Keeping Experts Honest: The Necessity of Override*

Ricardo Fonseca†

July 2026

Abstract

A principal commits to a dynamic mechanism and repeatedly decides whether to fund projects after consulting an expert biased toward approval. Transfers are unavailable, and realized project quality never reveals the signal observed by the expert, so future decisions become the instrument for rewarding honest advice. I begin from an unrestricted measurable mechanism class and prove an exact reduction: when the expert's degree of bias is known, every attainable outcome can be implemented by a finite recursive direct mechanism with binary reports. Mechanisms that never fund after unfavorable advice on the truthful path face a payoff ceiling that remains bounded away from first best, no matter how patient the principal becomes. Unrestricted mechanisms, by contrast, approach first best. Above an explicit threshold, every globally optimal mechanism must sometimes override unfavorable advice, meaning fund after a truthful negative recommendation. Thus, retained discretion has a nonvanishing incentive value even when override occurs only rarely on the equilibrium path. I characterize the minimum-funding frontier that governs the delivery of promised decision authority and extend the analysis to privately known expert bias, where bad-project exposure becomes the relevant screening allocation.

Keywords: dynamic mechanism design; repeated communication; biased advice; nonmonetary incentives; screening.
JEL codes: C73, D82, D83, D86.

1 Introduction

Organizations often rely on experts whose close involvement gives them valuable information but may also tilt their judgments. Scientific funding makes this tradeoff concrete. Li (2017) finds that National Institutes of Health (NIH) reviewers who are intellectually closer to an application are both more favorable toward it and better able to distinguish high- from low-quality projects. In her setting, the informational benefit of expertise weakly dominates the cost of bias, so blanket recusal would discard valuable information together with the conflict.

*I am grateful for comments from Jack Fanning, Teddy Mekonnen, Bobby Pakzad-Hurson, Kareen Rozen, Roberto Serrano, and seminar participants. All remaining errors are my own.

†Department of Economics, Pontificia Universidad Javeriana. E-mail: baricardo@javeriana.edu.co.

Review systems also differ in how tightly final decisions track expert evaluations. In the historical arrangements motivating the application, NIH study-section scores were close to decisive, while the National Science Foundation (NSF) inserted a program-officer recommendation and senior administrative review between external evaluation and award (Li, 2017; National Science Foundation, 2016). Section 8 documents this contrast in detail. It isolates the margin studied here: does an unfavorable evaluation nearly determine the action, or does the organization retain room to approve despite it?

That second architecture raises a question distinct from ordinary delegation. An expert may truthfully report that the evidence is unfavorable while still preferring the project to be funded. Why would an organization ever choose the expert's preferred action after receiving her truthful bad news? Such decisions are usually attributed to superior administrative information, programmatic objectives, favoritism, or error. This paper develops a different possibility: retained discretion can be part of the incentive system that makes unfavorable advice credible in the first place.

Motivated by this tension, I study a research fund that repeatedly consults the same expert over an infinite discounted horizon. Each period brings a new project of binary quality, and the expert privately observes a binary signal before recommending whether it should be funded. The expert is biased toward approval, while the principal values funding only sufficiently promising projects. If a project is funded, realized quality never reveals which signal the expert observed. The principal can commit at the outset to a public dynamic mechanism, but transfers are unavailable. The expert may be an individual adviser or a standing review unit with overlapping membership and a stable mandate; for expositional simplicity, I use feminine pronouns for either interpretation.[1]

Because the interaction repeats, the fund can make future influence depend on the public record generated by current advice. To make an unfavorable recommendation credible today, it may promise the reviewer greater influence over future funding. Yet future influence is a claim on actual decisions, not a bookkeeping transfer. When that promise comes due, honoring it may require funding a proposal even after the reviewer reports unfavorable evidence. The paper asks whether such override is merely one possible incentive device or instead an unavoidable feature of optimal governance.

The paper makes two main contributions. The first is methodological. I begin from the unrestricted class of committed measurable public mechanisms and prove an exact payoff-equivalent reduction to a finite recursive direct mechanism. When the expert's degree of bias is known, binary reports suffice, promised utility is a sufficient recursive state, and a public lottery over at most three current plans implements every attainable payoff pair. The reduction preserves complete continuation behavior after every public history, including zero-probability histories. It therefore turns the unrestricted problem into a tractable recursive program without narrowing the payoff set. In particular, richer communication and public randomization cannot eliminate the need for override.

[1] Under the review-unit interpretation, the model abstracts from internal aggregation and treats the unit's recommendation and funding preference as the relevant reduced-form objects.

The second contribution is economic. Call a mechanism *advice-obedient* if it never funds after unfavorable advice on the truthful path. This restriction gives negative advice effective veto power. Advice-obedient mechanisms face a payoff ceiling that remains bounded away from first best, no matter how patient the principal becomes. Unrestricted committed mechanisms, by contrast, approach the first-best policy of funding exactly after favorable signals. Combining these facts yields the paper's central conclusion: above a unique threshold discount factor, every global optimum must sometimes *override unfavorable advice*, meaning fund after a truthful negative recommendation at a history reached on the equilibrium path. Retained discretion therefore has a nonvanishing incentive value even when override occurs only rarely.

The mechanism behind this result is an intertemporal promise. A continuation promise discourages exaggeration today, but at the same time creates a claim on future decisions. Because observed outcomes do not verify the expert's signal, the principal cannot wait to detect a lie and then cancel that claim. As patience grows, promises can be delivered at vanishing welfare cost, but they still have to be honored. Near-first-best value can coexist with override along histories reached under honest reporting.

To characterize how promises are honored, I index the expert's conflict by λ: a good funded project gives her payoff one, whereas a bad funded project gives $1-2\lambda$, so lower λ means a stronger taste for approval. I call an expert *conflicted* when she prefers funding even after an unfavorable signal. The recursive state is her promised normalized utility, interpreted as an *authority account*. For a known type, the least funding needed to deliver each account balance defines a convex frontier. Low balances are delivered by rationing favorable recommendations. As the balance rises, an exact capacity identifies the largest promise consistent with no current funding after unfavorable advice.

Beyond that capacity, some funding after unfavorable advice is unavoidable. For general conflicted types, current funding is determined by the interaction of action bounds, continuation-promise bounds, and pooling of outcome-contingent promises. At the benchmark $\lambda = 1/2$, these constraints line up, under mild patience conditions stated below, so that an optimum can be selected to move from rationing, to ordinary review (funding exactly after favorable advice), and finally to override, meaning funding despite a recommendation not to fund. These frontier results explain how a given promise is delivered. The upper-tail theorem instead concerns complete relationships and shows that override occurs on the truthful path of every sufficiently patient global optimum.

These conclusions hold for the full committed mechanism class, not only for stationary rules or finite message spaces. The principal may condition decisions on arbitrary public histories and may use rich communication and public randomization. Theorem 1 shows that every attainable pair of expert utility and total funding nevertheless has an exact recursive direct implementation with binary current reports and a public lottery over at most three current plans. These two coordinates are enough for the known-type problem because, for a fixed type, they jointly determine bad-project exposure and hence the principal's payoff. The theorem therefore bridges the unrestricted dynamic problem and a fixed finite recursive architecture without narrowing the payoff set. In particular, necessary override is a property of the unrestricted committed problem, not an artifact of the

recursion used to solve it.

The same delivery logic also creates a screening problem when the expert's degree of conflict is privately known. In that case, complete dynamic policies become menu options: each branch solves the delivery problem for some authority profile, and the menu assigns those branches across types. Two statistics organize the assignment. Total funding measures how much authority a policy delivers, whereas bad-project exposure (discounted funding directed to bad projects) measures its composition and which types value it most. Because exposure falls with alignment, it becomes the screening allocation. This observation yields three further results: the low-promise rationing segment cannot separate types, the indirect-utility envelope identifies exposure through its slope, and finite submenus approximate arbitrarily rich menus under the stated conditions.

The proof of the high-patience result mirrors this economic logic. First, a representation-independent argument gives a fixed upper bound for every advice-obedient mechanism. Second, an unrestricted construction keeps the authority account near the utility delivered by ordinary review until a rare exit, so its payoff converges to first best. Finally, monotonicity in patience extends the resulting strict comparison to every sufficiently high discount factor. Binary signals, binary quality, and independent and identically distributed (i.i.d.) arrivals keep this recursive state one-dimensional.

1.1 Related literature

The closest related paper is Rantakari (2023). In his model, an advocate observes project value, the principal observes a random outside option, and implementation subsequently reveals the project's value. His main analysis focuses on stationary truthful equilibria along the equilibrium path and disciplines detected exaggeration through reversion to uninformative communication. Rantakari also observes that, under commitment, history-dependent continuation promises may improve on stationary rules and formulates promised utility as a state variable. In the current environment, promised utility indeed provides the sufficient recursive state for solving the unrestricted committed problem. However, he leaves the principal-optimal unrestricted dynamic problem unresolved.

The present paper instead solves the unrestricted committed problem in a binary noisy-signal environment in which realized quality never reveals the expert's signal. This information structure creates a different incentive logic. In Rantakari (2023), implementation can expose exaggeration; here performance can discipline reports only statistically and can never identify a lie. As a result, the principal can approach first best even though every sufficiently patient optimum must still override unfavorable advice on path.

A related distinction concerns commitment. Rantakari (2021) studies accumulated influence in a relational environment, where the principal must find each promised rule sequentially credible. The paper derives an optimal stationary equilibrium and analyzes nonstationary influence, but it does not fully characterize the principal-optimal dynamic equilibrium. Here commitment removes that self-enforcement constraint and instead turns promised influence into a delivery obligation. The same contrast separates the present analysis from other models of relational and repeated delegation,

including Alonso and Matouschek (2007), Lipnowski and Ramos (2020), and the no-commitment benchmark of Kivinen and Kuzmics (2025).

Deb, Pai, and Said (2018) provide a complementary benchmark for dynamic evaluation without transfers. There, a principal uses forecasts about publicly observed outcomes to evaluate persistent forecasting quality, and the analysis separates a sharp binary benchmark from a robustness result for a more general signal environment. Here the friction is persistent conflict rather than forecasting ability, and the same incumbent repeatedly advises whether current projects should be funded. Because realized quality disciplines reports statistically without revealing the expert's signal, the problem becomes one of dynamically allocating decision authority rather than selecting a forecaster. The analysis characterizes the unrestricted committed optimum and shows that every sufficiently patient optimum must sometimes override unfavorable advice. The binary i.i.d. structure delivers the one-dimensional recursion and sharp frontier, while Section 8 separates those tractable features from the broader economic logic.

More broadly, several dynamic models replace transfers with future allocations. Discounted quotas in Frankel (2016) are especially close to the low-promise region here, while related mechanisms appear in Guo and Hörner (2020), Bird and Frug (2019), Bird and Frug (2025), Malenko (2019), Campbell (2021), Gupta et al. (2024), and de Clippel et al. (2021). The present model differs because the reward is a claim on a decision whose informational basis remains private. Even asymptotic efficiency need not eliminate the distortion used to honor that claim.

Methodologically, the recursive construction follows Abreu, Pearce, and Stacchetti (1990), while broader dynamic-design frameworks include Bergemann and Välimäki (2019) and Pavan, Segal, and Toikka (2014). Theorem 1 contributes an exact canonicalization result for the present full-commitment environment: unrestricted measurable protocols generate the same pairs of expert utility and funding as a fixed finite recursive direct class. Doval and Skreta (2022) also start from a broad mechanism class and prove outcome-equivalent canonicalization, but for limited commitment; their canonical mechanisms use type reports and posterior beliefs to encode the information learned by the designer. Here the canonical state is promised utility, and the additional conclusion is finite support over at most three current plans. Once bias is private, complete dynamic policies also become screening objects. That extension connects the paper to delegation under uncertain preferences (Frankel, 2014) and to nonmonetary screening (Ambrus and Egorov, 2017; Amador and Bagwell, 2020).

Interpreting promised utility as authority links the analysis to Aghion and Tirole (1997), Baker, Gibbons, and Murphy (1999), and especially Li, Matouschek, and Powell (2017). In their relational model, future authority rewards cooperative conduct because the principal later follows the agent's preferred recommendation. Here, the expert's informative report and preferred action point in opposite directions: she may truthfully recommend against funding while still preferring approval. Rewarding her with influence can therefore require the principal to fund despite her truthful recommendation not to do so. The relevant object is not simply delegated authority, but override of truthful advice. Experimental evidence likewise treats information sharing, advice, and delegation

as distinct organizational arrangements with different implications for trust and trustworthiness (Özer, Subramanian, and Wang, 2017). Scientific review provides the leading application (Li, 2017; Ham et al., 2021; Fehrler and Janas, 2020), while internal capital allocation supplies a parallel setting in which managers privately assess projects but may favor continuation (Malenko, 2019; Gupta et al., 2024).

The paper proceeds in the order suggested by this logic. Section 2 presents the model and shows why the unrestricted benchmark can be analyzed recursively. Sections 3–4 then study how promised utility is delivered, after which Section 5 compares advice-obedient and unrestricted mechanisms over the whole relationship. Section 6 makes the delivery path fully explicit in the ordered benchmark, and Section 7 turns from delivery to screening when bias is privately known. Section 8 closes with scope, organizational override, and implications. Proofs are in the Appendix.

2 Model and mechanism class

Time is discrete and infinite, with common discount factor $\delta \in (0,1)$. I write $a := 1-\delta$ and normalize each lifetime payoff by multiplying its discounted sum by a. In every period, a new project has quality $\theta \in \{0,1\}$, independently drawn with $\Pr(\theta = 1) = q$. Before the funding decision, the expert privately observes a signal $\eta \in \{0,1\}$ with accuracy $p \in (1/2, 1)$:

$$\Pr(\eta = \theta) = p.$$

After the decision, the project's quality becomes public regardless of whether this principal funded it.[2] Realized quality is correlated with the expert's private signal, but no outcome reveals whether a particular report was truthful.

Let $d \in \{0,1\}$ denote the funding decision. The principal's stage payoff is $d(2\theta - 1)$, whereas a type $\lambda \in [0,1]$ expert receives

$$d\{1 - 2\lambda(1-\theta)\}.$$

Every type receives payoff one from a good funded project, while a bad funded project yields $1-2\lambda$. Lower λ corresponds to a stronger taste for approval.

To keep advice decision-relevant after either signal, I maintain the intermediate-prior assumption

$$1 - p < q < p. \tag{A1}$$

Let

$$m_1 = pq + (1-p)(1-q), \qquad m_0 = (1-p)q + p(1-q),$$

and

$$\alpha_1 = \frac{pq}{m_1}, \qquad \alpha_0 = \frac{(1-p)q}{m_0}.$$

[2]This captures settings in which an unfunded proposal may be pursued with support from another organization or later evidence reveals the project's underlying quality.

Here $m_i = \Pr(\eta = i)$ is the probability of signal i, and $\alpha_i = \Pr(\theta = 1 \mid \eta = i)$ is the posterior probability of good quality after that signal. Then

$$0 < \alpha_0 < q < \alpha_1 < 1, \qquad \alpha_0 < \frac{1}{2} < \alpha_1, \qquad q = m_0\alpha_0 + m_1\alpha_1.$$

These primitives make promised utility the natural recursive state. I refer to that promised normalized discounted utility as the expert's *authority account*. I also use two decision labels throughout: *review* funds after a favorable report and rejects after an unfavorable report, whereas *override* assigns positive funding probability after an unfavorable report. Because quality is learned only later, override means funding after negative advice rather than knowingly funding a bad project.

2.1 The mechanism benchmark

A public history records all past public messages, funding decisions, realized qualities, and publicly observed randomization. A committed public mechanism specifies, after every public history and current message, a public lottery over the current funding action and continuation mechanism. Rules are complete after every public history, including histories reached with probability zero. Message and history spaces are standard Borel, and all kernels are measurable. A reporting strategy maps the expert's private signal and public history into a distribution over messages; an admissible outcome requires a complete sequentially optimal strategy after every finite private history.

The principal may use this unrestricted public mechanism class: histories, communication, and public lotteries need not be finite. This breadth matters because an ex ante protocol restriction would leave open whether richer mechanisms could avoid the distortion characterized below.[3] An outcome is admissible only if the selected strategy is sequentially optimal after every finite private history, including histories of zero initial probability. The benchmark therefore uses actual, attained best responses, not suprema over responses that need not exist. This attainment requirement is important because an unattained supremum could otherwise enter both the payoff set and its recursive compression. For a fixed known type λ, let $\mathcal{W}_\lambda^{SB}$ denote the set of pairs (u, A) generated by these unrestricted standard-Borel outcomes, where u is normalized expert utility and $A = a\mathbb{E}\sum_{t\geq 0}\delta^t d_t$ is normalized discounted funding. By comparison, let $\mathcal{W}_\lambda^{F}$ denote the set generated by a finite recursive class of public direct mechanisms with binary current-signal reports and a fixed ternary public label drawn before the current signal at every date. The label selects among at most three current funding-and-continuation plans, and every label, report, funding, and quality branch has a complete continuation rule.

[3] For a related canonicalization under limited commitment, see Doval and Skreta (2022), who replace general mechanism-selection games with mechanisms whose inputs are type reports and whose public outputs encode the designer's posterior beliefs. Theorem 1 addresses a different problem: under full commitment, it compresses the unrestricted outcome class here into a fixed finite recursive architecture.

2.2 A methodological foundation: payoff equivalence and finite recursion

The next theorem makes the unrestricted scope operational. Starting from arbitrary measurable public histories and message spaces, with public lotteries over current actions and continuation mechanisms, it shows that the same payoff pairs can nevertheless be generated by one fixed finite recursive class. Besides grounding the economic analysis below, the theorem supplies an exact bridge from a rich dynamic protocol problem to a tractable promised-utility recursion.

Theorem 1 (Payoff equivalence of unrestricted and recursive mechanisms)**.** *For every fixed known type λ,*

$$\mathcal{W}_\lambda^{SB} = \mathcal{W}_\lambda^{F}.$$

Rich measurable communication and public randomization therefore neither lower the minimum-funding frontier nor raise the optimal known-type payoff. Moreover, every attainable payoff pair has a binary-report implementation with a public lottery over at most three current funding-and-continuation plans. Its countable formal history tree carries complete continuation rules, including on zero-probability branches.

Necessary override should not arise because the analysis imposed stationarity, finite communication, or a recursive mechanism class in advance. Theorem 1 rules out that concern. Every payoff pair attainable under the full measurable benchmark has a binary-report implementation with a public lottery over at most three current plans. The support bound reflects convexification in two payoff coordinates. The theorem gives an exact payoff equivalence, including complete continuation rules after null histories.

The theorem goes beyond a one-period revelation principle: it preserves the complete dynamic payoff set and sequential optimality on the full history tree. The paper can therefore solve the finite recursion while retaining conclusions about the unrestricted committed class. This is the only step that uses the measurable apparatus; from the funding frontier onward, all economic characterizations are established within the finite recursion. In particular, richer communication and public randomization cannot provide a way around necessary override. The Appendix proves the theorem and supplies complete continuation rules after every public history, including histories reached with probability zero.

3 The funding-cost frontier

Under truthful reporting, define normalized discounted funding and bad-project exposure by[4]

$$A = a\mathbb{E}\sum_{t\geq 0}\delta^t d_t, \qquad B = a\mathbb{E}\sum_{t\geq 0}\delta^t d_t(1-\theta_t).$$

[4]Expectations are taken over project qualities, private signals, and any public randomization induced by the mechanism, conditional on the stated initial promise.

The expert and principal receive

$$U_\lambda = A - 2\lambda B, \qquad V = A - 2B.$$

Here A measures how much funding the mechanism delivers, whereas B records how much of that discounted funding is directed to bad projects.

Let

$$h := 2\lambda - 1, \qquad s_i(\lambda) := \alpha_i - h(1 - \alpha_i) = 1 - 2\lambda(1 - \alpha_i).$$

The quantity $s_i(\lambda)$ is the expert's expected payoff from funding after signal i.

Definition 1 (Conflicted type)**.** A type is conflicted if $s_0(\lambda) > 0$. Under (A1), $s_1(\lambda) > 0$ for every $\lambda \leq 1$, so conflict concerns only the unfavorable signal.

For a conflicted type, funding after either signal raises the expert's payoff. Full funding therefore delivers the maximal promise $\bar{u}_\lambda$, while public lotteries span the interval between silence and full funding. By contrast, if $s_0(\lambda) \leq 0$, funding after an unfavorable signal is not a reward, so full funding need not maximize expert utility. Proposition 3 treats that aligned region separately.

For the remainder of the delivery analysis, fix a conflicted known type and define

$$C_\lambda(u) := \inf\{A : \text{an admissible mechanism delivers expert utility } u\}.$$

3.1 The frontier

The first result shows that, for every feasible promise, the minimum funding cost is attained by an admissible mechanism. The frontier is convex, recursively implementable, and defined over the entire feasible authority interval.

Proposition 1 (Frontier foundation)**.** *For every conflicted known type, the minimum-funding frontier is well defined and attained on* $[0, \bar{u}_\lambda]$*, where*

$$\bar{u}_\lambda = q - h(1 - q) = 1 - 2\lambda(1 - q).$$

It is continuous and convex, with $C_\lambda(0) = 0$ *and* $C_\lambda(\bar{u}_\lambda) = 1$*. Every interior subgradient lies between* $1/s_1(\lambda)$ *and* $1/s_0(\lambda)$*. The frontier satisfies the one-period recursion below, and every initial promise admits an optimal truthful policy on a countable formal history tree with complete zero-probability-branch rules.*

Combining the endpoint values with the subgradient bounds gives the following global lower bounds:

$$C_\lambda(u) \geq \frac{u}{s_1(\lambda)}, \qquad C_\lambda(u) \geq 1 - \frac{\bar{u}_\lambda - u}{s_0(\lambda)}. \tag{1}$$

3.2 The funding-cost transformation

The frontier is useful because, once expert utility is fixed, total funding also determines bad-project exposure. The principal's conditional problem can therefore be rewritten as minimizing funding subject to delivering the promise. Let $P_\lambda(u)$ denote the principal's highest payoff among mechanisms that deliver expert utility u.

Proposition 2 (Funding-cost transformation)**.** *Fix $\lambda \in (0,1)$. Conditional on delivering expert utility u, maximizing the principal's payoff is equivalent to minimizing discounted funding. In particular,*

$$P_\lambda(u) = \frac{u - (1-\lambda)C_\lambda(u)}{\lambda}, \tag{2}$$

and the optimal known-type payoff is the maximum of this expression over $u \in [0, \bar{u}_\lambda]$.

