Keeping Experts Honest*

Ricardo Fonseca†

This version: April 2026
First version: October 2023

Abstract

Decision-makers often rely on advice from specialists who possess superior information but whose preferences are not perfectly aligned with those of the decision-maker. A canonical example is the funding of scientific research by agencies such as the NIH or NSF, which depend on peer reviewers who may be biased toward particular methodologies or research programs. This paper studies how a principal who lacks monetary transfers can discipline a biased expert through commitment to future decision rules. In an infinite-horizon, discrete-time model with noisy binary signals and a known state-independent bias, I show that a simple randomized grim-trigger mechanism is optimal: the principal follows the expert's recommendation whenever possible but, with a carefully chosen probability, commits to permanently ignoring the expert after any recommendation that proves incorrect. I characterize the optimal punishment probability as a function of the signal precision, prior, and bias, and establish optimality within the full class of history-dependent mechanisms under a natural patience condition.

Keywords: repeated games, mechanism design without transfers, expert advice, cheap talk, grim-trigger strategies, peer review.

JEL Codes: C73, D82, D83.

*I am grateful to Roberto Serrano's guidance and comments from Jack Fanning, Teddy Mekonnen, Bobby Pakzad-Hurson, Kareen Rozen, and Theory Lunch at Brown University participants. All remaining errors are my own.
†Pontificia Universidad Javeriana, Economics Department. E-mail: baricardo@javeriana.edu.co.

1 Introduction

Decision-makers routinely rely on the opinions of experts whose preferences differ from their own. A leading example is the allocation of scientific funding: agencies such as the National Institutes of Health (NIH) and the National Science Foundation (NSF) use peer review to evaluate proposals, yet reviewers may favor particular methodologies, research programs, or sub-communities, and therefore recommend funding projects that a disinterested decision-maker would reject.[1] The decision-maker cannot make contingent monetary transfers to reviewers, but he can commit to a policy mapping past recommendations and realized outcomes into future decisions. How should he design such a policy?

This paper studies a discrete-time, infinite-horizon principal–agent model in which a single long-lived expert observes a noisy binary signal about the quality of a project that arrives each period. The expert's preferences place weight $\lambda \in [0, 1]$ on the principal's payoff and weight $1 - \lambda$ on an action-contingent component that rewards the expert whenever a project is funded, regardless of quality. The parameter λ captures the degree of preference alignment and is common knowledge. The principal commits at $t = 0$ to a mechanism that maps the public history of recommendations, realized qualities, and randomization draws into current and future funding decisions.

The main result is that a simple randomized grim-trigger mechanism, which I call the *Almost-Unforgiving Information Transmission* mechanism (AUNT), is optimal whenever the principal is sufficiently patient. Under AUNT, the principal follows the expert's recommendation whenever he has not yet committed to ignore her. Following any funded project that turns out to be bad, the principal commits–with a punishment probability $\varepsilon(\lambda)$ chosen to just render the expert indifferent between reporting truthfully and lying when her signal is negative–to never fund any project in the future. This commitment is absorbing: once triggered, all subsequent payoffs for both players are zero. The result refines classical folk-theorem arguments for dynamic cheap talk by providing an exact, simple optimal mechanism.

The mechanism has several attractive features. It is Markovian: the same rule applies in every period. It uses only observed outcomes of funded projects, so the principal need not be able to verify the quality of projects that were rejected. Its punishment is extreme but rare: the expected duration until triggering is long when the expert is honest, because an honest expert is only punished when an informative but unlucky signal leads to a genuine misrecommendation. Finally, the structure extends in a transparent way: if the bias is toward rejecting projects, the mirror-image mechanism that commits to funding all projects after a wrong negative recommendation is optimal by the same logic.

The intuition behind optimality has two parts. First, a version of the revelation principle applies: it is without loss of generality to focus on mechanisms in which the expert reports truthfully

[1] Li (2017) documents that NIH reviewers are simultaneously more informed and more biased about proposals in their own areas of expertise; on net, expertise dominates, but bias is quantitatively significant. Related empirical work on the consequences of peer review design includes Azoulay et al. (2011), Boudreau et al. (2016), Pier et al. (2018), and Fang et al. (2016).

in equilibrium. Second, among truthful mechanisms, the principal's payoff is the probability that she follows the expert's recommendation, weighted by the informational content of that recommendation. Maximizing this probability subject to truth-telling constraints naturally pushes the principal toward mechanisms that reward accurate recommendations and punish inaccurate ones maximally. Because the expert receives a state-independent benefit from funding, the incentive problem is one-sided: only the signal realization $\eta_t = 0$ (indicating a bad project) tempts the expert to lie. The optimal punishment is therefore tailored to this single constraint, and the bang-bang logic familiar from Abreu et al. (1990) delivers the grim-trigger form: randomize between the harshest available continuation (never listen again) and the softest (continue following the expert), with the mixing probability set to make the relevant truth-telling constraint bind.

The paper proceeds as follows. Section 2 positions the contribution within the literature on repeated cheap talk, mechanism design without transfers, and review strategies. Section 3 introduces the model. Section 4 analyzes the first-best benchmark and establishes an impossibility result for imperfect precision. Section 5 defines the AUNT mechanism, computes its optimal punishment probability, and proves optimality among all feasible mechanisms. Section 6 concludes. An appendix contains the result on near-first-best payoffs under almost-perfect precision and defers proofs omitted from the main text.

