of the computed bounds. These are ordinary double-precision calculations, with the qualification about bounds stated above. All 6342 cases met the target (four required $H = 2000$); the maximum width across this full figure grid was 9.32×10^{-11}. All 42 grid maxima were separated, with minimum runner-up gap above 4.79×10^{-8}. Doubling the horizons for the LR and selected choices changed midpoints by at most 2.99×10^{-11}.

At the highlighted rewards 1.05, 1.15, and 1.30, the selected grid choices are 0.945, 0.875, and 0.805, with gains 0.005863, 0.017578, and 0.020505, respectively. In the near-participation check $r = 1.05$, $\lambda = 0.88$, the count-state probabilities are 0.671036993 and 0.665968500 for the tempered and untempered processes, giving gain 0.005068493. Curves join raw values without smoothing.