To see how the frontier reproduces itself over time, let ρ_r be current funding after report r, and let w_{r0} and w_{r1} be continuation promises after bad and good realized quality. Each report therefore determines both immediate funding and, once quality is observed, a new authority balance. The cost of a one-period decomposition is current expected funding plus the expected continuation cost evaluated on the same frontier. At the same time, promise keeping averages the report-contingent utilities under truthful signals, while the reporting constraints require each signal to prefer its assigned branch. It is convenient to collect current funding and continuation promises in the scores

$$x_r = -ah\rho_r + \delta w_{r0}, \qquad y_r = a\rho_r + \delta w_{r1}. \tag{3}$$

A signal-i expert who chooses report r receives $(1-\alpha_i)x_r + \alpha_i y_r$. The frontier obeys

$$C_\lambda(u) = \min_{\rho_r, x_r, y_r} \sum_{r=0}^{1} m_r \Bigg[a\rho_r + \delta(1-\alpha_r) C_\lambda\!\left(\frac{x_r + ah\rho_r}{\delta}\right) + \delta\alpha_r C_\lambda\!\left(\frac{y_r - a\rho_r}{\delta}\right) \Bigg], \tag{4}$$

subject to promise keeping, the two reporting constraints, $0 \le \rho_r \le 1$, and continuation promises in $[0, \bar{u}_\lambda]$.

The display suppresses the constraints, but writing them explicitly clarifies the geometry. Promise keeping is

$$u = \sum_{i=0}^{1} m_i \big[(1-\alpha_i)x_i + \alpha_i y_i\big],$$

and truth telling requires

$$(1-\alpha_0)x_0 + \alpha_0 y_0 \ge (1-\alpha_0)x_1 + \alpha_0 y_1,$$

$$(1-\alpha_1)x_1 + \alpha_1 y_1 \ge (1-\alpha_1)x_0 + \alpha_1 y_0.$$

These inequalities allow the two report branches to be viewed as affine payoff lines in the posterior α: truthfulness assigns the low posterior to the report-0 line and the high posterior to the report-1 line. The continuation arguments of C_λ in (4), in turn, simply invert (3). After report r, bad quality leaves promise $(x_r + ah\rho_r)/\delta$, whereas good quality leaves promise $(y_r - a\rho_r)/\delta$. Equation (4) is therefore not a relaxation of the original mechanism problem. Rather, it is a one-period decomposition of that same problem, with current funding paid immediately and every future obligation priced on the same minimum-funding frontier.

The same representation also explains the slope bounds in Proposition 1. If one additional unit of utility were delivered entirely through favorable-signal funding, it would cost $1/s_1(\lambda)$; if it were delivered entirely through unfavorable-signal funding, it would cost $1/s_0(\lambda)$. Convexity therefore places every interior marginal cost between these two extremes. Although the formal proof requires recursive self-generation and off-path completion, the economics of the frontier is already visible in the one-period program.

With this recursion in hand, the next section characterizes how different promise levels are delivered. Later, frontier attainment and convexity will also support the comparison of mechanism classes, for which no complete phase ordering is needed.

4 Delivering promised utility

The frontier turns the dynamic mechanism problem into a delivery problem: for each authority balance, the principal asks how little funding is needed to honor it. The answer changes with the size of the promise. Low balances lead to rationing, while a no-override capacity marks the largest balance that can still be delivered without current funding after unfavorable advice. Above the low region, a pointwise obstacle map governs conflicted types with $\lambda > 1/2$. The Appendix derives its exact boundaries; their global order, however, is not needed for the comparison in Section 5.

Under (A1), define

$$\bar{\lambda}(p, q) = \frac{1}{2(1 - \alpha_0)}.$$

4.1 Static alignment

The dynamic problem disappears when the expert dislikes bad funding sufficiently. In that case, she and the principal agree after both signals, so static recommendation-following is first best.

Proposition 3 (Static alignment threshold)**.** *If $\lambda \geq \bar{\lambda}(p, q)$, the expert and principal agree on funding if and only if the signal is favorable. The static first-best rule is truthful and optimal for every δ. If $\lambda < \bar{\lambda}(p, q)$, the expert wants funding after both signals and dynamic incentives may be valuable.*

The sharper delivery results use two patience conditions:

$$\delta q \geq a, \tag{A2}$$

and

$$\delta q \geq \alpha_0. \tag{A3}$$

Condition (A3) uniformly guarantees the low-authority construction throughout the conflicted region with $\lambda \geq 1/2$; for a fixed type, the primitive feasibility requirement is only $\delta \bar{u}_\lambda \geq s_0(\lambda)$. Condition (A2), by contrast, is used solely in the compatibility argument for the ordered benchmark.

4.2 Low promises

At small promises, favorable-signal funding is the cheapest way to deliver expert utility. The principal rations funding after favorable advice, rejects after unfavorable advice, and uses continuation utility to maintain truthful reporting.

Proposition 4 (General low-authority frontier)**.** *For every conflicted type $\lambda \in [1/2, \bar{\lambda})$, condition (A3) implies*

$$C_\lambda(u) = \frac{u}{s_1(\lambda)} \qquad 0 \leq u \leq a\bar{u}_\lambda.$$

An optimal implementation has

$$\rho_1(v) = \frac{v}{a\bar{u}_\lambda}, \qquad \rho_0(v) = 0,$$

zero continuation after report 1*, and the common continuation promise*

$$z(v) = \frac{s_0(\lambda)}{\delta \bar{u}_\lambda} v$$

after report 0 *and either quality realization.*

The endpoint of this linear segment is $u_L(\lambda) := a\bar{u}_\lambda$. Below it, favorable advice is rationed and unfavorable advice is rejected; at the endpoint itself, ordinary review becomes optimal.

Intermediate promises. Once the promise exceeds the low segment, the delivery problem becomes more intricate. For conflicted $\lambda > 1/2$, current funding can be selected at a current-action bound, a continuation bound, or the point at which the two outcome-contingent promises pool. Which obstacle binds need not vary monotonically with the promise. Still, one branch-level simplification survives: conditional on the low-signal payoff, the cheapest unfavorable-report branch pools its continuations and begins current funding only when that pooled promise reaches its upper bound. Whether this replacement fits the full mechanism then depends on the high-signal reporting constraint. The Appendix proves the resulting boundary map.

Even without a universal ordering of the intermediate obstacles, the low segment can be combined with an exact global capacity for policies that currently reject after unfavorable advice.

Theorem 2 (Authority at low and high promises)**.** *Under (A3), define*

$$u_L(\lambda) = a\bar{u}_\lambda, \qquad \widehat{u}_O(\lambda) = \delta \bar{u}_\lambda + am_1 \frac{\alpha_1 - \alpha_0}{1 - \alpha_0}.$$

Fix a conflicted type $\lambda \in [1/2, \bar{\lambda})$. Then $u_L(\lambda) < \widehat{u}_O(\lambda) < \bar{u}_\lambda$. Moreover:

(i) for $0 < u < u_L(\lambda)$, an explicit optimal rationing implementation exists;

(ii) at $u = u_L(\lambda)$, ordinary review is optimal;

(iii) if $\lambda > 1/2$, the pointwise boundary map summarized above applies for every $u > u_L(\lambda)$; if $\lambda = 1/2$, Section 6 gives the sharper ordered selection;

(iv) a truthful policy with zero current funding after the unfavorable signal exists for every $u \le \widehat{u}_O(\lambda)$, while every truthful policy delivering $u > \widehat{u}_O(\lambda)$ has positive current funding after that signal.

The threshold $\widehat{u}_O(\lambda)$ is the exact feasibility capacity of current no-override policies.[5] *A minimum-funding policy may start overriding below it.*

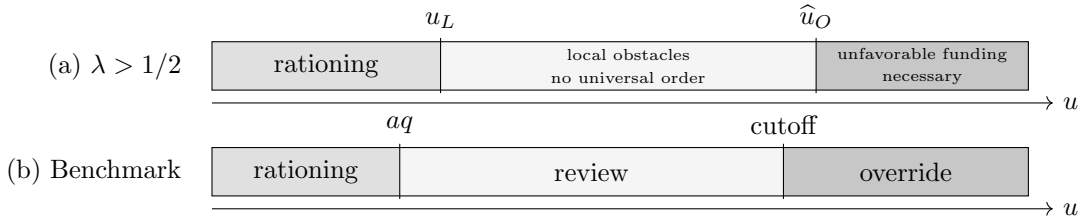

Figure 1: Authority delivery. Panel (a), for conflicted $\lambda > 1/2$ under (A3), shows the exact low rationing region, the pointwise obstacle map, and the current no-override capacity $\widehat{u}_O$. Every truthful policy delivering a promise above that line assigns positive current funding after the unfavorable signal, although a cost-minimizing policy may do so earlier. Panel (b) shows the benchmark optimal selection under (A2)–(A3): rationing, review, then override, up to boundary ties. Horizontal distances are schematic.

Initial promises. The low segment is useful off path, but it cannot contain the optimal initial promise. Indeed, the principal's objective is strictly increasing throughout that segment. Under (A3),

$$P_\lambda(u) = \frac{2\alpha_1 - 1}{s_1(\lambda)} u, \qquad 0 \le u \le a\bar{u}_\lambda,$$

Therefore, every optimal initial promise satisfies $u^*_\lambda \ge a\bar{u}_\lambda$. The relationship does not begin inside rationing, even though low promises remain useful as punishments after later histories.

5 Patience and necessary on-path override

The preceding results explain how to deliver a promise at a given history. The central welfare question, however, compares complete relationships: how well can advice-obedient mechanisms

[5] The displayed formula uses (A3). The Appendix derives the capacity without (A3), still for conflicted types with $\lambda \ge 1/2$. The low-region construction itself requires only the type-specific feasibility condition $\delta\bar{u}_\lambda \ge s_0(\lambda)$; (A2) is used only for the compatibility argument behind the ordered $\lambda = 1/2$ benchmark.

perform relative to unrestricted ones? This comparison does not require the detailed intermediate-promise obstacle map or either of assumptions (A2)–(A3).

Let

$$H := m_1(2\alpha_1 - 1) = p + q - 1$$

be the principal's payoff from funding exactly after favorable signals. Call an admissible outcome *advice-obedient* if, under truthful play, $d_t \mathbf{1}\{\eta_t = 0\} = 0$ almost surely at every date. This is a path-law property, not a restriction at every formal public history: equivalently, conditional funding after an unfavorable signal is zero at almost every public history under the truthful path law. For a conflicted type define

$$\overline{V}_\lambda^{AO} := H\left[1 - \frac{(1-\alpha_1)s_0(\lambda)}{(1-\alpha_0)s_1(\lambda)}\right]$$

and

$$\delta_\lambda^{AO} := \frac{s_0(\lambda)}{s_0(\lambda) + m_1(\alpha_1 - \alpha_0)}.$$

Let $P_\lambda^*(\delta)$ denote the optimal known-type payoff at discount factor δ. Proposition 1, Theorem 1, and compactness of the finite recursive class imply that this value is attained for every δ.

Proposition 5 (More patience expands attainable payoffs)**.** *Fix a known type λ and $0 < \delta < \delta' < 1$. Every normalized vector consisting of expert utility, principal payoff, funding, and bad-project exposure attainable at δ is attainable at δ'. It follows that*

$$P_\lambda^*(\delta') \geq P_\lambda^*(\delta),$$

and the minimum-funding frontier weakly falls at every promise common to the two feasible domains. The same inclusion holds within the advice-obedient class.

The proof embeds the lower-discount outcome in an independent public clock. Because the clock sometimes pauses, the expert may remember signals from dates when the virtual mechanism did not advance; the construction checks sequential optimality after every such private history. It preserves normalized payoffs even though it may change the public process. Together with global attainment, this monotonicity converts one strict value comparison into a statement about every optimum at all higher discount factors.

Theorem 3 (Upper-tail threshold for necessary override)**.** *Maintain (A1) and fix a conflicted known type λ. Then:*

(i) every advice-obedient admissible outcome gives the principal at most $\overline{V}_\lambda^{AO} < H$, and this bound is attained whenever $\delta \geq \delta_\lambda^{AO}$;

(ii)

$$\lim_{\delta \uparrow 1} P_\lambda^*(\delta) = H;$$

(iii) defining

$$\delta_\lambda^\dagger := \inf\left\{\delta \in [\delta_\lambda^{AO}, 1) : P_\lambda^*(\delta) > \overline{V}_\lambda^{AO}\right\},$$

we have $\delta_\lambda^\dagger < 1$. For $\delta_\lambda^{AO} \le \delta < \delta_\lambda^\dagger$, an advice-obedient optimum exists. For every $\delta > \delta_\lambda^\dagger$, every globally optimal admissible outcome satisfies $\Pr(d_t = 1, \eta_t = 0) > 0$ for some date t under truthful play. No assertion is made at $\delta = \delta_\lambda^\dagger$.

Corollary 1 (Negative advice cannot be a veto)**.** *Fix a conflicted known type and suppose $\delta > \delta_\lambda^\dagger$. Every admissible outcome in which truthful unfavorable advice rules out funding almost surely at every date gives the principal strictly less than $P_\lambda^*(\delta)$. No globally optimal review system can grant truthful unfavorable advice an effective veto over funding.*

Proof. Such outcomes are exactly the advice-obedient outcomes bounded in Theorem 3(i), whereas part (iii) gives $P_\lambda^*(\delta) > \overline{V}_\lambda^{AO}$. □

The theorem compares two mechanism classes, and each side requires a different argument. Begin with advice obedience. Along the truthful path, every funded date then follows signal 1, so the principal's and expert's flow payoffs are proportional. The problem consequently reduces to bounding how much utility can be promised to the expert. Let $\overline{U}$ be the supremum of advice-obedient expert utility. Choose an advice-obedient outcome within $\xi > 0$ of that supremum, freeze the complete current behavior prescribed after a favorable signal, and ask a low-signal expert to imitate it; letting $\xi \downarrow 0$ at the end removes the approximation. Conditional on true signal i, denote the induced public-history law by Q_i. Because neither the mechanism nor the frozen behavior observes the true signal, the two laws differ only through realized quality, and their likelihood ratio is

$$\frac{dQ_1}{dQ_0} = \begin{cases} \dfrac{1-\alpha_1}{1-\alpha_0}, & \theta = 0, \\ \dfrac{\alpha_1}{\alpha_0}, & \theta = 1. \end{cases}$$

Write L_i^1 for the expected loss, relative to $\overline{U}$, following the imitated report-1 behavior when the true signal is i, and $L_0^0 \ge 0$ for the corresponding truthful loss after signal 0. If current funding after report 1 is x, the low-signal incentive constraint gives

$$\delta(L_0^1 - L_0^0) \ge (1-\delta)x s_0(\lambda).$$

The lower likelihood ratio implies

$$L_1^1 \ge \frac{1-\alpha_1}{1-\alpha_0} L_0^1.$$

Taken together, these inequalities show why patience cannot eliminate the advice-obedient loss. The continuation punishment needed to deter exaggeration cannot be concentrated on the deviating signal; a fixed fraction must also fall on the truthful favorable-signal branch. Combining that

spillover with promise keeping and $x \leq 1$ yields

$$\overline{U} \leq m_1 \left[s_1(\lambda) - \frac{1-\alpha_1}{1-\alpha_0} s_0(\lambda) \right].$$

Multiplying this utility bound by the fixed principal-to-expert payoff ratio after favorable signals gives $\overline{V}_\lambda^{AO}$. Crucially, the resulting ceiling is independent of δ; greater patience cannot rescue advice obedience.

Unrestricted mechanisms escape that ceiling by using the authority account itself as a public state. Let $G_\lambda = m_1 s_1(\lambda)$ denote the promise delivered by static review, and keep the state inside a small interval I around G_λ. Within that interval, use the local funding rule

$$(\rho_0(u), \rho_1(u)) = \begin{cases} (0, u/G_\lambda), & u \leq G_\lambda, \\ ((u - G_\lambda)/(m_0 s_0(\lambda)), 1), & u \geq G_\lambda. \end{cases}$$

This rule delivers promises below G_λ by rationing favorable recommendations and promises above G_λ by adding some funding after unfavorable recommendations. Let $d(u) = \rho_1(u) - \rho_0(u)$. Continuation promises can then be chosen so that low-signal exaggeration binds, while the high-signal truth-telling gain remains strictly positive and equals

$$(1-\delta) d(u) \frac{\alpha_1 - \alpha_0}{1 - \alpha_0} > 0.$$

At the same time, the truthful mean continuation equals the current promise. The authority account is a martingale until it leaves I, with increments bounded by $M(1-\delta)$ for a constant M independent of δ. If τ denotes the first exit time and $X_t = U_{t \wedge \tau}$, then

$$\mathbb{E}[(X_t - G_\lambda)^2] \leq M^2 (1-\delta)^2 t, \qquad \mathbb{E}|X_t - G_\lambda| \leq M(1-\delta)\sqrt{t}.$$

These martingale bounds control both the local distortion and the probability of an early exit. Inside I, the principal's flow loss relative to H is at most linear in $|U_t - G_\lambda|$. If the state exits, a public lottery between full funding forever and silence delivers the realized promise exactly. Let V_δ^I denote the principal's payoff from this stopped construction. Summing the local losses and applying Doob's inequality to the exit event gives

$$0 \leq H - V_\delta^I \leq C\sqrt{1-\delta} + 2e^{-(1-\delta)^{-1/2}},$$

for a constant C independent of δ and all sufficiently patient principals. This proves that unrestricted value converges to H. The Appendix verifies the construction on the full formal history tree, supplies the exact continuation formulas, and establishes the bound uniformly.

Once the advice-obedient ceiling itself is attainable, Proposition 5 turns the strict comparison into an upper-tail result. Above the threshold, no optimum can remain advice-obedient; some history reached under truthful reporting must fund after an unfavorable signal. Below δ_λ^{AO}, other

reward mechanisms may govern the low-patience problem. The calibration in Section 6 shows, however, that override is already necessary at $\delta = 9/10$ for the benchmark parameters.

5.1 Overriding truthful bad news

Corollary 1 gives the theorem a direct institutional implication: once the relationship is sufficiently patient, a negative report cannot serve as a veto in any optimal review system. To see the underlying configuration, consider a reviewer who broadly favors experimentation but receives weak evidence about a particular proposal. She reports that weakness honestly. If the agency nevertheless funds, it is neither following her report nor pretending that the evidence was favorable. Rather, it is acting against the report's informational implication but in the direction she privately prefers. In the mechanism, that action delivers an outstanding promise of utility.

This configuration distinguishes the result from models of formal and informal authority and relational delegation, including Baker, Gibbons, and Murphy (1999), Alonso and Matouschek (2007), and Li, Matouschek, and Powell (2017). In the last of these, power is exercised when the principal rubberstamps the agent's preferred recommendation. Here the informative report itself points away from the action that rewards the expert. The organization must sometimes choose the action she privately likes precisely after she has supplied information against it. What looks ex post like disregard of advice can be the intertemporal price of obtaining candid advice earlier in the relationship.

Scientific review provides the leading application. The expert can represent a standing study section or another persistent review body whose membership overlaps and whose disciplinary mandate survives individual turnover. Members of such bodies can combine valuable expertise with preferences for particular fields, methods, or applicants (Li, 2017; Ham et al., 2021). A rule under which an unfavorable assessment mechanically eliminates any chance of funding makes candor especially costly when the review body remains broadly pro-funding. Retaining some capacity to approve despite a negative assessment gives the organization an additional way to compensate honest reporting. This rationale does not require administrators to possess superior project information: peer-review recommendations may predict later performance better than discretionary selections even when administrators sometimes depart from the ranking (Ginther and Heggeness, 2020).

The same logic applies to internal capital allocation. A division manager or technical lead may truthfully report weak project prospects while still preferring continuation because of private benefits from budget, staffing, or experimentation. Headquarters may continue the project because future allocation decisions are the currency with which candor is rewarded when transfers are limited. Dynamic capital-allocation models already emphasize future resources as nonmonetary incentives (Malenko, 2019; Gupta et al., 2024); the distinctive implication here is continuation after a truthful negative assessment. A blanket ban on action against unfavorable advice can improve ex post discipline while weakening the ex ante credibility of bad news.

The mechanism also has an observable signature. Superior administrative information predicts override when administrators possess favorable evidence of their own; portfolio objectives predict

it when the proposal advances program-level priorities; favoritism predicts benefits for connected applicants or reviewers. Here, by contrast, the distinctive force is relationship history. Conditional on the current proposal and report, approval after negative advice can vary with the expert's accumulated claim on future influence, and restricting override capacity should change reporting incentives as well as final decisions. Related models of informal authority and relational delegation also make history matter, but through earned deference to the agent's recommendation (Baker, Gibbons, and Murphy, 1999; Alonso and Matouschek, 2007; Li, Matouschek, and Powell, 2017); the present model instead predicts action against the current report's informational implication in the direction favored by the expert.

6 An ordered benchmark

The benchmark $\lambda = 1/2$ makes the delivery logic fully visible. Economically, a funded bad project then gives the expert zero payoff, so the case lies exactly between an ex post taste for funding bad projects and an ex post aversion to them. Algebraically, after suppressing the type subscript, $h = 0$, $\bar{u} = q$, and $P(u) = 2u - C(u)$. More importantly, holding a branch's total score fixed, additional current funding reduces the good-outcome continuation while leaving the bad-outcome continuation unchanged. That one-sided movement removes a pooling obstruction and allows the cost-minimizing selection to be ordered from rationing, to review, and then to override. For general λ, by contrast, both continuations move and pooling can bind.

Current funding. Because current funding no longer moves the bad-outcome continuation, an optimum can be selected with

$$\rho_1 = \min\left\{1, \frac{y_1}{a}\right\}, \qquad \rho_0 = \max\left\{0, \frac{y_0 - \delta q}{a}\right\}. \tag{5}$$

Only three current allocation forms remain. Rationing has $0 \leq \rho_1 < 1$ and $\rho_0 = 0$; review has $\rho_1 = 1$ and $\rho_0 = 0$; and override has $\rho_1 = 1$ and $\rho_0 > 0$. Under (A3), Proposition 4 specializes to $C(u) = u/\alpha_1$ on the low-promise interval $[0, aq]$.

The proof then coordinates three operations that cannot be performed independently: it binds the low-signal exaggeration constraint, orders the favorable-report continuations, and minimizes the unfavorable-report branch. Once those operations are made compatible, only the payoff assigned after an unfavorable signal remains to be chosen. Its least minimizer is monotone in promised authority, which produces a single review-to-override cutoff. The Appendix gives the details.

6.1 Ordered authority regions

Because the reduced low-signal payoff is monotone in the promise, the selected optimum can cross from review to override at most once.

Theorem 4 (Ordered authority regions)**.** *Under (A1)–(A3), there is a cutoff*

$$u_H \in \left[q - a\alpha_0,\ q - \frac{a\alpha_0(1-q)}{1-\alpha_0}\right]$$

such that, apart from possible boundary ties, a cost-minimizing decomposition can be selected at every promise according to

$$0 < u < aq \Rightarrow \textit{rationing}, \qquad aq < u < u_H \Rightarrow \textit{review}, \qquad u_H < u \leq q \Rightarrow \textit{override}.$$

For the least-minimizer selection constructed in the proof, the upper regime never switches back to review as promised utility rises. At $u = q$, full funding and continuation promise q after every public outcome are necessary.

Below the cutoff, continuation utility alone can deliver the low-signal payoff, so current advice is followed. Above the cutoff, by contrast, the cost-minimizing report-0 branch requires current funding. Affine portions of the frontier may create ties; for that reason, the theorem selects an ordered optimum rather than claiming that every optimum has the same current action.

To obtain this ordering without narrowing the mechanism class, the proof starts from a global minimizer of the frontier program. At $\lambda = 1/2$, the bad-outcome score simplifies to $x_r = \delta w_{r0}$. Holding (x_r, y_r) fixed while changing ρ_r leaves the bad-outcome continuation unchanged and moves only the good-outcome continuation. Convexity then pushes favorable-report funding upward and unfavorable-report funding downward until one of the bounds in (5) is reached. Because these transformations preserve promise keeping and both reporting constraints while weakly reducing cost, they carry a global minimizer into another global minimizer.

The proof next binds the low-signal incentive constraint. If that constraint is slack, the two bad-outcome scores can be moved toward one another without changing their contribution to promise keeping; doing so relaxes the high-signal constraint and, by convexity, weakly lowers cost. Once the low-signal constraint binds, the report-0 branch can be replaced by the cheapest branch delivering the same payoff to posterior α_0. The report-1 branch has enough score spread to preserve high-signal truth telling. Finally, the reduced cost has decreasing differences and its feasible interval shifts upward with promised authority, so the least minimizer is monotone.[6] That last step rules out a return from override to review and yields the single cutoff u_H.

A finite-patience certificate. Although Theorem 3 is asymptotic, the same force is already active at a transparent finite discount factor. At $p = 3/4$, $q = 2/5$, $\delta = 9/10$, and $\lambda = 1/2$, every globally optimal admissible outcome overrides unfavorable advice with positive probability under truthful reporting. The comparison is exact: every advice-obedient outcome gives the principal at

[6] This is the standard monotone-comparative-statics logic for ordered solution selections; see Milgrom and Shannon (1994) and Topkis (1998).

most $2/15$, whereas a seven-state mechanism gives

$$\frac{162749}{1211600} = \frac{2}{15} + \frac{3607}{3634800}.$$

Because the known-type optimum is attained, this strict comparison applies to every optimum. Proposition 5 then propagates it to every higher discount factor. The Appendix gives the complete mechanism and exhibits a positive-probability route to override.

7 Private conflict: screening over delivery policies

When conflict is privately known, the delivery problem becomes an assignment problem. Every complete dynamic policy generates two payoff-relevant numbers (total funding A and bad-project exposure B), and type λ values that pair as $A - 2\lambda B$. A menu can be studied through the geometry of the exposure pairs its branches generate. Types evaluate those pairs affinely, so the menu value is their upper envelope; wherever that envelope is differentiable, its slope identifies bad-project exposure.[7] Geometry alone is not enough, however, because implementation also requires an actual branch and a complete sequentially optimal strategy. This section separates geometric value from sequential implementation. It characterizes the exposure envelope, explains why the complete-information low frontier cannot screen types, and proves finite-menu value density together with comparative results for diagnostics. A full characterization of the globally optimal private-type menu remains open.

Formally, let the type support be a compact interval $\Lambda = [\lambda_L, \lambda_H] \subseteq [0,1]$. After learning her type, the expert publicly chooses one complete dynamic policy from a committed menu $\mathcal{M} = (J, (M_j)_{j\in J})$; the selected branch then runs. To make this choice part of one well-defined extensive form, the common index and history spaces satisfy the same measurability convention as the unrestricted benchmark. The Appendix states the primitive conditions. For a branch M_j and Borel behavioral strategy σ, write

$$A_{M_j}(\sigma) = a\mathbb{E}_\sigma \sum_{t\geq 0} \delta^t d_t, \qquad B_{M_j}(\sigma) = a\mathbb{E}_\sigma \sum_{t\geq 0} \delta^t d_t(1-\theta_t).$$

A type λ evaluates the resulting pair as $A - 2\lambda B$, whereas the principal evaluates it as $A - 2B$. Let

$$Z(M_j) = \{(A_{M_j}(\sigma), B_{M_j}(\sigma)) : \sigma \text{ Borel behavioral}\}, \qquad D(\mathcal{M}) = \operatorname{cl}\operatorname{co}\left(\bigcup_{j\in J} Z(M_j)\right),$$

where $\operatorname{cl}\operatorname{co}$ denotes the closed convex hull, and define the expert's indirect-utility envelope by

$$U(\lambda) = \sup_{j\in J,\,\sigma} \{A_{M_j}(\sigma) - 2\lambda B_{M_j}(\sigma)\}.$$

[7] The use of convex potentials, support functions, and subgradient geometry in mechanism design follows a well-established line including Rochet (1987) and Krishna and Maenner (2001).

To interpret the envelope as equilibrium value, the maximizing objects must actually be selected. A menu is screening-admissible when a Borel branch choice and complete sequentially optimal strategies attain the envelope almost everywhere. It is *branchwise sequentially regular* when, for every branch, such a complete strategy can be selected measurably as a function of the expert's type and private history. For a type distribution F, let $\Pi_F(\mathcal{M})$ denote the principal's resulting expected payoff. The Appendix states the timing and measurability conditions. Notice that finiteness refers only to the number of menu branches: each branch may still have infinite histories and unrestricted within-branch communication.

7.1 Exposure and indirect utility

Because types differ only in how they value bad funding, private information is screened through bad-project exposure rather than through total funding itself.

Theorem 5 (Exposure envelope under private bias)**.** *For every nonempty indexed menu $\mathcal{M}$, $D(\mathcal{M})$ is nonempty, compact, and convex, and*

$$U(\lambda) = \max_{(A,B)\in D(\mathcal{M})} \{A - 2\lambda B\}.$$

The function U is convex, weakly decreasing, absolutely continuous, and $2(1-q)$-Lipschitz. For a screening-admissible menu, U is equilibrium indirect utility, and at every differentiability point every attained equilibrium outcome gives

$$B(\lambda) = -\frac{1}{2}U'(\lambda), \qquad A(\lambda) = U(\lambda) - \lambda U'(\lambda).$$

For such a menu, every equilibrium exposure selection is weakly decreasing in λ. If, in addition, F is atomless, then

$$\Pi_F(\mathcal{M}) = \int_\Lambda \left[U(\lambda) + (1-\lambda)U'(\lambda)\right] dF(\lambda),$$

independently of equilibrium tie breaking.

Bad-project exposure is the allocation statistic conjugate to private alignment: more aligned types select weakly lower exposure. The Appendix develops the corresponding geometry at kinks and along maximizing faces. When the type distribution has an atom, however, the principal's value can depend on which attained branch–strategy outcome is selected at that kink.

Separation and finite menus. The exposure envelope also clarifies why the complete-information low frontier cannot by itself screen conflict. If every assigned truthful outcome lies on the explicit low-frontier ray, then

$$B(\lambda) = (1-\alpha_1)A(\lambda), \qquad U(\lambda) = A(\lambda)s_1(\lambda).$$

Because all types then rank outcomes through the same funding index, incentive compatibility forces $A(\lambda)$ to be constant across the assigned types. Separation requires outcomes outside that ray.

Although the benchmark menu may contain uncountably many branches, only finitely many are needed to approximate its value. Write $U_{\mathcal{M}}$ for the indirect-utility envelope generated by menu $\mathcal{M}$. If F is atomless and $D_\Lambda = \lambda_H - \lambda_L$, then for every $K \geq 1$ and $\varepsilon > 0$ one can retain at most $K+1$ branches and obtain the following uniform bound. Denote such a submenu by $\mathcal{M}_{K,\varepsilon}$:

$$0 \leq U_{\mathcal{M}}(\lambda) - U_{\mathcal{M}_{K,\varepsilon}}(\lambda) \leq \frac{2(1-q)D_\Lambda}{K} + \varepsilon.$$

Under screening admissibility and branchwise sequential regularity, the retained submenus can also be chosen so that expected principal value converges. Within that class, the optimal value over indexed menus equals the supremum over finite menus.[8]

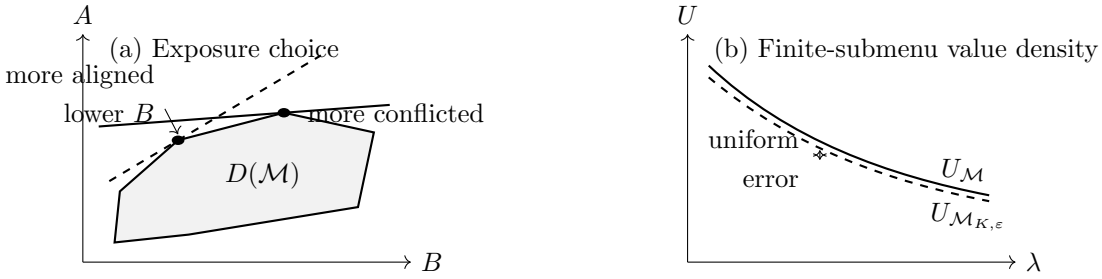

Figure 2: Private conflict is screened through bad-project exposure. Panel (a) shows the geometric hull and its maximizing faces; equilibrium tie breaking uses the outcomes actually attained by the menu. Panel (b) shows support-function value density. One dynamic branch can generate many exposure outcomes, so a finite-submenu envelope need not be polygonal.

Diagnostics. A public conflict diagnostic refines this assignment problem by changing the posterior over types before the principal chooses a category-specific menu. Because the diagnostic is realized before branch choice, it changes the menu offered but does not create a new state within a branch. More informative finite diagnostics weakly raise value in the Blackwell order, since the principal can always garble them (Blackwell, 1953). Moreover, if every positive-probability posterior is atomless, categorywise finite-menu values converge to the unrestricted diagnostic-contingent value. Under stronger regularity, the Appendix also gives an inverse-square-root approximation rate.[9] Better diagnostics raise value, although they may move the selected policy in either direction.

In scientific funding, observable reviewer–proposal proximity, disciplinary fit, conflict disclosures, and panel composition can serve as diagnostics of the expert's likely conflict. Their value need not come from improving the agency's own assessment of project quality. They allow the agency to tailor the menu of future influence to a more informative posterior over reviewer bias.

[8]This is value density, not density of allocations or mechanisms. The Appendix states the common-index measurability convention, selector and nonattainment qualifications, and the role of atomlessness; with atoms, equilibrium tie breaking can remain payoff relevant.

[9]The quantitative rate uses bounded posterior densities and Lipschitz indirect-utility envelopes.

8 Discussion and conclusion

8.1 Scientific review and retained authority

Scientific funding illustrates the paper's institutional margin. As Li (2017) shows, NIH reviewers who are intellectually closer to an application are both more favorable toward it and better able to distinguish high- from low-quality projects. In her setting, the informational benefit of expertise weakly dominates the cost of bias. Blanket recusal can discard valuable information together with the conflict, leaving the organization to decide how much control it should retain after consulting the expert. In this application, the repeated expert is naturally interpreted as a standing study section or other persistent review unit: individual members may rotate while the unit's mandate, procedures, and relationship with the funding agency endure.

A historical NIH–NSF comparison makes that choice concrete. During the period studied by Li (2017), NIH institutes lined up applications by study-section score and largely funded them in that order; fewer than four percent were funded out of order. The NSF process described in its 2016 proposal guide, by contrast, assigned program officers a separate recommendation role after external review, with senior staff reviewing award recommendations (National Science Foundation, 2016). Both systems relied on specialists, but expert evaluation was much closer to decisive in the NIH setting documented by Li, whereas the NSF architecture visibly retained an administrative decision between review and award. This is precisely the distinction between giving negative advice effective veto power and preserving room to approve despite it.

That retained layer can also use information about the review process itself. Program officers may observe disciplinary fit, panel composition, disclosed conflicts, and other features informative about how strongly a recommendation reflects expertise or a pro-funding preference. In the language of Section 7, these observables are conflict diagnostics: they permit the agency to condition the incentive regime on a more informative posterior over reviewer bias without requiring administrators to possess a better signal of scientific quality.

The agencies' different scopes offer a plausible reason why the value of this retained layer may vary. NIH applications in Li's setting were routed to standing study sections organized around specialized biomedical themes. NSF spans a much wider range of science and engineering and often assesses proposals that cross disciplinary boundaries. Consistent with the difficulty of evaluation across cognitive distance, Bromham, Dinnage, and Hua (2016) find in a comprehensive study of Australian grant applications that greater interdisciplinary distance is associated with lower funding success. Broad portfolios may place greater value on a decision layer that is not mechanically tied to any one panel's ranking.

Administrative discretion need not rest on administrators knowing more than specialist reviewers. Ginther and Heggeness (2020) study a two-stage NIH fellowship process in which administrative choices depart from peer-review rankings and find that peer-review recommendations better predict later scientific performance than discretionary selections. That finding sharpens the mechanism here. Retained authority can be valuable even when the current review is the better forecast, because

future decisions are the currency used to reward honest advice. A commitment never to fund after unfavorable advice removes one way of honoring those promises and can thereby reduce the credibility of the advice itself.

This interpretation is distinct from the authority and relational-delegation mechanisms in Baker, Gibbons, and Murphy (1999), Alonso and Matouschek (2007), and Li, Matouschek, and Powell (2017). In Li, Matouschek, and Powell, the principal eventually defers to the agent's preferred recommendation. Here the expert can report unfavorable evidence truthfully and still prefer funding, so the principal rewards her by acting against the current report's informational content. The model speaks to the more puzzling pattern in which an organization approves after candid negative evaluation. Its distinctive empirical implication is history dependence: conditional on the current proposal and report, departures from negative advice should depend on the expert's accumulated influence, and institutional limits on override should affect what experts are willing to report.

8.2 Scope and future directions

The exact delivery characterization is strongest for conflicted types with $\lambda \geq 1/2$, and the complete phase ordering is proved at the benchmark $\lambda = 1/2$. By contrast, the upper-tail patience theorem applies to every conflicted known type and uses neither (A2) nor (A3).[10]

Patience monotonicity concerns inclusion of attainable payoff vectors, and the unique override boundary is an upper-tail statement beginning once the advice-obedient ceiling is attainable. With private types, the exposure geometry is unconditional; implementation and expected-value convergence use the admissibility and regularity conditions stated in Section 7 and the Appendix.[11]

Binary quality and signals, public ex post quality, and i.i.d. arrivals keep the state one-dimensional. Richer or persistent information would generally require additional beliefs or continuation statistics. Limited commitment would make the delivery of promised authority itself an enforcement problem, while multiple experts would introduce aggregation and cross-report incentives. These extensions may alter the form of the optimal mechanism and provide natural settings in which to study whether the logic of history-dependent override survives.

The paper's main results can be read through this institutional lens. Without transfers, the principal rewards honest advice through future decisions, and the minimum-funding frontier measures the cost of doing so. Low authority balances are delivered by rationing favorable advice, whereas balances above the no-override capacity require current funding after unfavorable advice. At the level of the whole relationship, advice-obedient mechanisms face a fixed payoff ceiling, while unrestricted committed mechanisms approach the signal-contingent first best. Yet once patience exceeds the

[10] Payoff equivalence is typewise and uses attained sequential best responses. Within its proof, the reduction from arbitrary current messages to direct signal reports is established for finite current-message alphabets. For $\lambda > 1/2$, the Appendix derives the pointwise obstacle map even though the main text does not impose a universal ordering on its intermediate boundaries.

[11] In particular, indexed menus use common measurable index and history spaces, while finite-menu expected-value convergence invokes screening admissibility, branchwise sequential regularity, and atomlessness. Diagnostics add Blackwell monotonicity and categorywise convergence; the quantitative rate requires the stronger conditions stated above.

threshold, every global optimum must still fund after unfavorable advice at some history reached under truthful reporting. The welfare loss vanishes with patience, even though the action used to honor earlier promises remains on the truthful path.

Private bias adds assignment to this delivery problem. Funding and bad-project exposure become screening allocations over complete dynamic policies; exposure falls with alignment, the low-frontier ray cannot separate types, and finite menus approximate the value of richer indexed mechanisms under the stated regularity conditions. A complementary empirical direction is to test whether override depends on an expert's accumulated authority, conditional on the current proposal and report.

References

Abreu, D., Pearce, D., and Stacchetti, E. (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* 58(5), 1041–1063.

Aghion, P., and Tirole, J. (1997). Formal and real authority in organizations. *Journal of Political Economy* 105(1), 1–29.

Aliprantis, C. D., and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. Springer, Berlin.

Alonso, R., and Matouschek, N. (2007). Relational delegation. *RAND Journal of Economics* 38(4), 1070–1089.

Amador, M., and Bagwell, K. (2020). Money burning in the theory of delegation. *Games and Economic Behavior* 121, 382–412.

Ambrus, A., and Egorov, G. (2017). Delegation and nonmonetary incentives. *Journal of Economic Theory* 171, 101–135.

Baker, G., Gibbons, R., and Murphy, K. J. (1999). Informal authority in organizations. *Journal of Law, Economics, and Organization* 15(1), 56–73.

Bergemann, D., and Välimäki, J. (2019). Dynamic mechanism design: An introduction. *Journal of Economic Literature* 57(2), 235–274.

Bird, D., and Frug, A. (2019). Dynamic non-monetary incentives. *American Economic Journal: Microeconomics* 11(4), 111–150.

Bird, D., and Frug, A. (2025). A theory of front-line management. *Management Science*, Articles in Advance.

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

Bromham, L., Dinnage, R., and Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature* 534, 684–687.

Campbell, A. (2021). Spending political capital. *Economic Journal* 131(640), 3103–3121.

de Clippel, G., Eliaz, K., Fershtman, D., and Rozen, K. (2021). On selecting the right agent. *Theoretical Economics* 16, 381–402.

Deb, R., Pai, M., and Said, M. (2018). Evaluating strategic forecasters. *American Economic Review* 108(10), 3057–3103.

Doval, L., and Skreta, V. (2022). Mechanism design with limited commitment. *Econometrica* 90(4), 1463–1500.

Fehrler, S., and Janas, M. (2020). Delegation to a group. *Management Science* 67(6), 3714–3743.

Frankel, A. (2014). Aligned delegation. *American Economic Review* 104(1), 66–83.

Frankel, A. (2016). Discounted quotas. *Journal of Economic Theory* 166, 396–444.

Ginther, D. K., and Heggeness, M. L. (2020). Administrative discretion in scientific funding: Evidence from a prestigious postdoctoral training program. *Research Policy* 49(4), 103953.

Guo, Y., and Hörner, J. (2020). Dynamic allocation without money. Toulouse School of Economics Working Paper 20-1133.

Gupta, S., Bansal, S., Dawande, M., and Janakiraman, G. (2024). Trust-and-evaluate: A dynamic nonmonetary mechanism for internal capital allocation. *Management Science* 70(11), 7811–7828.

Ham, S. H., Koch, I., Lim, N., and Wu, J. (2021). Conflict of interest in third-party reviews: An experimental study. *Management Science* 67(12), 7535–7559.

Kallenberg, O. (2021). *Foundations of Modern Probability*, 3rd ed. Springer, Cham.

Kechris, A. S. (1995). *Classical Descriptive Set Theory*. Springer, New York.

Kivinen, S., and Kuzmics, C. (2025). Renegotiation-proof cheap talk. Graz Economics Papers 2025-11, University of Graz.

Krishna, V., and Maenner, E. (2001). Convex potentials with an application to mechanism design. *Econometrica* 69(4), 1113–1119.

Li, D. (2017). Expertise versus bias in evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics* 9(2), 60–92.

Li, J., Matouschek, N., and Powell, M. (2017). Power dynamics in organizations. *American Economic Journal: Microeconomics* 9(1), 217–241.

Lipnowski, E., and Ramos, J. (2020). Repeated delegation. *Journal of Economic Theory* 188, 105040.

Malenko, A. (2019). Optimal dynamic capital budgeting. *Review of Economic Studies* 86(4), 1747–1778.

Milgrom, P., and Shannon, C. (1994). Monotone comparative statics. *Econometrica* 62(1), 157–180.

National Science Foundation. (2016). *Proposal and Award Policies and Procedures Guide* (NSF 16-1). National Science Foundation, Arlington, VA.

Özer, Ö., Subramanian, U., and Wang, Y. (2017). Information sharing, advice provision, or delegation: What leads to higher trust and trustworthiness? *Management Science* 64(1), 474–493.

Pavan, A., Segal, I., and Toikka, J. (2014). Dynamic mechanism design: A Myersonian approach. *Econometrica* 82(2), 601–653.

Rantakari, H. (2021). Relational influence. Working paper, University of Rochester, November 12.

Rantakari, H. (2023). How to reward honesty? *Journal of Economic Behavior & Organization* 207, 129–145.

Rochet, J.-C. (1987). A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics* 16(2), 191–200.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton.

Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton University Press, Princeton.

Appendix to "Keeping Experts Honest: The Necessity of Override"

Proofs and supplementary results

This appendix collects proofs and supplementary results for the main paper. It uses the primitives, assumptions, and notation defined there, and its numbering is independent of the main text.

Proof guide and conventions

Section 1 proves payoff equivalence between unrestricted and recursive mechanisms and establishes current-signal revelation. Section 2 develops the recursive frontier and exact recursion. Sections 3–5 prove the known-type results, while Sections 6 and 7 treat indexed private-type menus and conflict diagnostics. *Numbering crosswalk.* Appendix results use section-prefixed numbering, independently of the main text. The main paper's central formal statements are proved here as follows:

Main paper	Appendix proof result
Theorem 1	Theorem 1.1
Propositions 1–2	Propositions 2.1–2.2
Propositions 3–4 and Theorem 2	Propositions 3.1–3.2 and Theorem 3.1
Proposition 5, Theorem 3, and Corollary 1	Proposition 5.1, Theorem 5.1, and Corollaries 5.1–5.2
Theorems 4–5	Theorems 4.1 and 6.1

Throughout, date-by-date spaces are standard Borel, kernels are defined on the full formal history tree, and selected strategies are complete and sequentially optimal after every finite private history, including null histories. Finite ties are resolved by the least indexed maximizer. When a proof averages continuation payoff pairs, the average is implemented by a newly selected complete continuation mechanism. We write (u, A) for expert utility and discounted funding, B for bad-project exposure, C_λ for the minimum-funding frontier, and (x_r, y_r) for current-plus-continuation branch scores. Quality and signals are binary, ex post quality is public, and arrivals are i.i.d. The additional assumptions used by particular results are stated with those results.

1 Payoff equivalence of unrestricted and recursive mechanisms

1.1 The unrestricted outcome class

This section answers one question: does solving the finite promised-utility recursion discard payoffs available under richer history-dependent mechanisms? The answer is no under the maintained attained-best-response benchmark. The proof preserves expert utility and funding, not the original public process. Its two substantive steps are local. First, after a public history, inherited private signals do not change maximal continuation utility, though they may change which optimal strategy is selected and therefore the funding path. Second, continuation payoff pairs can be averaged and

then implemented anew by the finite recursive mechanism. Lemma 1 establishes the first point, and Lemma 2 the second.

Fix a known type λ. At each date the public history records past public messages, funding decisions, qualities, and public randomization; private history additionally records the expert's signals. The date-by-date spaces are fixed and standard Borel, and mechanism and strategy kernels are specified at every finite history. This is where the external measure theory enters: disintegration supplies the conditional kernels used in the compression, and Ionescu–Tulcea supplies the induced continuation laws; see, for example, Kallenberg (2021). No equivalence claim is made for an arbitrary collection of unrelated history-indexed measurable spaces.

Definition 1 (Admissible standard-Borel outcome)**.** An admissible outcome is a standard-Borel public mechanism together with a Borel behavioral strategy that is sequentially optimal after every finite private history. Sequential optimality is pointwise, including at histories having zero probability under the initial path law. The requirement that the selected best response be attained is part of the class.

Let $\mathcal{W}_\lambda^{SB}$ be the set of normalized expert-utility/funding pairs (u, A) induced by admissible outcomes. Let $\mathcal{W}_\lambda^{F}$ be the corresponding set generated by public direct mechanisms with binary current-signal reports and one fixed ternary public label drawn before the signal at every date. The label selects among at most three current decompositions; all label, report, funding, and quality branches have complete continuation rules. Lemma 3 shows that this fixed alphabet already implements every public mixture needed below.

The exact equivalence proof uses two facts that must be kept separate. Continuation *utility* after a public history is independent of inherited private signals. Continuation *funding* may depend on inherited private signals because different optimal continuation strategies can induce different funding paths.

Lemma 1 (Inherited-private-history lemma)**.** *Fix an admissible standard-Borel outcome and a date-t continuation public history h. Let $k \in \{0,1\}^t$ be any formal inherited signal history. The mechanism and selected strategy are defined at (h, k) whether or not that private history has positive probability under the initial path law.*

(i) The expert's maximal continuation utility from h is a number $V(h)$ independent of k.

(ii) Hard-code k into the original continuation strategy: on a fresh future private signal history s, use the action prescribed originally after the concatenated private history $(h; k, s)$. The resulting shifted strategy is Borel and sequentially optimal after every fresh future private history, including fresh null histories.

(iii) If $A(h, k)$ is the normalized continuation funding induced by that shifted strategy, then $(V(h), A(h, k)) \in \mathcal{W}_\lambda^{SB}$.

(iv) The maps $V(h)$ and $A(h, k)$ are Borel on the corresponding finite-history spaces.

Proof. Conditional on a public history h, future primitives and public rules are independent of inherited signals, and the feasible future strategy sets are canonically identical across formal strings k. This proves the common value $V(h)$. Prefixing (h,k) to the original public and private kernels gives Borel shifted kernels; every fresh history maps to a formal original history where sequential optimality holds, so the shifted strategy is optimal everywhere and generates $(V(h), A(h,k))$.

Finite-horizon evaluations of the shifted kernels are Borel and converge uniformly because discounted tails are bounded by a constant times δ^T. Hence $A(h,k)$ is Borel. The utility limit is constant in k; evaluating it at the canonical string $(0,\dots,0)$ gives a Borel version of $V(h)$ without selecting a compatible private history. □

Lemma 2 (Barycenters of continuation outcomes)**.** *Let Y be a standard-Borel random element and let $Z(Y) = (V(Y), A(Y))$ be a bounded Borel random vector such that $Z(Y) \in \mathcal{W}_\lambda^{SB}$ almost surely. Then*

$$\mathbb{E}Z(Y) \in \operatorname{cl}\operatorname{co}\mathcal{W}_\lambda^{SB}.$$

No measurable selection of continuation mechanisms or strategies is required.

Proof. If $b = \mathbb{E}Z(Y)$ lay outside the closed convex set $\operatorname{cl}\operatorname{co}\mathcal{W}_\lambda^{SB}$, a separating hyperplane c would satisfy $c \cdot b > \sup_{w \in \mathcal{W}_\lambda^{SB}} c \cdot w$. But $Z(Y) \in \mathcal{W}_\lambda^{SB}$ almost surely, so $c \cdot b = \mathbb{E}[c \cdot Z(Y)]$ cannot exceed that supremum. □

1.2 Finite recursive maximality

Fix once and for all the alphabets

$$L = \{1,2,3\}, \qquad R = D = \Theta = \{0,1\}.$$

At the beginning of each date a public label $\ell \in L$ is drawn, then the expert observes her signal and reports $r \in R$, funding $d \in D$ is realized, and quality $\theta \in \Theta$ is observed. Thus the date-t public histories form the finite set $(L \times R \times D \times \Theta)^t$, and their union over dates is countable. A mechanism specifies a label distribution at every pre-label public history and a funding probability after every public history, current label, and report. Truthful reporting is the designated strategy. Rules are specified on the entire formal history tree, whether or not a label, report, or funding realization has positive probability.

For compact $W \subset \mathbb{R}^2$, an *elementary decomposition* consists of one current binary-report rule, continuation pairs in W after every (r,d,θ), exact promise keeping for its induced pair, and both current-report inequalities. Let $E(W)$ be the compact set of payoff pairs generated by elementary decompositions, and define

$$\mathcal{D}(W) := \operatorname{co} E(W).$$

Because the ambient payoff space is $\mathbb{R}^2$, every point of $\mathcal{D}(W)$ is a mixture of at most three elementary decompositions. The public label $\ell \in L$ implements that mixture before the signal is observed.

Conditional on each realized label, the selected elementary decomposition is itself truthful. If W is compact and convex, then $\mathcal{D}(W)$ is compact and convex.

Lemma 3 (Finite recursive maximality)**.** *The set $\mathcal{W}_\lambda^F$ is nonempty, compact, and convex and satisfies*

$$\mathcal{W}_\lambda^F = \mathcal{D}(\mathcal{W}_\lambda^F).$$

If W is any compact convex set with $W \subseteq \mathcal{D}(W)$, then $W \subseteq \mathcal{W}_\lambda^F$. The implementation is exact and supplies pointwise rules after every finite public and private history.

Proof. With the fixed alphabets, the mechanism space is a countable product of finite-dimensional simplexes and intervals. Discounted payoff and one-period-deviation maps are uniform limits of finite-horizon continuous maps; truthful mechanisms form a closed set, so their payoff image is compact. Carathéodory's theorem in $\mathbb{R}^2$ implements every convex combination with the same ternary root label, proving convexity.

Decomposing the first date gives $\mathcal{W}_\lambda^F \subseteq \mathcal{D}(\mathcal{W}_\lambda^F)$. Conversely, a point of $\mathcal{D}(\mathcal{W}_\lambda^F)$ is a mixture of at most three elementary decompositions. The root label selects one, and each formal child receives a complete truthful continuation attaining its assigned pair. Once continuation payoffs are identified, the elementary report inequalities and the pointwise one-period-deviation principle give sequential truthfulness, proving the reverse inclusion.

For maximality, fix $w_\varnothing \in W$ and recursively choose, at every node of the full countable tree $\bigcup_t (L \times R \times D \times \Theta)^t$, a Carathéodory decomposition of the assigned pair in W. Countable choice suffices. Writing $w_h = (u_h, A_h)$ and g_h for current normalized flows, promise keeping implies

$$w_h = g_h + \delta \mathbb{E}_h w_{H_1}.$$

After T iterations the residual is $\delta^T \mathbb{E}_h w_{H_T}$, which vanishes uniformly because W is bounded. Thus the constructed mechanism delivers every assigned pair exactly from every formal node. The local report constraints therefore use actual continuation utilities, and the one-period-deviation principle gives sequential truthfulness at reached and null histories. Hence $W \subseteq \mathcal{W}_\lambda^F$. □

Theorem 1 (Payoff equivalence of unrestricted and recursive mechanisms)**.** *For every fixed known type λ,*

$$\mathcal{W}_\lambda^{SB} = \mathcal{W}_\lambda^F.$$

The equality preserves the expert-utility/funding pair and hence the principal payoff, but need not preserve the original message process.

Proof. Let $K = \operatorname{cl}\operatorname{co} \mathcal{W}_\lambda^{SB}$. It is compact and convex. We show $K \subseteq \mathcal{D}(K)$. Take $(u, A) \in \mathcal{W}_\lambda^{SB}$. At the initial history let μ_i be the selected distribution of the original current message after signal i. Combine all current public randomization before funding into a record c with kernel $\Gamma(dc, d \mid m)$,

and let $H(m, c, d, \theta)$ be the resulting continuation public history. Put

$$q_{rd} = \int \mu_r(dm)\Gamma(\mathsf{C}, d \mid m).$$

For $q_{rd} > 0$, define the continuation-history law

$$\nu_{rd\theta}(B) = \frac{1}{q_{rd}} \int \mu_r(dm) \int \Gamma(dc, d \mid m)\mathbf{1}\{H(m, c, d, \theta) \in B\}. \qquad \text{(SB-branch)}$$

This is a normalized restriction of primitive kernels, not a conditional-law choice on a null event. Reporting r under true signal i reaches (d, θ) with probability $q_{rd}\beta_i(\theta)$. Full support of β_i makes null branches common to all signals choosing report r.

For $q_{rd} > 0$, average $(V(h), A(h, (r)))$ under $\nu_{rd\theta}$. Lemmas 1–2 place the resulting pair in K; when $q_{rd} = 0$, assign any pair in K. The truthful branches preserve (u, A). If true signal i reports r, the binary device reproduces the feasible original deviation that draws $m \sim \mu_r$ and then uses a continuation best response for the actual inherited history. Its continuation utility is $V(h)$, so original sequential optimality gives both binary report inequalities. Hence $\mathcal{W}^{SB}_\lambda \subseteq \mathcal{D}(K)$.

Compact convexity of $\mathcal{D}(K)$ implies $K \subseteq \mathcal{D}(K)$, and Lemma 3 gives $K \subseteq \mathcal{W}^F_\lambda$. The reverse inclusion is immediate. The constructed finite mechanism implements the averaged pairs anew on its full formal tree; it does not reproduce the original process. Attainment of the selected pointwise best response is therefore part of the theorem's admissible-outcome class. □

1.3 Finite current-signal revelation

Lemma 4 (Current-signal dynamic revelation)**.** *Fix a public mechanism whose current-message alphabet is finite at every date, together with a sequentially optimal, possibly mixed behavioral reporting strategy that may depend on the entire private history. There exists a public direct mechanism in which the expert reports only the current signal and truthful reporting is sequentially optimal after every finite private history. Under truthful play, the direct mechanism reproduces the finite-dimensional distributions of the original public messages, public randomization outcomes, decisions, qualities, and hence discounted payoffs. The conclusion is truthful-path distributional equivalence; it does not identify the original private strategy or reproduce off-path behavior after deviations.*

Proof. At date t, let h be the original public history and let the finite message set be $\mathcal{M}_t$. By Lemma 1, maximal continuation utility $V_t(h)$ is independent of inherited signals and is Borel. Let $Q_{t,i}(h, m)$ be the payoff to true signal i from current message m followed by continuation value V_{t+1}; these functions are Borel. The least element $b_{t,i}(h)$ of the finite argmax is therefore Borel.

Choose a Borel regular conditional law $\widehat{\mu}^i_{t,h}$ of the original message given (h, i). It is specified only almost surely. The set

$$G_{t,i} = \{h : \widehat{\mu}^i_{t,h}(\arg\max_m Q_{t,i}(h, m)) = 1\}$$

is Borel and has full truthful-path probability. Define $\mu^i_{t,h} = \widehat{\mu}^i_{t,h}$ on $G_{t,i}$ and $\delta_{b_{t,i}(h)}$ otherwise. This is a Borel kernel supported on current best replies at every formal history.

After direct report r, publicly draw a simulated original message from $\mu^r_{t,h}$, record it in the simulated original public history, and apply the original current kernels. Future rules condition on the entire simulated history. The original value function is a pointwise fixed point of the direct truthful-evaluation contraction, hence is the actual direct continuation value everywhere. Truthful report i mixes only maximizers of $Q_{t,i}$; any other report gives another mixture of the same values and cannot improve. The one-period-deviation principle yields sequential truthfulness at every private history. Since $\mu^i_{t,h}$ equals the original predictive message law almost surely on truthful histories, induction reproduces all finite-dimensional truthful public distributions. Retaining past simulated messages preserves their serial correlation. The construction does not reproduce the original private strategy or off-path process. □

2 Finite recursive frontier foundations

Write $a = 1 - \delta$, $h = 2\lambda - 1$, and

$$s_i(\lambda) = 1 - 2\lambda(1 - \alpha_i).$$

Throughout this section the type is conflicted, so $s_0(\lambda) > 0$. Then both signal-contingent funding payoffs are positive and maximal expert utility is

$$\bar{u}_\lambda = m_0 s_0(\lambda) + m_1 s_1(\lambda) = 1 - 2\lambda(1 - q),$$

achieved by funding every project. The restriction to conflicted types is essential: outside this region unconditional funding need not maximize expert utility.

Proposition 1 (Frontier attainment and recursion)**.** *For a conflicted known type, the utility projection of $\mathcal{W}^F_\lambda = \mathcal{W}^{SB}_\lambda$ is $[0, \bar{u}_\lambda]$. The minimum-funding frontier C_λ is attained, continuous, and convex. Every initial promise has an optimal truthful policy on a countable finite-alphabet formal history tree with complete rules on zero-probability branches.*

Proof. Silence and unconditional funding attain $(0, 0)$ and $(\bar{u}_\lambda, 1)$, and public mixing fills the utility interval. Since $s_0, s_1 > 0$, no history can generate current expected expert payoff above $\bar{u}_\lambda$; iteration gives the global bound. Compactness, convexity, and continuity follow from Lemma 3.

At a frontier pair, replace each continuation by a frontier mechanism with the same utility and select a minimizing one-period decomposition. Recursively apply this selection at every node of the full countable formal tree. The bounded residual argument in Lemma 3 identifies promised and actual continuation pairs exactly, and pointwise one-period deviations give sequential truthfulness. The frontier graph need not be convex; the construction follows selected frontier decompositions rather than applying maximality to the graph itself. □

Proposition 2 (Funding-cost transformation)**.** *For $\lambda \in (0,1)$ and fixed expert utility u, principal payoff is*

$$V = \frac{u - (1-\lambda)A}{\lambda}.$$

Thus minimizing A maximizes V, and

$$P_\lambda(u) = \frac{u - (1-\lambda)C_\lambda(u)}{\lambda}.$$

Proof. Eliminate B from $u = A - 2\lambda B$ and $V = A - 2B$. □

Proposition 3 (Convexity and slope bounds)**.** *For every conflicted type,*

$$C_\lambda(0) = 0, \qquad C_\lambda(\bar{u}_\lambda) = 1,$$

$$C_\lambda(u) \geq \frac{u}{s_1(\lambda)}, \qquad C_\lambda(u) \geq 1 - \frac{\bar{u}_\lambda - u}{s_0(\lambda)}.$$

Every nondegenerate secant of C_λ, including a secant with an endpoint at 0 or $\bar{u}_\lambda$, has slope in $[1/s_1(\lambda), 1/s_0(\lambda)]$. Hence every interior subgradient lies in the same interval.

Proof. For any outcome, expert utility is at most $s_1(\lambda)$ times total funding because signal 1 gives the highest utility per funded project. This gives the first bound. Relative to unconditional funding, every unit of omitted funding removes expected expert payoff at least $s_0(\lambda)$, because both signal-contingent funding payoffs are at least $s_0(\lambda)$. Thus $\bar{u}_\lambda - u \geq s_0(\lambda)(1 - A)$, giving the second bound.

It remains to justify the secant statement used below. Write $C = C_\lambda$ and $\bar{u} = \bar{u}_\lambda$. For a convex function with $C(0) = 0$, the right derivative at zero is the infimum of the endpoint secants:

$$C'_+(0) = \inf_{v>0} \frac{C(v)}{v} \geq \frac{1}{s_1(\lambda)}.$$

Convex secant slopes are ordered from left to right, so every secant on $[0, \bar{u}]$ is at least $C'_+(0)$. Likewise, since $C(\bar{u}) = 1$,

$$C'_-(\bar{u}) = \sup_{u<\bar{u}} \frac{1 - C(u)}{\bar{u} - u} \leq \frac{1}{s_0(\lambda)},$$

where the inequality follows from the second frontier bound. Every secant on $[0, \bar{u}]$ is at most $C'_-(\bar{u})$. These arguments include secants touching either endpoint. Interior subgradients are limits of nondegenerate secants and therefore satisfy the same bounds. □

To connect the recursive problem to the fixed-alphabet class, distinguish the root label $k \in K = \{1, 2, 3\}$, drawn publicly before the current signal, from an optional finite post-report label ℓ. Conditional on k and report r, let $\pi_{kr}(\ell)$ be the post-report label law, let $\rho_{kr\ell}$ be the funding probability, and initially allow a post-outcome public lottery over complete continuation outcomes after every (k, r, ℓ, d, θ).

Lemma 5 (Public-lottery and continuation compression)**.** *At every frontier promise one common global minimizer can be selected with a degenerate root label, degenerate post-report labels and post-outcome continuation lotteries, frontier continuation mechanisms, report-dependent funding probabilities ρ_r, and continuation promises $w_{r\theta}$ independent of the funding-lottery realization. The compression preserves the promised utility and the payoff from each report for each true signal, and weakly lowers expected funding. It is a payoff-pair reduction, not a public-process equivalence.*

Proof. Start from one complete global minimizer. After each realized branch (k, r, ℓ, d, θ), replace any continuation lottery (u_z, A_z) by one complete frontier mechanism delivering its mean promise $\widetilde{w}_{kr\ell d\theta} = \sum_z \xi_z u_z$. Convexity gives

$$C_\lambda(\widetilde{w}_{kr\ell d\theta}) \le \sum_z \xi_z C_\lambda(u_z) \le \sum_z \xi_z A_z.$$

This is a new continuation implementation; no barycenter is identified with an original strategy.

For fixed (k, r) jointly average the post-report label and funding realization:

$$\bar{\rho}_{kr} = \sum_\ell \pi_{kr}(\ell)\rho_{kr\ell}, \qquad \bar{w}_{kr\theta} = \sum_\ell \pi_{kr}(\ell)[\rho_{kr\ell}\widetilde{w}_{kr\ell 1\theta} + (1-\rho_{kr\ell})\widetilde{w}_{kr\ell 0\theta}]. \tag{C1}$$

The weights are common to every true signal choosing report r, so for each i

$$G_i^k(r) = as_i(\lambda)\bar{\rho}_{kr} + \delta[(1-\alpha_i)\bar{w}_{kr0} + \alpha_i \bar{w}_{kr1}] \tag{C2}$$

equals the original report payoff. Jensen's inequality weakly lowers continuation funding. This joint formula is essential when funding is label dependent.

Finally average the pre-signal root label, with law γ:

$$\rho_r = \sum_k \gamma(k)\bar{\rho}_{kr}, \qquad w_{r\theta} = \sum_k \gamma(k)\bar{w}_{kr\theta}. \tag{C3}$$

Because each root-label policy is truthful, averaging preserves both report inequalities and promise keeping; Jensen again weakly lowers cost. Null labels and funding branches have zero weight and receive arbitrary complete rules. Every transformed mechanism is feasible at the original promise, so starting from a global minimizer forces equality in aggregate. Applying the transformations sequentially yields one common minimizer with all stated properties. □

Define scores

$$x_r = -ah\rho_r + \delta w_{r0}, \qquad y_r = a\rho_r + \delta w_{r1}.$$

Truthfulness is

$$(1-\alpha_0)x_0 + \alpha_0 y_0 \ge (1-\alpha_0)x_1 + \alpha_0 y_1, \tag{IC0}$$

$$(1-\alpha_1)x_1 + \alpha_1 y_1 \ge (1-\alpha_1)x_0 + \alpha_1 y_0. \tag{IC1}$$

Lemma 6 (Score and outcome ordering)**.** *Every truthful reduced form has $x_0 \geq x_1$ and $y_1 \geq y_0$. A frontier optimum can be selected with*

$$w_{10} \leq w_{11}, \qquad w_{00} \geq w_{01}.$$

Proof. Let $D(\alpha) = (1-\alpha)(x_1 - x_0) + \alpha(y_1 - y_0)$ be the payoff from report 1 minus the payoff from report 0 at posterior α. Truthfulness gives $D(\alpha_0) \leq 0 \leq D(\alpha_1)$. Since D is affine and $\alpha_1 > \alpha_0$, its slope is nonnegative. If $x_1 > x_0$, then this nonnegative slope would imply $D(\alpha_0) > 0$; hence $x_0 \geq x_1$. If $y_1 < y_0$, then the same slope restriction gives $x_1 - x_0 \leq y_1 - y_0 < 0$, which would imply $D(\alpha_1) < 0$; hence $y_1 \geq y_0$.

For outcome ordering, pooling is enough. If $w_{10} > w_{11}$, replace both report-1 continuations by

$$z_1 = (1-\alpha_1)w_{10} + \alpha_1 w_{11}.$$

The report-1 payoff at posterior α_1 and its promise contribution are unchanged, while its payoff at α_0 falls by $\delta(\alpha_1 - \alpha_0)(w_{10} - w_{11})$. Thus (IC0) relaxes and (IC1) is unchanged. Jensen's inequality weakly lowers continuation cost. If $w_{00} < w_{01}$, replace both report-0 continuations by

$$z_0 = (1-\alpha_0)w_{00} + \alpha_0 w_{01}.$$

The report-0 payoff at α_0 and its promise contribution are unchanged, while its payoff at α_1 falls, so (IC1) relaxes and (IC0) is unchanged. Again Jensen's inequality weakly lowers continuation cost. Starting from a global minimizer, each weak cost reduction must be equality, giving the stated optimal selection. □

The exact recursion is therefore

$$C_\lambda(u) = \min_{\rho_r, x_r, y_r} \sum_{r=0}^{1} m_r \Bigg[a\rho_r + \delta(1-\alpha_r) C_\lambda\!\left(\frac{x_r + ah\rho_r}{\delta}\right) + \delta\alpha_r C_\lambda\!\left(\frac{y_r - a\rho_r}{\delta}\right) \Bigg], \tag{1}$$

subject to

$$u = \sum_{r=0}^{1} m_r \big[(1-\alpha_r)x_r + \alpha_r y_r\big],$$

(IC0)–(IC1), $0 \leq \rho_r \leq 1$, and both continuation arguments in $[0, \bar{u}_\lambda]$.

Lemma 5 proves necessity. Conversely, any feasible tuple is implemented with a degenerate root label, Bernoulli funding ρ_r, and complete frontier continuations

$$w_{r0} = \frac{x_r + ah\rho_r}{\delta}, \qquad w_{r1} = \frac{y_r - a\rho_r}{\delta}.$$

Promise keeping and (IC0)–(IC1), followed by the pointwise one-period-deviation principle, give a complete truthful mechanism. Thus the recursion is exact.

3 Known conflict: low authority, obstacles, and the high ceiling

The section begins with the closed-form low segment, then gives a pointwise obstacle map at an arbitrary frontier promise, and finally derives a no-override capacity that applies to every truthful mechanism. The obstacle map is local and does not imply a global phase ordering.

Proposition 4 (Static alignment threshold)**.** *Let $\bar{\lambda} = [2(1-\alpha_0)]^{-1}$. If $\lambda \geq \bar{\lambda}$, the expert and principal agree on signal-following, which is truthful and first best. If $\lambda < \bar{\lambda}$, the expert prefers funding after both signals.*

Proof. The low-signal funding payoff $s_0(\lambda)$ is nonpositive exactly at and above $\bar{\lambda}$. Under (A1), $s_1(\lambda) > 0$. The conclusion follows directly. □

Proposition 5 (General low-authority frontier)**.** *If $\delta\bar{u}_\lambda \geq s_0(\lambda)$, then*

$$C_\lambda(u) = \frac{u}{s_1(\lambda)}, \qquad 0 \leq u \leq a\bar{u}_\lambda.$$

An attaining recursive rule has

$$\rho_1(v) = \frac{v}{a\bar{u}_\lambda}, \qquad \rho_0(v) = 0,$$

zero continuation after report 1, and common report-0 continuation

$$z(v) = \frac{s_0(\lambda)}{\delta\bar{u}_\lambda} v.$$

Condition $\delta q \geq \alpha_0$ implies the displayed fixed-type condition throughout the conflicted region with $\lambda \geq 1/2$.

Proof. The first slope bound gives $C_\lambda(u) \geq u/s_1(\lambda)$. The stated rule remains in the interval because $z(v) \leq v$. A low-signal expert obtains $\delta z(v) = s_0 v/\bar{u}_\lambda$ from either report. A high-signal expert obtains $s_1 v/\bar{u}_\lambda$ truthfully and $s_0 v/\bar{u}_\lambda$ by understating. Promise keeping follows from $m_0 s_0 + m_1 s_1 = \bar{u}_\lambda$.

If $A(v)$ is recursive funding cost,

$$A(v) = am_1 \frac{v}{a\bar{u}_\lambda} + \delta m_0 A(z(v)).$$

The bounded solution is $A(v) = v/s_1(\lambda)$; uniqueness follows by iterating the homogeneous difference, whose sup norm is multiplied by $\delta m_0 < 1$.

For the uniform condition, put $D(\lambda) = \delta\bar{u}_\lambda - s_0(\lambda)$. At $\lambda = 1/2$, $D = \delta q - \alpha_0 \geq 0$, while $D'(\lambda) = 2\{(1-\alpha_0) - \delta(1-q)\} > 0$ because $\alpha_0 < q$ and $\delta < 1$. □

For $h > 0$ and fixed scores (x, y) define

$$\underline{\rho}(x,y) = \max\left\{0, -\frac{x}{ah}, \frac{y-\delta\bar{u}_\lambda}{a}\right\},$$

$$\overline{\rho}(x,y) = \min\left\{1, \frac{\delta\bar{u}_\lambda - x}{ah}, \frac{y}{a}\right\}, \qquad \rho^P(x,y) = \frac{y-x}{2a\lambda}.$$

Proposition 6 (Exact four-obstacle map)**.** *At every frontier promise for a conflicted $\lambda > 1/2$, one common global minimizer can be selected that is outcome ordered and satisfies*

$$\rho_1^* = \min\{\overline{\rho}(x_1,y_1), \rho^P(x_1,y_1)\}, \qquad \rho_0^* = \max\{\underline{\rho}(x_0,y_0), \rho^P(x_0,y_0)\}.$$

The scores in these formulas are the scores of that selected minimizer. Ties among funding, continuation, and pooling obstacles are allowed.

Proof. Fix a promise and start from one outcome-ordered global minimizer. Holding (x_r, y_r) fixed gives

$$w_{r0}(\rho) = \frac{x_r + ah\rho}{\delta}, \qquad w_{r1}(\rho) = \frac{y_r - a\rho}{\delta}.$$

Feasibility is exactly $\rho \in [\underline{\rho}(x_r,y_r), \overline{\rho}(x_r,y_r)]$, and $w_{r0} \le w_{r1}$ exactly when $\rho \le \rho^P(x_r,y_r)$.

For report r, write the conditional branch cost at fixed scores as

$$\psi_r(\rho) = a\rho + \delta(1-\alpha_r)C_\lambda(w_{r0}(\rho)) + \delta\alpha_r C_\lambda(w_{r1}(\rho)).$$

For report 1, move ρ upward toward $\widehat{\rho}_1 = \min\{\overline{\rho}, \rho^P\}$. Along that feasible ordered segment,

$$w_{10}(\rho) \le w_{10}(\rho+\Delta) \le w_{11}(\rho+\Delta) \le w_{11}(\rho).$$

Let c_b, c_g be the corresponding normalized bad- and good-continuation secants. Convexity gives $c_b \le c_g$, and Proposition 3 gives $c_g \ge 1/s_1$. Hence

$$\frac{\psi_1(\rho+\Delta) - \psi_1(\rho)}{a\Delta} = 1 + h(1-\alpha_1)c_b - \alpha_1 c_g \le 1 - s_1 c_g \le 0.$$

For report 0, move ρ downward toward $\widehat{\rho}_0 = \max\{\underline{\rho}, \rho^P\}$. The ordered segment reverses the secant ordering, $c_b \ge c_g$, while $c_g \le 1/s_0$, so

$$\frac{\psi_0(\rho) - \psi_0(\rho-\Delta)}{a\Delta} \ge 1 - s_0 c_g \ge 0.$$

The slope bounds include endpoint secants.

Both moves keep all four scores fixed, so promise keeping and both report constraints are unchanged; each segment remains inside the continuation and pooling bounds. Apply the two projections sequentially. Since the starting point is globally minimizing, neither weak cost reduction can be strict. The resulting common minimizer satisfies both formulas; affine pieces and coincident

obstacles only create ties. □

The fixed-score projection uses $h > 0$; Section 4 treats $h = 0$ directly.

Lemma 7 (Isolated unfavorable-report branch)**.** *Fix a conflicted type $\lambda > 1/2$ and a low-signal payoff g assigned to report 0. Ignoring the other report branch, the minimum conditional funding cost is*

$$K_0(g) = \begin{cases} \delta C_\lambda(g/\delta), & 0 \le g \le \delta\bar{u}_\lambda, \\ \delta + \dfrac{g - \delta\bar{u}_\lambda}{s_0(\lambda)}, & \delta\bar{u}_\lambda \le g \le \delta\bar{u}_\lambda + a s_0(\lambda). \end{cases}$$

An attaining selection pools the two continuation promises and uses

$$\widehat{\rho}_0(g) = \max\left\{0, \frac{g - \delta\bar{u}_\lambda}{a s_0(\lambda)}\right\}.$$

Let $\ell_1(\alpha) = (1-\alpha)x_1 + \alpha y_1$, $S_1 = y_1 - x_1$, and $\sigma_0 = g - \ell_1(\alpha_0) \ge 0$. Replacing report 0 by the isolated solution preserves the full truthful decomposition if and only if

$$S_1 - 2a\lambda\widehat{\rho}_0(g) \ge \frac{\sigma_0}{\alpha_1 - \alpha_0}. \tag{R0-COMP}$$

Whenever this condition holds, the replacement weakly lowers total funding cost; from a global minimizer it produces another global minimizer.

Proof. For fixed ρ, Jensen's inequality pools the continuations at $z = (g - a\rho s_0)/\delta$. The resulting cost is $a\rho + \delta C_\lambda(z)$. Every frontier secant has slope at most $1/s_0$, so this cost is weakly increasing in ρ and is minimized at the least feasible funding probability, giving the displayed formulas. The pooled branch has score spread $2a\lambda\widehat{\rho}_0(g)$ and high-signal payoff $g + (\alpha_1 - \alpha_0)2a\lambda\widehat{\rho}_0(g)$. Comparing it with $\ell_1(\alpha_1) = g - \sigma_0 + (\alpha_1 - \alpha_0)S_1$ yields (R0-COMP). Promise keeping, the report-1 branch, and (IC0) are unchanged. □

Proposition 7 (Exact current no-override capacity)**.** *For a conflicted type $\lambda \ge 1/2$, define*

$$Q(c) = \begin{cases} \dfrac{\alpha_1}{\alpha_0}c, & 0 \le c \le \alpha_0\bar{u}_\lambda, \\ \alpha_1\bar{u}_\lambda + \dfrac{1-\alpha_1}{1-\alpha_0}(c - \alpha_0\bar{u}_\lambda), & \alpha_0\bar{u}_\lambda \le c \le \bar{u}_\lambda, \end{cases}$$

and

$$\rho^N = \min\left\{1, \frac{\delta\bar{u}_\lambda}{a s_0(\lambda)}\right\}.$$

The largest promise deliverable by a truthful outcome with zero current funding after the unfavorable signal is

$$\widehat{u}_O = m_0\delta\bar{u}_\lambda + m_1\left[a\rho^N s_1(\lambda) + \delta Q\left(\bar{u}_\lambda - \frac{a\rho^N s_0(\lambda)}{\delta}\right)\right]. \tag{R0-CAP}$$

Every promise in $[0, \widehat{u}_O]$ is attainable with zero current unfavorable-signal funding, and every larger promise requires positive current funding after that signal. Under (A3), the capacity simplifies to

$$\widehat{u}_O = \delta \bar{u}_\lambda + a m_1 \frac{\alpha_1 - \alpha_0}{1 - \alpha_0}. \tag{R0-CAP-A3}$$

Proof. At a fixed public history, compress each current behavior to its funding probability and its two quality-contingent mean continuation utilities. If the signal-0 behavior funds with probability zero, its truthful payoff is at most $\delta \bar{u}_\lambda$. By (IC0), the signal-0 payoff from the signal-1 behavior is also at most this amount. If that behavior funds with probability ρ, its continuation mean under posterior α_0 is at most $c(\rho) = \bar{u}_\lambda - a\rho s_0/\delta$. For a given c, the largest continuation mean under posterior α_1 is $Q(c)$: continuation capacity is assigned first to the good-quality state because $\alpha_1/\alpha_0 > (1-\alpha_1)/(1-\alpha_0)$.

Feasibility gives $0 \le \rho \le \rho^N$. On the two pieces of Q, the signal-1 payoff $a\rho s_1 + \delta Q(c(\rho))$ has derivatives $a(\alpha_1 - \alpha_0)/(1-\alpha_0)$ and $ah(\alpha_1 - \alpha_0)/\alpha_0$, respectively. The first is positive and the second is weakly positive for $\lambda \ge 1/2$, so the payoff is maximized at ρ^N, which gives (R0-CAP). The bound is attained by assigning the report-0 behavior continuation $\bar{u}_\lambda$ after both qualities and choosing the report-1 continuations that attain Q; (IC0) binds and (IC1) holds. A public lottery with silence spans every lower promise.

Under (A3), $\rho^N = 1$ and $\delta \bar{u}_\lambda (1 - \alpha_0) > a s_0(\lambda)$, so the second piece of Q applies and yields (R0-CAP-A3). □

Theorem 2 (Low and high promises)**.** *Under (A3), let*

$$u_L = a \bar{u}_\lambda, \qquad \widehat{u}_O = \delta \bar{u}_\lambda + a m_1 \frac{\alpha_1 - \alpha_0}{1 - \alpha_0}.$$

Fix a conflicted $\lambda \in [1/2, \bar{\lambda})$. Then $u_L < \widehat{u}_O < \bar{u}_\lambda$. Rationing is optimal below u_L and review is optimal at u_L. If $\lambda > 1/2$, the obstacle map applies above u_L; if $\lambda = 1/2$, Section 4 gives the sharper ordered selection. The interval $[0, \widehat{u}_O]$ is exactly the set of promises feasible with zero current funding after the unfavorable signal.

Proof. The low statements follow from Proposition 5, and the exact capacity follows from Proposition 7. To compare the thresholds, write

$$\widehat{u}_O - u_L = \delta \bar{u}_\lambda - \frac{a m_0}{p} s_0(\lambda) > 0,$$

where $m_0 < p$ under (A1), and (A3) together with $\lambda \ge 1/2$ implies $\delta \bar{u}_\lambda \ge s_0(\lambda)$. Also

$$\bar{u}_\lambda - a m_0 s_0(\lambda) - \widehat{u}_O = a m_1 \frac{1 - \alpha_1}{1 - \alpha_0} s_0(\lambda) > 0,$$

while $\bar{u}_\lambda - a m_0 s_0(\lambda) < \bar{u}_\lambda$. Hence $\widehat{u}_O < \bar{u}_\lambda$. □

Under the scope of Theorem 2, substitution on the low frontier into the funding-cost transforma-

tion gives

$$P_\lambda(u) = \frac{2\alpha_1 - 1}{s_1(\lambda)} u,$$

which is strictly increasing. Hence an optimal initial promise is never strictly inside the rationing interval.

4 The ordered benchmark $\lambda = 1/2$

Here $h = 0$, $\bar{u} = q$, and $P(u) = 2u - C(u)$. The proof proceeds in four steps. Proposition 8 selects maximal favorable-report funding and minimal unfavorable-report funding. Lemmas 8 and 9 make binding exaggeration compatible with ordered favorable-report continuations. Lemma 10 adds the cheapest unfavorable-report branch and checks that this final replacement preserves the high-signal incentive constraint. Lemma 11 then reduces the review–override transition to a monotone scalar choice.

For a fixed promise u, let $\mathcal{D}(u)$ be the set of reduced decompositions

$$d = (\rho_0, \rho_1, w_{00}, w_{01}, w_{10}, w_{11}) \in [0,1]^2 \times [0,q]^4$$

satisfying promise keeping and (IC0)–(IC1), and let $J(d)$ be the one-period frontier cost. The constraint set $\mathcal{D}(u)$ is compact, because it is a closed subset of a compact box, and J is continuous. Define

$$\mathcal{M}(u) = \arg\min_{d \in \mathcal{D}(u)} J(d), \qquad \mathcal{M}(u;\rho) = \{d \in \mathcal{M}(u) : (\rho_0(d), \rho_1(d)) = \rho\}.$$

Lemma 5 guarantees that the original frontier problem has a minimizer represented in $\mathcal{D}(u)$. The first set is therefore nonempty and compact; the second is compact and is nonempty whenever the funding rule ρ is taken from a global minimizer. Every transformation below preserves feasibility and weakly lowers J. Therefore, when it is applied to a point of $\mathcal{M}(u)$, its image remains in $\mathcal{M}(u)$; transformations that hold funding fixed remain in $\mathcal{M}(u;\rho)$.

Proposition 8 (Extremal current funding)**.** *An optimum can be selected with*

$$\rho_1 = \min\{1, y_1/a\}, \qquad \rho_0 = \max\{0, (y_0 - \delta q)/a\}.$$

Proof. Holding y fixed, branch cost is

$$\phi_r(\rho; y) = a\rho + \delta\alpha_r C((y - a\rho)/\delta).$$

Every secant slope of C lies in $[1/\alpha_1, 1/\alpha_0]$. Increasing favorable-report funding weakly lowers ϕ_1; increasing unfavorable-report funding weakly raises ϕ_0. Move to the corresponding feasible endpoint. □

Lemma 8 (Binding exaggeration selection)**.** *Fix a promise and current funding rule. Among cost-minimizing truthful decompositions using that rule, one can be selected with (IC0) binding.*

Proof. Put $X = x_0 - x_1$ and $Y = y_1 - y_0$. Truthfulness is equivalent to

$$\frac{\alpha_0}{1-\alpha_0} Y \le X \le \frac{\alpha_1}{1-\alpha_1} Y.$$

If (IC0) is slack, move x_0 downward and x_1 upward while preserving

$$m_0(1-\alpha_0)x_0 + m_1(1-\alpha_1)x_1,$$

the bad-score contribution to promise keeping. This decreases X continuously until the lower displayed bound is reached, so (IC0) binds and (IC1) relaxes. Throughout the move $x_0 \ge x_1$ continues to hold. Because $h = 0$, $x_r = \delta w_{r0}$, and the two bad-outcome continuation promises move toward one another and remain in their original closed interval. Convexity weakly lowers their weighted continuation cost. Hence the endpoint is feasible and, starting from a fixed-funding global minimizer, belongs to $\mathcal{M}(u;\rho)$. □

Lemma 9 (Alternating projection)**.** *Fix current funding probabilities and a global minimizer with (IC0) binding. If $w_{10} > w_{11}$, alternating report-1 continuation pooling and rebinding (IC0) produces global minimizers with the same funding probabilities and multiplies the ordering violation by*

$$\kappa = \frac{m_0(\alpha_1 - \alpha_0)}{1-q} \in (0,1).$$

The compact limit has binding (IC0) and $w_{10} \le w_{11}$.

Proof. Let $v = w_{10} - w_{11} > 0$. Pool the two report-1 continuations at $z = (1-\alpha_1)w_{10} + \alpha_1 w_{11}$. The high-signal report-1 payoff and promise contribution are unchanged; the low-signal exaggeration payoff falls by $\delta(\alpha_1 - \alpha_0)v$. Jensen weakly lowers cost.

Keep both good-outcome scores fixed and move the bad scores according to

$$x_1' = \delta z + \Delta, \qquad x_0' = x_0 - \frac{m_1(1-\alpha_1)}{m_0(1-\alpha_0)}\Delta,$$

choosing Δ to restore (IC0). This preserves the bad-score contribution to promise keeping. At the endpoint the two report payoff lines agree at α_0. As shown below, $x_0' \ge x_1'$, while the unchanged good scores satisfy $y_1 \ge y_0$ by Lemma 6. Hence the report-1 line has weakly larger slope, so (IC1) holds. Direct calculation gives

$$\Delta = \delta \frac{m_0(\alpha_1 - \alpha_0)}{1-q} v,$$

so the new violation is $v' = \Delta/\delta = \kappa v$.

For completeness, $0 < \kappa < 1$ because

$$m_0(\alpha_1 - \alpha_0) < m_0(1-\alpha_0) < 1 - q.$$

The continuation bounds are also preserved. Score ordering before the pooling step gives $x_0 \geq x_1 = \delta w_{10}$, and direct substitution yields

$$x_0' - x_1' \geq \delta \frac{\alpha_0(1-\alpha_1)}{1-\alpha_0} v \geq 0.$$

Moreover,

$$\frac{x_1'}{\delta} = w_{11} + (1 - \alpha_1 + \kappa)v \leq w_{10} \leq q,$$

where $\kappa \leq \alpha_1$ follows from $m_0(\alpha_1 - \alpha_0) \leq \alpha_1(1-q)$. Thus $x_1' \geq 0$, $x_0' \geq x_1'$, and $x_0' \leq x_0 \leq \delta q$, so both bad-outcome continuation promises remain in $[0, q]$. The good-outcome continuations and funding probabilities, including boundary probabilities 0 and 1, are unchanged. The rebinding move takes a weighted contraction of x_0 and x_1, so convexity weakly lowers cost.

Consequently the composite map sends the closed set of fixed-funding global minimizers with binding (IC0) into itself. Since $v_n = \kappa^n v_0 \to 0$, every convergent subsequence has violation zero. Compactness supplies such a subsequence, and closedness gives one common limiting minimizer with binding (IC0) and $w_{10} \leq w_{11}$. □

Define the minimum report-0 branch cost for assigned low-signal payoff g by

$$K(g) = \begin{cases} \delta C(g/\delta), & 0 \leq g \leq \delta q, \\ \delta + (g - \delta q)/\alpha_0, & \delta q \leq g \leq \delta q + a\alpha_0. \end{cases}$$

Lemma 10 (Compatibility)**.** *Under $\delta q \geq a$, every frontier promise has a global minimizer satisfying simultaneously: (IC0) binds; $w_{10} \leq w_{11}$; and the report-0 branch uses $K(g)$ conditional on its assigned payoff g.*

Proof. Select extremal funding by Proposition 8; use Lemmas 8–9 without changing funding. Conditional on the report-0 payoff g, pool its two continuations. If current funding is ρ_0, the common continuation must be

$$t = \frac{g - a\alpha_0\rho_0}{\delta},$$

and the branch cost is $a\rho_0 + \delta C(t)$. This cost is weakly increasing in ρ_0, because every secant slope of C is at most $1/\alpha_0$. The least feasible funding probability is 0 for $g \leq \delta q$ and $(g - \delta q)/(a\alpha_0)$ for $g \geq \delta q$, yielding exactly $K(g)$. The replacement preserves the report-0 payoff at posterior α_0, and therefore preserves promise keeping and binding (IC0).

Let $S_r = y_r - x_r$. The selected report-1 branch has $S_1 = a\rho_1 + \delta(w_{11} - w_{10}) \geq a\rho_1$. The $K(g)$ branch has spread $S_0^K = 0$ below δq and $(g - \delta q)/\alpha_0 \leq a$ above it. Under review or override, $\rho_1 = 1$, so $S_1 \geq a \geq S_0^K$. Under rationing, extremal funding and score ordering imply $y_0 \leq y_1 < a \leq \delta q$ and $x_0 \leq \delta q$, hence $g = (1-\alpha_0)x_0 + \alpha_0 y_0 \leq \delta q$ and $S_0^K = 0$. Since the branches agree at posterior α_0, $S_1 \geq S_0^K$ implies (IC1). Thus the replacement is feasible and weakly cheaper; because its input is a global minimizer, its output is the same global minimum.

The maintained condition $\delta q \geq a$ is sufficient rather than sharp for the rationing comparison: $\delta q \geq a\alpha_0$ is enough, and (A3) also implies that inequality. □

Preservation of compatibility under report-0 replacement. Binding (IC0) fixes the two branch payoff lines at posterior α_0. For two affine payoff lines that agree at α_0, the high-posterior type prefers report 1 exactly when the report-1 line has weakly larger slope. Those slopes are the score spreads S_1 and S_0^K. The case analysis above proves $S_1 \geq S_0^K$ in rationing, review, and override. Thus replacing report 0 by $K(g)$ cannot destroy (IC1); this is the compatibility step on which the scalar reduction rests.

Put $k = q - \alpha_0$. Binding (IC0) and promise keeping give $u = g + kS$. In the review–override domain,

$$w_{10}(u,g) = \frac{qg - \alpha_0 u}{\delta k}, \qquad w_{11}(u,g) = \frac{(1-\alpha_0)u - (1-q)g - ak}{\delta k},$$

and reduced cost is

$$F(u,g) = am_1 + m_0 K(g) + \delta m_1(1-\alpha_1)C(w_{10}) + \delta m_1 \alpha_1 C(w_{11}).$$

Lemma 11 (Monotone reduced choice)**.** *For every $u \in [aq, q]$, compatible feasible g form a nonempty compact interval $G(u) = [L(u), U(u)]$ moving upward in the strong set order. The reduced cost has decreasing differences. Hence the least minimizer $g(u)$ is weakly increasing by the standard monotone-comparative-statics argument (Milgrom and Shannon 1994; Topkis 1998).*

Proof. The constraints $0 \leq g \leq \delta q + a\alpha_0$, $S \geq a$, and $w_{10}, w_{11} \in [0,q]$ are equivalent to

$$L(u) = \max\left\{0, \frac{\alpha_0}{q}u, \frac{(1-\alpha_0)u - (a+\delta q)(q-\alpha_0)}{1-q}\right\},$$

$$U(u) = \min\left\{\delta q + a\alpha_0, u - a(q-\alpha_0), \frac{\alpha_0}{q}u + \delta(q-\alpha_0), \frac{(1-\alpha_0)u - a(q-\alpha_0)}{1-q}\right\}.$$

Conversely, every g between these endpoints places both report-1 continuations in $[0,q]$, gives $S = (u-g)/(q-\alpha_0) \geq a$, and places g in the domain of K. The $K(g)$ report-0 branch has score spread in $[0,a]$, while the report-1 branch has spread at least a; because their payoff lines agree at α_0, both report constraints hold. Thus the endpoint inequalities are also sufficient. Nonemptiness follows from Lemma 10; at $u = aq$ the explicit low-frontier mechanism supplies the boundary case.

Every affine component defining L and U is weakly increasing. Hence L and U are weakly increasing, and the intervals move upward in the strong set order. Explicitly, if $u' < u''$, $g' \in G(u')$, $g'' \in G(u'')$, and $g' > g''$, then

$$L(u') \leq L(u'') \leq g'' < g' \leq U(u') \leq U(u''),$$

so each choice is feasible at the other promise.

For the first continuation term, w_{10} is increasing in g and decreasing in u; for the second, w_{11} is

decreasing in g and increasing in u. Monotonicity of convex secant slopes therefore gives decreasing differences for each composed C term. The term $K(g)$ has zero cross difference, so F has decreasing differences. Continuity of F and compactness of $G(u)$ make the argmin nonempty and compact, and its least element is well defined.

If $u' < u''$ but the least minimizers satisfy $g(u') > g(u'')$, strong-set-order feasibility makes the choices cross-feasible. Add the two optimality inequalities. Decreasing differences gives the reverse weak inequality, so equality must hold throughout. In particular, $g(u'')$ is also a minimizer at u', contradicting the definition of $g(u')$ as the least minimizer. This argument permits affine frontier segments and nonunique minimizers: in those cases the comparison may bind, and leastness rather than uniqueness delivers the monotone selection. □

Theorem 3 (Ordered authority regions)**.** *Under (A1)–(A3), there is*

$$u_H \in \left[q - a\alpha_0,\ q - \frac{a\alpha_0(1-q)}{1-\alpha_0}\right]$$

such that an optimum can be selected as rationing below aq, review between aq and u_H, and override above u_H, with possible boundary ties.

Proof. If $\rho_1 = 1$, binding (IC0) gives $u = x_1 + qS \geq aq$. If $\rho_1 < 1$, extremal funding gives $w_{11} = 0$, outcome ordering gives $w_{10} = 0$, and $u = qS < aq$. Thus rationing is exact below aq and favorable advice is fully funded above it.

In the review–override domain, override is equivalent to $g > \delta q$. Choose the least-minimizer selection from Lemma 11; its monotonicity implies a single switch. At $u = aq$, the low-frontier rule has $g = a\alpha_0 < \delta q$. At $u = q$, maximal expert utility forces full funding and all continuation promises equal to q, so $g = \delta q + a\alpha_0 > \delta q$.

If $g > \delta q$ and $S \geq a$, then $u = g + (q - \alpha_0)S > q - a\alpha_0$, giving the lower cutoff bound. Under review, $g \leq \delta q$ and feasibility imply

$$u \leq \min\{\delta q + (q - \alpha_0)S,\ \delta q + a - (1-q)S\}.$$

The maximum occurs at $S = a/(1-\alpha_0)$ and equals $q - a\alpha_0(1-q)/(1-\alpha_0)$, giving the upper bound. □

Corollary 1 (Cutoff bounds at the benchmark calibration)**.** *Let $p = 3/4$, $q = 2/5$, $\delta = 9/10$, and $\lambda = 1/2$. Then $\alpha_0 = 2/11$ and the review–override cutoff satisfies*

$$\frac{21}{55} \leq u_H \leq \frac{29}{75}.$$

The two endpoints are approximately 0.381818 *and* 0.386667*, respectively.*

The bounds are the specialization of the general interval in Theorem 3; the exact seven-state certificate below concerns the same calibration but is logically separate from the cutoff calculation.

5 On-path override: general patience result and exact calibration

This section studies how patience changes the known-type problem. A lazy-transition construction shows that the entire attainable payoff set expands with the discount factor. The advice-obedient subproblem is then solved exactly. Above the point at which its discount-independent ceiling is attainable, monotonicity and a local near-first-best construction produce a unique upper-tail threshold for necessary override. The final subsection returns to the benchmark calibration. It records a distinct low-patience effect in the appendix and verifies the exact seven-state certificate at $\delta = 9/10$.

Write

$$t_i := 2\alpha_i - 1, \qquad G_\lambda := m_1 s_1(\lambda), \qquad H := m_1 t_1 = p + q - 1.$$

The quantity H is the principal's payoff from funding exactly after favorable signals.

Lemma 12 (Signal-contingent first-best bound)**.** *For every admissible outcome and every $\delta \in (0,1)$, the principal's payoff is at most H.*

Proof. At a public history let $x_i \in [0,1]$ be the conditional probability of funding given current signal i under the selected behavioral strategy. The current expected principal payoff is

$$m_1 x_1 t_1 + m_0 x_0 t_0.$$

Under (A1), $t_1 > 0 > t_0$, so this expression is at most $m_1 t_1 = H$. Averaging over histories and dates preserves the bound. □

Proposition 9 (Payoff-set monotonicity in patience)**.** *Fix a known type λ and let $0 < \delta < \delta' < 1$. Every normalized vector consisting of expert utility, principal payoff, funding, and bad-project exposure attainable at δ is attainable at δ'. The same inclusion holds for advice-obedient outcomes. Consequently, the optimal principal payoff is weakly increasing in δ, and the minimum-funding frontier weakly falls at every promise common to the two feasible domains.*

Proof. Fix an attained outcome at δ and its selected strategy. At transformed date t, let N_t be the number of prior public advances. The mechanism applies the original date-N_t current rule to the virtual history. After the complete current original-format outcome, it draws an independent public bit $c_t \sim \text{Bernoulli}(\zeta)$: if $c_t = 1$, append the current outcome and signal to the virtual public and private histories; if $c_t = 0$, hold both virtual histories fixed. The full actual history retains every outcome and signal. At date t use the finite tagged spaces

$$\widetilde{M}_t = \bigsqcup_{n \le t} \{n\} \times M_n, \qquad \widetilde{C}_t = \bigsqcup_{n \le t} \{n\} \times C_n \times \{0,1\},$$

decoding a mismatched tag as a fixed default original message. These are standard Borel, and the virtual-history update is Borel. The bit is included in the public-randomization tag, so this update is a Borel function of the full history.

Set

$$\kappa = \frac{1-\delta'}{1-\delta}, \qquad \zeta = \frac{(1-\delta')\delta}{(1-\delta)\delta'},$$

so $\delta'\zeta = \kappa\delta$ and $\delta'(1-\zeta) = 1-\kappa$. For any selected normalized continuation coordinate X (expert utility, principal payoff, funding, or bad-project exposure), with original Bellman equation

$$X = (1-\delta)g + \delta\mathbb{E}X',$$

the transformed evaluation at the same virtual history is

$$(1-\delta')g + \delta'[(1-\zeta)X + \zeta\mathbb{E}X'] = X.$$

Contraction identifies this bounded fixed point as the actual transformed continuation value at every full history.

A one-period deviation induces an original current deviation at the virtual history. If Δg and ΔW are the original truthful-minus-deviation current and continuation differences, the transformed gain is

$$(1-\delta')\Delta g + \delta'\zeta\Delta W = \kappa[(1-\delta)\Delta g + \delta\Delta W] \geq 0.$$

This holds at every full private history, including histories differing only in held-date signals; the selected strategy ignores those signals in its virtual statistic but perfect recall is retained. The one-period-deviation principle gives sequential optimality. Conditional on $N_t = n$, the embedded virtual path has the original truthful date-n law, so advice obedience is preserved. Thus each attained payoff vector at δ is attained at δ', yielding the value and frontier monotonicity claims. The construction is outcome-specific and does not preserve the calendar-time public process. □

Call an admissible outcome *advice-obedient* if, under truthful play,

$$d_t\mathbf{1}\{\eta_t = 0\} = 0 \qquad \text{almost surely at every date } t.$$

Equivalently, conditional funding after signal 0 is zero at almost every public history under the truthful path law. The definition imposes no action restriction at formal histories outside that path-law support. It is representation independent and remains meaningful with nonatomic public randomization.

Lemma 13 (General advice-obedient upper bound)**.** *Fix a conflicted type. Every advice-obedient admissible outcome gives the principal at most*

$$\overline{V}_\lambda^{AO} = H\left[1 - \frac{(1-\alpha_1)s_0(\lambda)}{(1-\alpha_0)s_1(\lambda)}\right] < H. \tag{2}$$

Proof. Under advice obedience, every funded truthful-path date has signal 1, so principal and expert stage payoffs are in the fixed ratio t_1/s_1; hence $V = (t_1/s_1)U$. Let $\overline{U}$ be the supremum advice-obedient expert utility. At almost every truthful public history, hard-coding any inherited

private string in conditional support gives a fresh advice-obedient continuation, so its common maximal continuation utility is at most $\overline{U}$.

Choose an outcome with utility at least $\overline{U} - \xi$, and let x be current funding after signal 1. Freeze the complete current behavior used on that branch. Conditional on true signal i, the complete current public history has the factored law

$$Q_i(dh, d\theta) = K_\theta(dh)\beta_i(d\theta),$$

where the kernel K_θ is common to both signals because the frozen behavior and public mechanism do not observe the true signal. Thus Q_0, Q_1 are equivalent and their likelihood ratio equals $(1-\alpha_1)/(1-\alpha_0)$ on bad quality and α_1/α_0 on good quality. This transfers the almost-sure continuation bound to the signal-0 imitation branch.

Let $L^1 = \overline{U} - V(H)$ on the signal-1 current behavior, with means L_i^1 under Q_i, and let $L_0^0 \geq 0$ be the truthful signal-0 continuation loss. A low-signal expert can imitate the complete signal-1 current behavior and then use a continuation best response for her actual inherited history; its utility is the same $V(H)$. Hence

$$\delta(L_0^1 - L_0^0) \geq (1-\delta)xs_0.$$

Nonnegativity and the lower likelihood ratio give

$$L_1^1 \geq \frac{1-\alpha_1}{1-\alpha_0} L_0^1.$$

Promise keeping, $x \leq 1$, and $\xi \downarrow 0$ therefore imply

$$\overline{U} \leq m_1 \left[s_1 - \frac{1-\alpha_1}{1-\alpha_0} s_0 \right].$$

Multiplication by t_1/s_1 gives the bound, which is strict because the type is conflicted. □

Proposition 10 (Exact advice-obedient benchmark)**.** *Fix a conflicted type and write*

$$\Delta_\alpha := \alpha_1 - \alpha_0, \qquad D := \frac{\Delta_\alpha}{1-\alpha_0}, \qquad \overline{U}^{AO} := m_1 D,$$

$$\delta_\lambda^{AO} := \frac{s_0(\lambda)}{s_0(\lambda) + m_1 \Delta_\alpha}.$$

Let $U_\lambda^{AO,*}(\delta)$ be the maximum expert utility in the advice-obedient class. Then the maximum principal payoff in that class is

$$P_\lambda^{AO,*}(\delta) = \frac{t_1}{s_1(\lambda)} U_\lambda^{AO,*}(\delta).$$

If $\lambda \le 1/2$, then

$$U_\lambda^{AO,*}(\delta) = \begin{cases} 0, & 0 < \delta < \delta_\lambda^{AO}, \\ \overline{U}^{AO}, & \delta_\lambda^{AO} \le \delta < 1. \end{cases} \tag{AO1}$$

If $\lambda > 1/2$, define

$$\delta_\lambda^L := \frac{s_0(\lambda)}{\bar{u}_\lambda} = \frac{s_0(\lambda)}{s_0(\lambda) + 2\lambda m_1 \Delta_\alpha}.$$

Then $\delta_\lambda^L < \delta_\lambda^{AO}$ and

$$U_\lambda^{AO,*}(\delta) = \begin{cases} 0, & 0 < \delta < \delta_\lambda^L, \\ \dfrac{(1-\delta)m_1(2\lambda-1)\Delta_\alpha}{\alpha_0(1-\delta) - \delta m_1 \Delta_\alpha}, & \delta_\lambda^L \le \delta < \delta_\lambda^{AO}, \\ \overline{U}^{AO}, & \delta_\lambda^{AO} \le \delta < 1. \end{cases} \tag{AO2}$$

In particular, the discount-independent upper bound in Lemma 13 is attained for every $\delta \ge \delta_\lambda^{AO}$.

Proof. The payoff ratio follows from the same path-law argument as in Lemma 13: under truthful play every funded date has signal 1, so principal and expert expected stage payoffs are in the fixed ratio $t_1/s_1(\lambda)$. It remains to maximize expert utility.

Let $\overline{U}$ be the supremum advice-obedient utility. Choose a maximizing sequence of initial advice-obedient outcomes with utilities $U_n \uparrow \overline{U}$, and pass to a subsequence along which the current funding probabilities after signal 1 satisfy $x_n \to x \in [0,1]$. If $x = 0$, promise keeping and the fact that every continuation utility is at most $\overline{U}$ give

$$U_n \le \delta\overline{U} + (1-\delta)m_1 x_n s_1(\lambda),$$

so $\overline{U} \le \delta\overline{U}$ and therefore $\overline{U} = 0$. Hence suppose $x > 0$.

For outcome n, let $Q_{i,n}$ be the law of the complete current public history generated when the true signal is i and the equilibrium current behavior used after signal 1 is followed. The two laws have the likelihood ratio in (AO-RN). Let $V_n(H)$ be the expert's continuation utility at current public history H and put

$$L_n(H) := \overline{U} - V_n(H).$$

The continuation bound holds $Q_{1,n}$-almost surely because the signal-1 branch is truthful and, after conditioning on the realized public history and any inherited private string in its conditional support, the selected continuation remains advice-obedient. Its continuation utility is nonnegative because every future truthful-path stage funds only after signal 1, whose conditional expert payoff $s_1(\lambda)$ is positive. Mutual absolute continuity in (AO-RN) transfers both bounds to $Q_{0,n}$-almost every history. The transfer concerns the common continuation utility only; continuation funding and the selected continuation strategy need not coincide across inherited private histories. Thus $0 \le L_n \le \overline{U}$

under both laws. Write

$$\ell_{i,n} := \mathbb{E}_{Q_{i,n}} L_n.$$

Let $y_n \geq 0$ be the expected continuation loss on the truthful signal-0 branch. Low-signal incentive compatibility gives

$$\ell_{0,n} - y_n \geq c x_n, \qquad c := \frac{(1-\delta)s_0(\lambda)}{\delta}, \tag{AO3}$$

This inequality also implies $y_n + c x_n \leq \overline{U}$, because $\ell_{0,n} \leq \overline{U}$. We next solve the continuation-loss allocation problem without imposing any restriction on public labels, funding randomization, or the number of continuation states. For any random variable L with $0 \leq L \leq u$ under the current signal-1 branch, define

$$z_0 := \mathbb{E}_{Q_0}[L\mathbf{1}\{\theta = 0\}], \qquad z_1 := \mathbb{E}_{Q_0}[L\mathbf{1}\{\theta = 1\}].$$

Then

$$0 \leq z_0 \leq (1-\alpha_0)u, \qquad 0 \leq z_1 \leq \alpha_0 u, \qquad z_0 + z_1 = z := \mathbb{E}_{Q_0} L,$$

and (AO-RN) gives

$$\mathbb{E}_{Q_1} L = \frac{1-\alpha_1}{1-\alpha_0} z_0 + \frac{\alpha_1}{\alpha_0} z_1.$$

Because the first coefficient is smaller, the minimum puts as much loss as possible on bad-quality histories. Hence

$$\psi_u(z) := \begin{cases} \dfrac{1-\alpha_1}{1-\alpha_0} z, & 0 \leq z \leq (1-\alpha_0)u, \\ \dfrac{\alpha_1}{\alpha_0} z + \left(1 - \dfrac{\alpha_1}{\alpha_0}\right) u, & (1-\alpha_0)u \leq z \leq u \end{cases} \tag{AO4}$$

is the exact lower envelope of the bounded-loss linear program. As a statement about random variables, its endpoint profiles use only losses 0 and u, with randomization within the relevant quality cell. At this stage u may be only a supremum, so this geometric observation is not a claim that a continuation mechanism delivering utility u already exists. Dynamic attainment is established separately below by the explicit active-state mechanisms at the candidate boundary values. Correlation with funding, messages, or additional public labels cannot improve the lower bound because those variables have the same conditional law under Q_0 and Q_1 once quality is fixed.

Promise keeping for outcome n is

$$U_n \leq \delta\overline{U} + (1-\delta) m_1 x_n s_1(\lambda) - \delta\big(m_0 y_n + m_1 \ell_{1,n}\big).$$

For fixed x_n, (AO3), monotonicity of $\psi_{\overline{U}}$, and $y_n \geq 0$ imply

$$m_0 y_n + m_1 \ell_{1,n} \geq m_0 y_n + m_1 \psi_{\overline{U}}(y_n + c x_n) \geq m_1 \psi_{\overline{U}}(c x_n).$$

Thus rewarding the truthful signal-0 branch with the maximal continuation utility, $y_n = 0$, is not an assumption: it minimizes the total continuation loss compatible with low-signal truthfulness.

Letting $n \to \infty$ and using continuity and homogeneity of ψ gives, with $v := \overline{U}/x$,

$$(1-\delta)v \leq (1-\delta)m_1 s_1(\lambda) - \delta m_1 \psi_v(c), \qquad v \geq c. \tag{AO5}$$

Because $x \leq 1$, every upper bound on feasible v is also an upper bound on $\overline{U} = xv$. The boundary values identified below are attained with $x = 1$ by explicit two-state mechanisms. Thus (AO5) is used only as a necessary upper bound; the matching constructions, rather than any sufficiency claim for the abstract loss program, make the solution exact.

If $c \leq (1-\alpha_0)v$, the first line of (AO4) yields

$$v \leq m_1 \left[s_1(\lambda) - \frac{1-\alpha_1}{1-\alpha_0} s_0(\lambda) \right] = \overline{U}^{AO}.$$

The endpoint $v = \overline{U}^{AO}$ lies in this region exactly when

$$c \leq (1-\alpha_0)\overline{U}^{AO} = m_1 \Delta_\alpha,$$

which is equivalent to $\delta \geq \delta_\lambda^{AO}$.

If $c > (1-\alpha_0)v$, the second line of (AO4) gives

$$\left[(1-\delta) - \delta m_1 \left(\frac{\alpha_1}{\alpha_0} - 1 \right) \right] v \leq (1-\delta) m_1 (2\lambda - 1) \left(\frac{\alpha_1}{\alpha_0} - 1 \right). \tag{AO6}$$

Put

$$B_\delta := \alpha_0(1-\delta) - \delta m_1 \Delta_\alpha.$$

For $\lambda > 1/2$, conflictedness implies $0 < s_0(\lambda) < \alpha_0$. Hence

$$\delta_\lambda^L = \frac{s_0(\lambda)}{\bar{u}_\lambda} < \frac{s_0(\lambda)}{s_0(\lambda) + m_1 \Delta_\alpha} = \delta_\lambda^{AO} < \frac{\alpha_0}{\alpha_0 + m_1 \Delta_\alpha},$$

so $B_\delta > 0$ throughout the middle interval. Solving (AO6) gives the middle expression $U_\lambda^M(\delta)$ in (AO2), and

$$U_\lambda^M(\delta) - c = \frac{\alpha_0(1-\delta)[\delta \bar{u}_\lambda - s_0(\lambda)]}{\delta B_\delta}.$$

Thus a positive feasible solution first appears at $\delta = \delta_\lambda^L$. At that endpoint $U^M = c$ and the good- and bad-quality histories both move to silence. At $\delta = \delta_\lambda^{AO}$,

$$U_\lambda^M(\delta) = \overline{U}^{AO}, \qquad c = (1-\alpha_0)\overline{U}^{AO},$$

so the middle construction joins continuously with the ceiling construction.

For $\lambda \leq 1/2$, the right-hand side of (AO6) is nonpositive. If its coefficient is nonnegative, no $v > 0$ satisfies the inequality. If the coefficient is negative, (AO6) gives $v \geq U_\lambda^M(\delta)$, while direct

subtraction yields

$$U_\lambda^M(\delta) - \frac{c}{1-\alpha_0} = \frac{\alpha_0(1-\delta)[\delta m_1 \Delta_\alpha - (1-\delta)s_0(\lambda)]}{\delta(1-\alpha_0)B_\delta} > 0$$

for $\delta < \delta_\lambda^{AO}$ and $B_\delta < 0$. This contradicts the second-region requirement $v < c/(1-\alpha_0)$. Hence no positive advice-obedient utility is feasible below δ_λ^{AO} when $\lambda \leq 1/2$.

It remains to verify attainment and sequential truthfulness. Both mechanisms have an active public state and absorbing silence. In the ceiling region, assign utility $\overline{U}^{AO}$ to the active state. After report 0, reject and remain active. After report 1, fund; following bad quality, enter silence with probability

$$\varepsilon_\delta = \frac{(1-\delta)s_0(\lambda)}{\delta(1-\alpha_0)\overline{U}^{AO}},$$

and otherwise remain active. The condition $\varepsilon_\delta \leq 1$ is exactly $\delta \geq \delta_\lambda^{AO}$. The low-signal constraint binds, the high-signal gain is $(1-\delta)\overline{U}^{AO}/m_1 > 0$, and promise keeping delivers $\overline{U}^{AO}$.

For the middle region, assign the displayed value $U = U_\lambda^M(\delta)$ to the active state. After report 0, reject and remain active. After report 1, fund; after bad quality enter silence for sure, and after good quality enter silence with probability

$$\gamma_\delta = \frac{c - (1-\alpha_0)U}{\alpha_0 U} \in [0,1].$$

The expected signal-0 loss after report 1 is c, so the low-signal constraint binds. Equation (AO6) at equality is promise keeping, and the high-signal gain is $(1-\delta)U/m_1 > 0$. At every private history whose public state is active, future primitives are i.i.d. and the same two inequalities apply independently of inherited signals; at silence both reports give zero. The mechanism specifies these same rules at every formal active-state history, including histories that are null under truthful play. The pointwise one-period-deviation principle therefore gives sequential optimality after every private history. Silence attains the zero regions. This proves (AO1)–(AO2). □

Theorem 4 (Near-first-best authority under patience)**.** *Maintain (A1) and fix a conflicted known type λ. Let $P_\lambda^*(\delta)$ be the optimal known-type payoff at discount factor δ. Then*

$$\lim_{\delta \uparrow 1} P_\lambda^*(\delta) = H.$$

Proof. Lemma 12 gives the upper bound H. Put $G_\lambda = m_1 s_1(\lambda)$ and choose $r_0 > 0$ with $I = [G_\lambda - r_0, G_\lambda + r_0] \subset (0, \bar{u}_\lambda)$. For $u \in I$ define

$$(\rho_0(u), \rho_1(u)) = \begin{cases} (0, u/G_\lambda), & u \leq G_\lambda, \\ ((u - G_\lambda)/(m_0 s_0), 1), & u \geq G_\lambda. \end{cases} \tag{NF1}$$

Then $d(u) = \rho_1(u) - \rho_0(u)$ is uniformly positive on I and $m_0\rho_0 s_0 + m_1\rho_1 s_1 = u$. With $a = 1-\delta$, set

$$z(u) = \frac{d(u)s_0}{\delta(1-\alpha_0)}, \qquad c(u) = -m_0(1-\alpha_0)z(u),$$

and

$$w_{00} = u + a(c+z), \qquad w_{01} = w_{10} = w_{11} = u + ac. \tag{NF2}$$

The truthful mean continuation is u, so $u = au + \delta\mathbb{E}w_{\eta\theta}$. Low-signal exaggeration binds, while the high-signal truthful gain is

$$ad(u)\left[s_1 - \frac{1-\alpha_1}{1-\alpha_0}s_0\right] = ad(u)\frac{\alpha_1-\alpha_0}{1-\alpha_0} > 0. \tag{NF3}$$

Let b_0 be the distance from I to the endpoints of $[0, \bar{u}_\lambda]$. For $\delta \geq 1/2$, all increments in (NF2) have absolute value at most $M(1-\delta)$ for $M = 2s_0/(1-\alpha_0)$. Hence

$$\delta_0 = \max\{1/2, 1 - b_0/M\} < 1$$

ensures every continuation promise lies in the feasible authority interval. The public state is the promise U_t. At $u \in I$ use (NF1)–(NF2); at $v \notin I$, publicly choose full funding forever with probability $v/\bar{u}_\lambda$ and silence otherwise. All remaining formal histories lead to silence. The exit lottery delivers promise v exactly and is sequentially truthful.

The truthful evaluation operator equals $Tf(u) = au + \delta\mathbb{E}f(w_{\eta\theta}(u))$ on I and $Tf(v) = v$ outside. It is a contraction, and (NF2) makes the identity function its fixed point. Thus the public state is the actual continuation utility at every formal state; (NF3) and the one-period-deviation principle establish sequential truthfulness.

For the public filtration immediately before each report, let $\tau = \inf\{t : U_t \notin I\}$. Since U_t is publicly known, τ is a stopping time and the exit promise U_τ is known before the exit lottery. Put $X_t = U_{t\wedge\tau}$. It is a bounded martingale with differences bounded by Ma, so

$$\mathbb{E}[(X_t - G_\lambda)^2] \leq M^2a^2t, \qquad \mathbb{E}|X_t - G_\lambda| \leq Ma\sqrt{t}. \tag{NF4}$$

The local principal flow is

$$\phi(u) = \begin{cases} Hu/G_\lambda, & u \leq G_\lambda, \\ H + t_0(u - G_\lambda)/s_0, & u \geq G_\lambda, \end{cases}$$

and $0 \leq H - \phi(u) \leq L|u - G_\lambda|$ for finite L. The exit lottery has principal value $J(v) = v(2q-1)/\bar{u}_\lambda \leq H$. With the convention $\delta^\infty = 0$, exact stopping-time accounting gives

$$V_\delta^I = \mathbb{E}\left[a\sum_{t<\tau}\delta^t\phi(U_t) + \delta^\tau J(U_\tau)\right],$$

and therefore

$$H - V_\delta^I \le LMa^2 \sum_{t\ge 0} \delta^t \sqrt{t} + 2\mathbb{E}[\delta^\tau]. \tag{NF5}$$

The sum is $O(a^{-3/2})$. With $T_a = \lceil a^{-3/2} \rceil$, Doob's inequality and (NF4) give

$$\Pr(\tau \le T_a) \le \frac{M^2}{r_0^2}(\sqrt{a} + a^2), \qquad \mathbb{E}[\delta^\tau] \le \Pr(\tau \le T_a) + e^{-a^{-1/2}}.$$

Thus, for every δ on a full tail,

$$0 \le H - V_\delta^I \le C\sqrt{1-\delta} + 2e^{-(1-\delta)^{-1/2}} \to 0.$$

The local flow is counted only for $t < \tau$ and the normalized exit continuation starts at τ, so (NF5) contains no double counting. This proves the full limit, not merely a subsequential result. □

Corollary 2 (Unique upper-tail threshold)**.** *Maintain (A1) and fix a conflicted known type. Define*

$$\delta_\lambda^\dagger := \inf\left\{\delta \in [\delta_\lambda^{AO}, 1) : P_\lambda^*(\delta) > \overline{V}_\lambda^{AO}\right\}.$$

Then $\delta_\lambda^\dagger < 1$. For $\delta_\lambda^{AO} \le \delta < \delta_\lambda^\dagger$, the advice-obedient mechanism in Proposition 10 is globally optimal. For every $\delta > \delta_\lambda^\dagger$, every globally optimal outcome satisfies $\Pr(d_t = 1, \eta_t = 0) > 0$ for some date t under truthful play. No claim is made at $\delta = \delta_\lambda^\dagger$.

Proof. For each fixed δ, existence of a global optimum follows from Proposition 1: the attained frontier is continuous on the compact promise interval, so the principal's transformed objective attains its maximum. Proposition 10 supplies an advice-obedient outcome worth $\overline{V}_\lambda^{AO}$ for every $\delta \ge \delta_\lambda^{AO}$, and Lemma 13 gives the matching upper bound within that class.

Let

$$S := \{\delta \in [\delta_\lambda^{AO}, 1) : P_\lambda^*(\delta) > \overline{V}_\lambda^{AO}\}.$$

Theorem 4 and $\overline{V}_\lambda^{AO} < H$ imply that S is nonempty. Proposition 9 implies that S is an upper set: if $\delta \in S$ and $\delta' > \delta$, then $\delta' \in S$. Hence its infimum $\delta_\lambda^\dagger$ is the unique boundary of this upper set; the argument does not determine membership of the boundary point itself.

If $\delta_\lambda^{AO} \le \delta < \delta_\lambda^\dagger$, then $\delta \notin S$, so $P_\lambda^*(\delta) \le \overline{V}_\lambda^{AO}$. The feasible advice-obedient mechanism attains the reverse inequality, and is therefore globally optimal. If $\delta > \delta_\lambda^\dagger$, the definition of the infimum supplies some $\hat{\delta} \in S$ with $\hat{\delta} < \delta$; monotonicity then gives $P_\lambda^*(\delta) \ge P_\lambda^*(\hat{\delta}) > \overline{V}_\lambda^{AO}$. No globally optimal outcome can be advice-obedient, because every such outcome is bounded by $\overline{V}_\lambda^{AO}$. Failure of advice obedience means that for some date t the nonnegative random variable $d_t \mathbf{1}\{\eta_t = 0\}$ is not almost surely zero, equivalently $\Pr(d_t = 1, \eta_t = 0) > 0$. The endpoint is left open because monotonicity alone does not determine whether the strict inequality holds at the infimum. □

5.1 A low-patience caveat at the benchmark

The main text emphasizes the upper-tail threshold. At the benchmark, a separate low-patience construction shows why that threshold should not be described as a global one-crossing result.

Proposition 11 (Low-patience override at the benchmark)**.** *Fix $p = 3/4$, $q = 2/5$, and $\lambda = 1/2$. Then the optimal advice-obedient payoff is zero for $0 < \delta < 5/11$ and equals $2/15$ for $\delta \geq 5/11$. Nevertheless, unrestricted optimal value is strictly positive for every $\delta > 0$. Hence every optimum overrides on path for*

$$0 < \delta < \frac{5}{11}.$$

Combined with the exact certificate at $\delta = 9/10$ and Proposition 9, every optimum also overrides for every $\delta \geq 9/10$. No conclusion is asserted for the intervening discount factors.

Proof. At the benchmark,

$$s_0 = \frac{2}{11}, \qquad s_1 = \frac{2}{3}, \qquad \bar{u}_\lambda = \frac{2}{5}, \qquad H = \frac{3}{20}, \qquad 2q - 1 = -\frac{1}{5}.$$

Proposition 10 gives $\delta_\lambda^{AO} = 5/11$ and the stated advice-obedient values.

For unrestricted value, choose any

$$0 < x \leq \min\left\{1, \frac{\delta \bar{u}_\lambda}{(1-\delta)s_0}\right\}, \qquad z := \frac{(1-\delta)xs_0}{\delta}.$$

At the initial state, fund with probability x after report 1 and then enter silence. After report 0, reject and use a public lottery that enters perpetual full funding with probability $z/\bar{u}_\lambda$ and silence otherwise. The low-signal reporting constraint binds, while the high-signal constraint is strict because $s_1 > s_0$.

The principal's normalized payoff is

$$\begin{aligned}(1-\delta)xH + \delta m_0 \frac{z}{\bar{u}_\lambda}(2q-1) &= (1-\delta)x\left[H + (2q-1)\frac{m_0 s_0}{\bar{u}_\lambda}\right] \\ &= \frac{(1-\delta)x}{10} > 0.\end{aligned}$$

Thus unrestricted value is positive while advice-obedient value is zero below $5/11$. The high-patience statement follows from the exact $\delta = 9/10$ comparison below and payoff-set monotonicity. □

This low-patience mechanism uses a rare full-funding continuation as a blunt reporting reward. It is economically distinct from the near-first-best authority account that drives the upper-tail theorem and is therefore kept out of the main narrative.

5.2 Exact benchmark certificate

Set

$$p = \frac{3}{4}, \quad q = \frac{2}{5}, \quad \delta = \frac{9}{10}, \quad \lambda = \frac{1}{2},$$

so

$$m_0 = \frac{11}{20}, \quad m_1 = \frac{9}{20}, \quad \alpha_0 = \frac{2}{11}, \quad \alpha_1 = \frac{2}{3},$$

and the joint probabilities of $(\eta, \theta) = (0,0), (0,1), (1,0), (1,1)$ are $9/20, 1/10, 3/20, 3/10$.

Lemma 14 (Advice-obedient bound at the benchmark calibration)**.** *Every advice-obedient admissible outcome at this calibration gives the principal at most* $2/15$*.*

Proof. At the calibration, $H = 3/20$, $s_0 = \alpha_0 = 2/11$, and $s_1 = \alpha_1 = 2/3$. Substituting these values into Lemma 13 gives

$$\overline{V}^{AO}_{1/2} = \frac{3}{20}\left[1 - \frac{(1/3)(2/11)}{(9/11)(2/3)}\right] = \frac{2}{15}.$$

□

The public states are

$$\mathcal{S} = \left\{0, \frac{1}{5}, \frac{13}{50}, \frac{7}{25}, \frac{57}{200}, \frac{3}{10}, \frac{2}{5}\right\}.$$

State 0 is absorbing silence and state $2/5$ is absorbing full funding. At every other state, funding is deterministic and equals the report: $d = r$. Within a period, the expert reports, the funding decision is made, quality is publicly revealed, and then the public transition lottery in Table 1 is drawn. Thus each table cell conditions only on public variables available at the transition date. "Stay" assigns the residual probability to the current state. At the two absorbing states every report–quality pair returns to the same state, so the table together with these absorption rules is a complete public kernel after every formal history, including histories reached only after a deviation.

The transition table is:

Table 1: Seven-state authority mechanism

State	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$1/5$	0 w.p. $5/81$, else stay	stay	0 w.p. $5/27$, else stay	stay
$13/50$	stay	stay	$1/5$ w.p. $40/81$, else stay	stay
$7/25$	$57/200$ w.p. $40/81$, else stay	stay	$1/5$ w.p. $1/27$; $13/50$ w.p. $26/27$	stay
$57/200$	$3/10$ w.p. $1/243$, else stay	$3/10$	$13/50$ w.p. $22/27$; $7/25$ w.p. $5/27$	stay
$3/10$	stay	$2/5$ w.p. $5/18$, else stay	$7/25$ w.p. $19/27$; $57/200$ w.p. $8/27$	stay

Let u_s and P_s denote expert and principal continuation values. Under truthful reporting, the seven expert Bellman equations are

$$\begin{aligned} u_0 &= \delta u_0, \\ u_{1/5} &= \frac{3}{100} + \delta\left(\frac{1}{18}u_0 + \frac{17}{18}u_{1/5}\right), \end{aligned}$$

$$
\begin{aligned}
u_{13/50} &= \frac{3}{100} + \delta\left(\frac{2}{27}u_{1/5} + \frac{25}{27}u_{13/50}\right),\\
u_{7/25} &= \frac{3}{100} + \delta\left(\frac{1}{180}u_{1/5} + \frac{13}{90}u_{13/50} + \frac{113}{180}u_{7/25} + \frac{2}{9}u_{57/200}\right),\\
u_{57/200} &= \frac{3}{100} + \delta\left(\frac{11}{90}u_{13/50} + \frac{1}{36}u_{7/25} + \frac{101}{135}u_{57/200} + \frac{11}{108}u_{3/10}\right),\\
u_{3/10} &= \frac{3}{100} + \delta\left(\frac{19}{180}u_{7/25} + \frac{2}{45}u_{57/200} + \frac{37}{45}u_{3/10} + \frac{1}{36}u_{2/5}\right),\\
u_{2/5} &= \frac{1}{25} + \delta u_{2/5}.
\end{aligned}
$$

Their unique solution is exactly the state labeling $u_s = s$. The continuation system is a discounted finite-state contraction, so these labels are the actual truthful continuation utilities from every formal public state, not merely a solution to promise equations along the initial path.

For true signal i and report r, let $G_i(r \mid s)$ be normalized current expert payoff plus discounted expected continuation utility, using the transition kernel in the (r, θ) cells. The report-payoff checks are shown in Table 2. Each entry is exact. The low-signal differences $G_0(0 \mid s) - G_0(1 \mid s)$ are zero except at $13/50$, where they equal $1/275 > 0$. The high-signal differences $G_1(1 \mid s) - G_1(0 \mid s)$ are, state by state, $0, 8/135, 13/225, 8/135, 173/3375, 2/45, 0$. Hence truthful reporting is a one-period best reply at every public state, including both absorbing states.

Table 2: Exact one-period reporting checks

State	$G_0(0)$	$G_0(1)$	$G_1(1)$	$G_1(0)$
0	0	0	0	0
1/5	47/275	47/275	53/225	119/675
13/50	117/500	1267/5500	1313/4500	117/500
7/25	349/1375	349/1375	39/125	853/3375
57/200	259/1000	259/1000	2851/9000	7169/27000
3/10	151/550	151/550	149/450	43/150
2/5	104/275	104/275	32/75	32/75

Because future projects are i.i.d. and the public state is a sufficient statistic for the mechanism, the same one-period inequalities apply after every inherited private history leading to a given public state. Bounded discounted payoffs and the pointwise one-period-deviation principle therefore upgrade these checks to sequential truthfulness after every private history, including histories reached only after a prior deviation.

Under truthful play, the exact public-state transition matrix, in the state order displayed above,

is

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/18 & 17/18 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2/27 & 25/27 & 0 & 0 & 0 & 0 \\ 0 & 1/180 & 13/90 & 113/180 & 2/9 & 0 & 0 \\ 0 & 0 & 11/90 & 1/36 & 101/135 & 11/108 & 0 \\ 0 & 0 & 0 & 19/180 & 2/45 & 37/45 & 1/36 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The only recurrent classes are the two absorbing states 0 and 2/5; the five interior states are transient.

The seven principal Bellman equations have the same continuation coefficients. At an interior state current normalized principal payoff is 3/200; at full funding it is −1/50:

$$\begin{aligned} P_0 &= \delta P_0, \\ P_{1/5} &= \frac{3}{200} + \delta\left(\frac{1}{18}P_0 + \frac{17}{18}P_{1/5}\right), \\ P_{13/50} &= \frac{3}{200} + \delta\left(\frac{2}{27}P_{1/5} + \frac{25}{27}P_{13/50}\right), \\ P_{7/25} &= \frac{3}{200} + \delta\left(\frac{1}{180}P_{1/5} + \frac{13}{90}P_{13/50} + \frac{113}{180}P_{7/25} + \frac{2}{9}P_{57/200}\right), \\ P_{57/200} &= \frac{3}{200} + \delta\left(\frac{11}{90}P_{13/50} + \frac{1}{36}P_{7/25} + \frac{101}{135}P_{57/200} + \frac{11}{108}P_{3/10}\right), \\ P_{3/10} &= \frac{3}{200} + \delta\left(\frac{19}{180}P_{7/25} + \frac{2}{45}P_{57/200} + \frac{37}{45}P_{3/10} + \frac{1}{36}P_{2/5}\right), \\ P_{2/5} &= -\frac{1}{50} + \delta P_{2/5}. \end{aligned}$$

Solving gives

$$P_0 = 0, \quad P_{1/5} = \frac{1}{10}, \quad P_{13/50} = \frac{13}{100}, \quad P_{7/25} = \frac{162749}{1211600},$$
$$P_{57/200} = \frac{1261599}{9692800}, \quad P_{3/10} = \frac{52131}{484640}, \quad P_{2/5} = -\frac{1}{5}.$$

In particular,

$$P_{7/25} - \frac{2}{15} = \frac{3607}{3634800} > 0.$$

A concrete three-step witness route is as follows. From 7/25, event $(\eta, \theta) = (0, 0)$ followed by the 40/81 transition lottery reaches 57/200 with probability 2/9. From 57/200, event $(0, 1)$ reaches 3/10 with probability 1/10. From 3/10, event $(0, 1)$ followed by the 5/18 lottery reaches 2/5 with probability 1/36. The selected route therefore has probability

$$\frac{2}{9} \cdot \frac{1}{10} \cdot \frac{1}{36} = \frac{1}{1620} > 0.$$

(The total one-step transition probability from 57/200 to 3/10 is 11/108, because the $(0,0)$ cell supplies an additional 1/540 route.) Once the full-funding state is reached, the next period has $(d,\eta)=(1,0)$ with probability $m_0 = 11/20$. Thus the displayed route alone gives

$$\Pr(d_3 = 1, \eta_3 = 0) \geq \frac{1}{1620}\frac{11}{20} = \frac{11}{32400} > 0,$$

with dates indexed from the initial state. Hence override occurs on the truthful path with strictly positive probability.

Corollary 3 (On-path override)**.** *Every globally optimal admissible outcome at this calibration satisfies* $\Pr(d_t = 1, \eta_t = 0) > 0$ *for some date* t *under truthful play.*

Proof. Known-type compactness gives existence of a global optimum. If a global optimum were advice obedient, Lemma 14 would bound its value by 2/15. The verified seven-state mechanism is an admissible truthful outcome with value

$$\frac{162749}{1211600} = \frac{2}{15} + \frac{3607}{3634800} > \frac{2}{15},$$

a contradiction. The witness route above separately verifies that the comparison mechanism itself reaches override on path; the conclusion for every optimum follows from the strict value comparison, not from the witness mechanism's reachability alone. □

6 Private bias and finite-menu density

6.1 Primitive indexed menus and timing

The private type is drawn first and observed by the expert. The organization commits ex ante to the menu, or to a diagnostic-contingent family of menus. If a public diagnostic is present, its category is then drawn and observed by both parties; the corresponding committed menu becomes active before the expert chooses a branch. After the public branch index is recorded, the selected dynamic branch runs: in each period quality is drawn, the expert privately observes the signal, sends the branch message, public randomization and funding occur, quality is publicly revealed, and the branch continues. The initial branch choice and every within-branch action are part of one behavioral strategy.

The economic reduction is straightforward. A branch and strategy produce an exposure pair (A, B); type λ values it as $A - 2\lambda B$; and the menu takes the upper envelope over all such pairs. Geometry alone describes that envelope and its slopes. Implementation is a separate question: the maximizing pair must be generated by an identifiable branch and a complete sequentially optimal strategy. The definitions below impose exactly the common measurability needed to make that branch–strategy choice.

A primitive menu is an *indexed measurable menu*

$$\mathcal{M} = (J, (M_j)_{j \in J}).$$

Formally, J and the fixed date-by-date spaces are standard Borel, and branch kernels are jointly Borel in the index and public history; branch-specific alphabets are Borel subsets of common spaces with fixed default rules. Thus the menu is one extensive form rather than an unstructured collection of policies. Screening admissibility below separately requires a measurable maximizing branch choice and a complete parameterized best-response kernel.

Let $\Lambda = [\lambda_L, \lambda_H] \subseteq [0,1]$ and let $\mathcal{M} = (J, (M_j)_{j \in J})$ be a nonempty indexed standard-Borel menu. For a branch M_j, let $\Sigma(j)$ be its Borel behavioral strategies. For $\sigma \in \Sigma(j)$ write

$$z_j(\sigma) = (A_{M_j}(\sigma), B_{M_j}(\sigma)), \qquad \ell_\lambda(A, B) = A - 2\lambda B.$$

Define the raw exposure set and branch value by

$$Z(j) = \{z_j(\sigma) : \sigma \in \Sigma(j)\}, \qquad v_j(\lambda) = \sup_{z \in Z(j)} \ell_\lambda(z).$$

The supremum here is over strategies and need not be attained. Let $R_j(\lambda)$ be the set of exposure pairs generated by strategies that attain $v_j(\lambda)$ and are sequentially optimal after every finite private history, including histories of zero initial probability. Thus $R_j(\lambda)$ may be empty even though $v_j(\lambda)$ is well defined.

The menu envelope and its geometric exposure hull are

$$U_{\mathcal{M}}(\lambda) = \sup_{j \in J} v_j(\lambda), \qquad D(\mathcal{M}) = \operatorname{cl}\operatorname{co}\left(\bigcup_{j \in J} Z(j)\right).$$

Finally define the attained equilibrium exposure correspondence

$$\Gamma_{\mathcal{M}}(\lambda) = \bigcup_{j \in J:\ v_j(\lambda) = U_{\mathcal{M}}(\lambda)} \{j\} \times R_j(\lambda).$$

The menu is screening-admissible for F when there exist a Borel branch-choice map $j_F : \Lambda \to J$ and a complete behavioral strategy kernel, jointly Borel in type and private history, such that for F-almost every type they attain the menu envelope and are sequentially optimal after every finite private history. Equivalently, admissibility selects a point of $\Gamma_{\mathcal{M}}$ together with the complete strategy that generates it. A Borel exposure selection alone does not identify a branch index and is insufficient. The menu is branchwise sequentially regular on Λ when, for every $j \in J$, there exists a complete strategy kernel $\sigma_t^j(dm \mid \lambda, h_t^p)$, Borel jointly in (λ, h_t^p) at every date, that attains $v_j(\lambda)$ and is sequentially optimal after every finite private history for every type. No measurable space of complete strategies is invoked: regularity is stated directly as the existence of the sequence

of parameterized behavioral kernels. The universal quantifier over j is load-bearing because the grid approximation may select any near-envelope branch. No selector is required for the geometric support-function identity itself.

6.2 Sequential verification and measurable selection

For one fixed branch, let H_t^p be its standard-Borel private-history space before the date-t message. A complete parameterized behavioral strategy is a sequence of kernels

$$\sigma_t(dm \mid \lambda, h_t^p), \qquad t \geq 0,$$

Borel on $\Lambda \times H_t^p$. This object is stronger than a Borel exposure selector and avoids treating the generally unwieldy collection of all Borel kernels as a standard-Borel strategy space.

Lemma 15 (One-period deviations and a verifiable regular subclass)**.** *Fix a branch with bounded normalized discounted payoffs.*

(i) If a complete behavioral strategy weakly dominates every one-period deviation after every finite private history, then it is sequentially optimal after every such history, including histories of zero initial probability.

(ii) If the branch has a finite message alphabet at every date, then it is branchwise sequentially regular on every compact type interval. More generally, the same conclusion holds whenever the current-message maximization has a Borel argmax selector at every history and the associated Bellman operator preserves bounded Borel functions.

Proof. For (i), truncate any alternative strategy after T dates and use the candidate thereafter. Backward induction with the one-period inequalities gives dominance over every truncated deviation; bounded tails differ by at most a constant times δ^T, proving pointwise sequential optimality as $T \to \infty$.

For (ii), on Λ times the disjoint union of date-indexed histories, the Bellman operator maps bounded Borel functions into themselves: current-message payoffs are jointly Borel in type and history, and the finite pointwise maximum is Borel. It contracts with modulus δ, so its fixed point is Borel. The least indexed maximizing message is Borel jointly in type and history and gives a complete strategy kernel satisfying the one-period inequalities everywhere; part (i) proves sequential optimality and attainment. The same proof works whenever a Borel current argmax selector exists and the Bellman operator preserves bounded Borel functions. □

Standard-Borel kernels and pointwise attainment alone do not imply a parameterized Borel selector; Lemma 15(ii) gives the model-relevant sufficient subclass used below.

For a fixed branch, $v_j(\lambda)$ is nevertheless a finite convex Lipschitz function of λ, hence Borel, because it is a supremum of affine exposure payoffs. The paper does not claim that $(j, \lambda) \mapsto v_j(\lambda)$ or the full best-response graph is Borel merely from joint Borel branch kernels. Screening admissibility

supplies the needed menu-level Borel branch and strategy selection; branchwise regularity supplies selectors only after a finite set of branches has been retained.

A universally measurable selector would be enough to integrate behavior after completing a single fixed type distribution in some static problems. It is not the maintained object here: the completion can depend on the prior or posterior, does not give one Borel kernel for every null private history, and is not automatically stable under diagnostic garbling and finite tagging. The paper therefore retains Borel strategy kernels.

Theorem 5 (Exposure envelope)**.** *For every nonempty indexed menu $\mathcal{M}$, $D(\mathcal{M})$ is a nonempty compact convex subset of*

$$\mathcal{P} = \left\{ \left(m_0 x_0 + m_1 x_1, m_0(1-\alpha_0)x_0 + m_1(1-\alpha_1)x_1 \right) : (x_0, x_1) \in [0,1]^2 \right\}.$$

Moreover

$$U_{\mathcal{M}}(\lambda) = \max_{(A,B)\in D(\mathcal{M})} \ell_\lambda(A,B). \tag{E1}$$

The maximum in (E1) is a maximum on the geometric closed convex hull; it need not be generated by a branch or strategy in the original menu. The function $U_{\mathcal{M}}$ is convex, nonincreasing, absolutely continuous, and L-Lipschitz for $L = 2(1-q)$.

Let

$$\mathcal{F}_D(\lambda) = \arg \max_{(A,B)\in D(\mathcal{M})} \ell_\lambda(A,B)$$

be the geometric maximizing face. Then

$$\partial U_{\mathcal{M}}(\lambda) = \{-2B : (A,B) \in \mathcal{F}_D(\lambda)\} = [U'_-(\lambda), U'_+(\lambda)]. \tag{E2}$$

At every differentiability point, $\mathcal{F}_D(\lambda)$ contains one exposure pair, and the exposure component of every element of $\Gamma_{\mathcal{M}}(\lambda)$ equals

$$B(\lambda) = -\frac{1}{2}U'_{\mathcal{M}}(\lambda), \qquad A(\lambda) = U_{\mathcal{M}}(\lambda) - \lambda U'_{\mathcal{M}}(\lambda). \tag{E3}$$

At a kink, every attained equilibrium exposure belongs to the geometric face and therefore satisfies

$$-\frac{1}{2}U'_+(\lambda) \le B(\lambda) \le -\frac{1}{2}U'_-(\lambda), \tag{E4}$$

but the original menu need not implement every point of the face or either endpoint. Any selection from $\Gamma_{\mathcal{M}}$ is nonincreasing in exposure: if $\lambda < \mu$, then $B(\lambda) \ge B(\mu)$.

For a screening-admissible menu and atomless F,

$$\Pi_F(\mathcal{M}) = \int_\Lambda \left[U_{\mathcal{M}}(\lambda) + (1-\lambda)U'_{\mathcal{M}}(\lambda) \right] dF(\lambda), \tag{E5}$$

independently of equilibrium tie breaking.

Proof. For each date define

$$x_i = a \sum_{t \geq 0} \delta^t \mathbb{E}[d_t \mid \eta_t = i] \in [0, 1].$$

Current quality is independent of past public and private histories conditional on the current signal, and current funding is chosen before current quality is observed. Hence

$$A = m_0 x_0 + m_1 x_1, \qquad B = m_0(1 - \alpha_0)x_0 + m_1(1 - \alpha_1)x_1.$$

Thus every raw exposure pair lies in the compact convex parallelogram $\mathcal{P}$. The closed convex hull of a nonempty subset of a compact set in $\mathbb{R}^2$ is compact, proving the first claim.

A linear functional has the same supremum on a set, its convex hull, and its closure. Compactness of $D(\mathcal{M})$ turns that common supremum into the maximum in (E1). This step is purely geometric: a maximizer created by closure or convexification need not be an original branch–strategy outcome.

Every affine function $\ell_\lambda(A, B)$ has slope $-2B \in [-2(1 - q), 0]$. Their supremum is therefore convex, nonincreasing, and L-Lipschitz, hence absolutely continuous. The standard support-function subgradient formula gives (E2) (see Rockafellar 1970).[12] Because the maximizing face is convex, its B-projection is an interval, so the displayed set of slopes already equals the full subdifferential.

If U is differentiable, (E2) fixes B throughout the maximizing face, and the equality $A - 2\lambda B = U(\lambda)$ then fixes A. This proves (E3) for every attained equilibrium choice. At a kink, the exposure projection of $\Gamma_{\mathcal{M}}(\lambda)$ is contained in $\mathcal{F}_D(\lambda)$, which gives (E4), but no reverse inclusion follows from closed convexification.

For monotonicity, take attained maximizing pairs (A_λ, B_λ) and (A_μ, B_μ) with $\lambda < \mu$. The two optimality inequalities are

$$A_\lambda - 2\lambda B_\lambda \geq A_\mu - 2\lambda B_\mu, \qquad A_\mu - 2\mu B_\mu \geq A_\lambda - 2\mu B_\lambda.$$

Adding them gives $(\mu - \lambda)(B_\lambda - B_\mu) \geq 0$.

A finite convex function on an interval has at most countably many nondifferentiability points. Atomless F gives that set measure zero. At every remaining type, all attained equilibrium outcomes have the unique pair in (E3), and their principal payoff is

$$A - 2B = U + (1 - \lambda)U'.$$

Integrating proves (E5) and tie-breaking independence. □

Proposition 12 (Finite-submenu density)**.** *Let $D_\Lambda = \lambda_H - \lambda_L$. For every indexed menu $\mathcal{M}$, every $K \geq 1$, and every $\varepsilon > 0$, there is a literal submenu $\mathcal{M}_{K,\varepsilon} \subseteq \mathcal{M}$ with at most $K + 1$ branches such that*

$$0 \leq U_{\mathcal{M}}(\lambda) - U_{\mathcal{M}_{K,\varepsilon}}(\lambda) \leq \frac{L D_\Lambda}{K} + \varepsilon = \frac{2(1 - q)D_\Lambda}{K} + \varepsilon \tag{D1}$$

[12] Convex-potential and cyclic-monotonicity methods have a long mechanism-design pedigree; see Rochet (1987) and Krishna and Maenner (2001).

uniformly on Λ.

If the original menu is screening-admissible and branchwise sequentially regular, one may choose $\varepsilon_K \downarrow 0$ *so that each retained finite submenu is screening-admissible and*

$$\Pi_F(\mathcal{M}_{K,\varepsilon_K}) \longrightarrow \Pi_F(\mathcal{M}) \tag{D2}$$

for every atomless F*. Consequently, when both sides are restricted to the branchwise sequentially regular screening class, indexed-menu and finite-menu principal-value suprema coincide. The result is support-function and expected-value density; it is not allocation, payoff-pair, or mechanism density.*

Proof. Partition Λ at $t_j = \lambda_L + jD_\Lambda/K$. For each grid point choose an actual branch r_j with

$$v_{r_j}(t_j) \geq U(t_j) - \varepsilon.$$

Retain these branches and delete duplicates. If t_j is nearest to λ, the L-Lipschitz property of U and every v_r gives

$$U(\lambda) - U_K(\lambda) \leq |U(\lambda) - U(t_j)| + \varepsilon + |v_{r_j}(t_j) - v_{r_j}(\lambda)| \leq \frac{LD_\Lambda}{K} + \varepsilon.$$

The explicit ε is needed because the branch supremum need not be attained.

Under branchwise sequential regularity, every retained branch has a complete Borel best-response kernel for every type. Relabel the finite family and select the least maximizing label; its maximizing regions are Borel, so finite case distinctions give a screening-admissible finite menu with all copied off-path rules intact. Choose $\varepsilon_K \downarrow 0$. Uniform convergence $U_K \to U$ and convexity imply convergence of derivatives at every differentiability point of U where the U_K are differentiable. The union of all kink sets is countable, hence null under atomless F; bounded slopes and dominated convergence in formula (E5) yield $\Pi_F(\mathcal{M}_{K,\varepsilon_K}) \to \Pi_F(\mathcal{M})$. □

Proposition 13 (Low-frontier assigned outcomes cannot separate)**.** *Consider an incentive-compatible direct menu of conflicted types. If every assigned truthful branch–strategy outcome lies on the explicit complete-information low-frontier ray, then assigned funding is constant across all types assigned such outcomes.*

Proof. Every assigned pair on that ray satisfies

$$B(\lambda) = (1 - \alpha_1)A(\lambda), \qquad U_\lambda = A(\lambda)s_1(\lambda).$$

Suppose $A(\mu) > A(\lambda)$. Type λ can choose type μ's branch and imitate the complete reporting strategy assigned to μ. This is a feasible deviation whether or not it is a best response for type λ, and it generates the same exposure pair. Its payoff is $A(\mu)s_1(\lambda) > A(\lambda)s_1(\lambda)$ because conflictedness implies $s_1(\lambda) > 0$. Hence incentive compatibility rules out unequal assigned funding. □

The proposition concerns the assigned exposure pairs on the explicit ray. It does not claim that

every complete-information optimal branch has the same off-equilibrium strategy set, and it does not rule out screening by branches or assigned outcomes outside the ray.

6.3 Atomic distributions

Let $F = F^c + \sum_{\lambda\in\mathcal{A}} m_\lambda \delta_\lambda$, where F^c is atomless and $\mathcal{A}$ is at most countable. For a screening-admissible menu and a selected attained equilibrium pair at an atom, put

$$s_\gamma(\lambda) = -2B_\gamma(\lambda).$$

Then

$$s_\gamma(\lambda) \in S^{\mathrm{eq}}_{\mathcal{M}}(\lambda) := \{-2B : (A,B) \in \Gamma_{\mathcal{M}}(\lambda)\} \subseteq [U'_-(\lambda), U'_+(\lambda)]. \tag{A1}$$

The selected expected principal payoff is

$$\Pi_F(\gamma) = \int_\Lambda [U + (1-\lambda)U']\, dF^c + \sum_{\lambda\in\mathcal{A}} m_\lambda \{U(\lambda) + (1-\lambda)s_\gamma(\lambda)\}. \tag{A2}$$

Thus the geometric one-sided derivatives give the bounds

$$U(\lambda) + (1-\lambda)U'_-(\lambda) \le P_\gamma(\lambda) \le U(\lambda) + (1-\lambda)U'_+(\lambda). \tag{A3}$$

They need not be attainable. If

$$\underline{s}^{\mathrm{eq}}(\lambda) = \inf S^{\mathrm{eq}}_{\mathcal{M}}(\lambda), \qquad \overline{s}^{\mathrm{eq}}(\lambda) = \sup S^{\mathrm{eq}}_{\mathcal{M}}(\lambda),$$

then the supremum-minus-infimum of expected principal payoffs over attained equilibrium selections is bounded by

$$\sum_{\lambda\in\mathcal{A}} m_\lambda(1-\lambda)\{\overline{s}^{\mathrm{eq}}(\lambda) - \underline{s}^{\mathrm{eq}}(\lambda)\} \le \sum_{\lambda\in\mathcal{A}} m_\lambda(1-\lambda)\{U'_+(\lambda) - U'_-(\lambda)\}. \tag{A4}$$

The first expression is an equality only when the relevant selection extrema exist or are interpreted as suprema and infima; the second is an equality only when the geometric endpoint slopes are implemented at every atom. Closed convexification alone does not provide those branches.

The bounds can be strict: distinct attained branches may tie in expert utility while generating different exposure, and a geometric endpoint slope may belong only to the closure of the exposure set. Thus atoms require the attained-equilibrium slope set in (A1), not merely one-sided derivatives.

7 Diagnostics and finite-menu approximation

7.1 Primitive diagnostic structure and categorywise assembly

Let the privately known type have prior F on Λ. A finite public diagnostic Z is generated before menu choice by a kernel $P_Z(z \mid \lambda)$. It is observed by both parties and carries no payoff-relevant information conditional on λ: future projects, signals, and payoffs are distributed as in the baseline model once the type is fixed. Write

$$\pi_z = \int_\Lambda P_Z(z \mid \lambda)\, dF(\lambda), \qquad F_z(E) = \frac{\int_E P_Z(z \mid \lambda)\, dF(\lambda)}{\pi_z}$$

when $\pi_z > 0$. Zero-probability categories are assigned a fixed default indexed menu and complete strategy. They never enter value or almost-sure screening constraints, but specifying them keeps the extensive form pointwise complete.

A diagnostic Z' Blackwell-dominates Z when there is a stochastic matrix $Q(z \mid z')$, independent of λ, satisfying

$$P_Z(z \mid \lambda) = \sum_{z'} Q(z \mid z') P_{Z'}(z' \mid \lambda) \quad \text{for every } \lambda. \tag{G1}$$

The organization can then generate the coarse category publicly from the fine one. Although the expert observes z', she already knows λ, and z' contains no additional payoff-relevant state conditional on that type. Thus the continuation problem under a coarse branch is the same for every fine category that is publicly garbled to it.

After category z, the organization activates an indexed branchwise sequentially regular screening-admissible menu $\mathcal{M}_z = (J_z, (M_{z,j})_{j \in J_z})$. Category-specific branches and complete sequential-best-response selections need not be common across categories. Because the diagnostic support is finite, the global index space

$$J^Z = \bigsqcup_{z \in \operatorname{supp} Z} \{z\} \times J_z$$

is standard Borel, and the global branch family is obtained by finite Borel gluing of the category families. Common date-by-date spaces can be finite tagged disjoint unions or common universal spaces. The selected tag (z, j) invokes that branch's own continuation rule at every reached or null history. A zero-probability category is assigned one fixed default index and complete strategy. Thus the diagnostic-contingent construction is one admissible standard-Borel mechanism, not merely a collection of posterior problems.

Let $V(Z)$ be the supremum of expected principal payoff over such diagnostic-contingent menus. Let $V_K(Z)$ impose at most $K+1$ branches in every positive-probability category. For a vector $\mathbf{K} = (K_z)_{z:\pi_z>0}$, let $V_{\mathbf{K}}(Z)$ impose at most $K_z + 1$ branches after category z.

Theorem 6 (Diagnostics and categorical authority)**.** *(i) If Z' Blackwell-dominates Z, then attainable value under Z' is weakly larger.*

(ii) If every positive-probability posterior is atomless, then $V_K(Z) \uparrow V(Z)$.

(iii) For each positive-probability category z, let I_z be the closed convex hull of $\operatorname{supp} F_z$ *and let $D_z > 0$ be its length. Suppose F_z is absolutely continuous with respect to Lebesgue measure on I_z, with density bounded by $\bar{f}_z$, and let L_z be a common Lipschitz bound for the posterior indirect-utility envelopes under consideration. Then*

$$0 \le V(Z) - V_{\mathbf{K}}(Z) \le \sum_{z:\pi_z>0} \pi_z \left(\frac{L_z D_z}{K_z} + \frac{8\sqrt{2}\,\bar{f}_z L_z D_z}{\sqrt{K_z}} \right). \tag{G2}$$

The right side is an infimum bound when a posterior unrestricted optimum is not attained. Singular atomless posteriors satisfy part (ii), but the density argument below does not provide the rate in part (iii).

Proof. Part (i): public garbling and null sets. Fix a feasible coarse diagnostic-contingent menu and a selected equilibrium outcome. Under Z', publicly draw z from $Q(\cdot \mid z')$ and then use the old menu and selected outcome for category z. Pointwise in type, the old branch-choice and within-branch deviation inequalities are unchanged: conditional on λ, the fine category affects neither the branch's rules nor any future payoff-relevant uncertainty. The copied complete strategy is therefore sequentially optimal after every private history, including histories having zero probability under the initial law. If the coarse equilibrium uses the Borel selector $j_z(\lambda)$, the fine mechanism uses the Borel tagged selector $(z', z, j_z(\lambda))$. Branchwise sequential regularity is preserved by finite tagging of the old complete outcome.

For completeness, let N_z be a type set on which the old category-z menu lacks the selected equilibrium outcome, and suppose $F_z(N_z) = 0$. Bayes' rule and (G1) give, for $\pi_z > 0$,

$$F_z = \sum_{z'} \omega(z' \mid z) F_{z'}, \qquad \omega(z' \mid z) = \frac{\pi_{z'} Q(z \mid z')}{\pi_z}. \tag{G3}$$

Because every summand is nonnegative, $F_z(N_z) = 0$ implies $F_{z'}(N_z) = 0$ for each z' with $\omega(z' \mid z) > 0$. Fine categories or garbling cells with zero joint probability impose no restriction and receive arbitrary specified rules. Summing over the finitely many coarse categories proves that the copied selection is admissible almost surely under every relevant fine posterior. This establishes the value inclusion.

Part (ii): posteriorwise approximation and one common cap. Fix $\eta > 0$. For each positive-probability category choose a branchwise sequentially regular screening menu whose posterior payoff is within $\eta/(2|Z_+|\pi_z)$ of the posterior supremum, where $Z_+ = \{z : \pi_z > 0\}$. Proposition 12 then gives, for that chosen menu, a finite literal submenu whose posterior payoff loss is below $\eta/(2|Z_+|\pi_z)$. Let $K_z + 1$ be its branch count and put $K = \max_z K_z$. The category menus need not share any branch or null-history strategy; each simply uses no more than the common cap $K + 1$. Their finite tagged disjoint union is one standard-Borel menu and loses at most η in total. Since the feasible classes are nested in K, $V_K(Z)$ increases to $V(Z)$. This order of choice handles nonattained posterior suprema: first choose near-optimal unrestricted menus, then finite submenus, then the maximum cap.

Part (iii): quantitative rate. Fix a positive-probability category and suppress its subscript. Choose a branchwise sequentially regular screening menu within posterior payoff tolerance $\tau > 0$ of the posterior supremum. Let f be its indirect-utility envelope on I, and let g be the envelope of the $K+1$ branch submenu supplied by Proposition 12. For every $\varepsilon > 0$,

$$\|f - g\|_\infty \le \Delta := \frac{LD}{K} + \varepsilon. \tag{G4}$$

The following lemma controls the exposure term.

Lemma 16 (Derivative stability for convex envelopes)**.** *Let f and g be convex nonincreasing functions on an interval of length $D > 0$, with almost-everywhere derivatives in $[-L, 0]$. If $\|f - g\|_\infty \le \Delta$, then*

$$\int |f' - g'|\, d\lambda \le 8\sqrt{2}\,\sqrt{L\Delta D}. \tag{G5}$$

If $L = 0$ or $\Delta = 0$, the left side is zero.

Proof. Translate the interval to $[0, D]$ and suppose $L\Delta > 0$. Put $p = f'$ and $q = g'$ at their almost-everywhere points of definition. Both are nondecreasing and take values in $[-L, 0]$. For $0 < r \le D/2$ and almost every $x \in [r, D - r]$, convexity and the uniform bound imply

$$(p(x) - q(x))_+ \le \frac{g(x + r) - 2g(x) + g(x - r) + 2\Delta}{r}.$$

Integrating the second difference leaves only two boundary increments, each bounded in absolute value by Lr^2; hence

$$\int_r^{D-r} (p - q)_+ \le 2Lr + \frac{2\Delta D}{r}.$$

Interchanging f and g gives the same bound for $(q - p)_+$. The two boundary strips have total length $2r$ and $|p - q| \le L$, so

$$\int_0^D |p - q| \le 6Lr + \frac{4\Delta D}{r}. \tag{G6}$$

If $r_* = \sqrt{2\Delta D/(3L)} \le D/2$, substitution in (G6) gives $4\sqrt{6}\sqrt{L\Delta D} < 8\sqrt{2}\sqrt{L\Delta D}$. If $r_* > D/2$, then $\Delta > 3LD/8$ and the trivial bound $\int |p - q| \le LD$ is also below the right side of (G5). This proves the claim, including endpoint effects. □

Because F has Lebesgue density at most $\bar{f}$, the atomless envelope formula and $0 \le 1 - \lambda \le 1$ give

$$|\Pi_F(f) - \Pi_F(g)| \le \Delta + \bar{f} \int_I |f' - g'|\, d\lambda \le \Delta + 8\sqrt{2}\,\bar{f}\sqrt{L\Delta D}. \tag{G7}$$

This is the only step using a bounded density. A singular atomless posterior still ignores the countable kink sets and hence gives qualitative convergence by dominated convergence, but Lebesgue L^1 control of derivatives cannot be converted into a rate for that posterior without an alternative domination assumption.

Let first $\varepsilon \downarrow 0$ in (G4)–(G7), then let the preliminary posterior tolerance $\tau \downarrow 0$. Since

$$\sqrt{L\left(\frac{LD}{K}\right)D} = \frac{LD}{\sqrt{K}},$$

the category loss is at most

$$\frac{LD}{K} + \frac{8\sqrt{2}\,\bar{f}LD}{\sqrt{K}}.$$

Summing with weights π_z proves (G2). A positive-probability singleton posterior is not absolutely continuous on an interval of positive length and is therefore outside part (iii); it is a known-type category rather than a bounded-density private-type category. □

Interpretation of the rate. Part (iii) is a quantitative refinement of the value-density result, not a characterization of the optimal allocation of a finite menu budget across diagnostic categories. Its role is only to show that, under bounded posterior densities and Lipschitz envelopes, categorywise approximation is of order $K_z^{-1/2}$. The qualitative convergence in part (ii) requires neither these stronger regularity conditions nor the explicit constant in (G2).