2 Related Literature

The paper contributes to three connected literatures: (i) repeated cheap talk and expert advice, (ii) dynamic mechanism design without transfers, and (iii) review strategies in repeated games with imperfect monitoring. It also speaks to an applied literature on the economics of scientific grant allocation.

Repeated cheap talk and expert advice. The starting point is the canonical cheap talk framework of Crawford and Sobel (1982), in which a biased sender transmits information only through coarse partitional equilibria. Sobel (1985) extended this setting dynamically, showing how an expert can build credibility and then strategically exploit it; the AUNT mechanism can be read as the principal's optimal response to exactly this dynamic exploitation pattern, in an environment where the principal has commitment power. Bénabou and Laroque (1992) introduced noisy private information into a reputational framework, establishing that manipulation can persist because signal imperfection masks dishonest reports—the same imperfection that motivates the present paper's probabilistic punishment.

Subsequent work in this tradition includes Ottaviani and Sørensen (2006a,b), who characterize the severity of information loss when experts have reputational concerns, and Morgan and Stocken (2003), who study analyst recommendations under uncertain analyst incentives. Morris (2001) identifies a countervailing force: even unbiased experts may distort reports to avoid appearing biased, a "political correctness" effect that any well-designed mechanism must accommodate. Krishna and Morgan (2001, 2004) show that multi-expert or multi-stage communication can improve upon Crawford-Sobel outcomes; the present paper pursues an orthogonal route, extracting information from a single expert across time rather than within periods.

Among recent contributions, Lipnowski and Ravid (2020) characterize sender-preferred equilibria with state-independent ("transparent") motives, providing the one-shot benchmark that the dynamic mechanism here improves upon. Kuvalekar et al. (2022) study repeated cheap talk when no outcome feedback is available and show that relational incentives cannot restore commitment in that environment; their result underscores the importance of outcome observability in the present setting, where the principal eventually observes realized project qualities. Best and Quigley (2024) analyze persuasion when credibility depends entirely on a record of accuracy, without commitment; their finding that on-path punishment is required when feedback is full parallels the grim-trigger structure of AUNT, though in a conceptually complementary setting where the sender, rather than the receiver, is the strategic designer. Mathevet et al. (2024) show that repeated play can endogenously approximate the Bayesian persuasion outcome as the sender becomes patient, providing a reputational foundation for commitment; the present paper takes the reverse perspective, asking what a receiver with commitment can achieve. Closest in environment, Kivinen and Kuzmics (2025) study the same repeated cheap-talk setting with i.i.d. decision problems but ask what grim-trigger-like strategies can sustain when players may renegotiate their continuation play; they show that no weakly renegotiation-proof equilibrium attains payoffs close to the receiver-optimal subgame-perfect payoff. Their negative result and the positive characterization here are complementary: under commitment (this paper), a grim-trigger-type mechanism attains the best receiver payoff; absent commitment and requiring renegotiation-proofness (Kivinen and Kuzmics, 2025), the same form of punishment is not credible and the receiver-optimal payoff is unachievable.

Mechanism design without transfers. The paper's closest methodological relatives study dynamic mechanisms without monetary transfers. Deb et al. (2018) consider a principal evaluating a forecaster of unknown precision, showing that an optimal deterministic mechanism evaluates a single, optimally-timed prediction. The present paper differs in two substantive dimensions: bias rather than precision is the informational friction (the expert's type is known), and the mechanism disciplines a biased known-precision expert rather than screening for quality. Guo and Hörner (2020) study dynamic allocation of a perishable good when the agent's value follows a Markov chain, finding that optimal mechanisms backload inefficiency through "quantified entitlement" that eventually tenures or terminates the agent. The structural parallel to AUNT is close: future decisions substitute for transfers as incentive tools, and optimal punishment is absorbing. Lipnowski and Ramos (2020) study repeated delegation without principal commitment and characterize optimal equilibria in which the agent progressively loses autonomy; here, the principal's commitment power and the expert's noisy signal yield a qualitatively different optimal structure.

Escobar and Toikka (2013) prove folk-theorem-style efficiency results for repeated Bayesian games with Markov private types, providing the asymptotic benchmark that the present paper's exact characterization refines. Margaria and Smolin (2018) establish a folk theorem for dynamic cheap talk with state-independent sender payoffs: for patient players, any feasible individually-rational payoff is achievable. The present paper sharpens this existence result to an exact optimal-mechanism characterization; moreover, my model features a state that is "fresh" each period (i.i.d. qualities with noisy signals), which is not covered by their folk theorem.[2] Che and Kartik (2009) show that centralization of authority is essential when advisor and principal disagree about the right action; the commitment assumption in the present paper plays the analogous role.

Krähmer (2021) studies a static environment in which a receiver with secret information-design power achieves near-complete-information payoffs even against strongly biased senders. The present paper's dynamic commitment plays an analogous role: the principal's ability to condition future decisions on observed outcomes substitutes for the ability to reshape information. Kamenica and Lin (2024) analyze how randomization interacts with commitment in communication games; AUNT's randomization over permanent exclusion is a specific dynamic instance of the general force they study. Kolotilin and Li (2021) characterize optimal relational communication with voluntary transfers; the no-transfer assumption here is the key constraint that pushes the optimum toward exclusion rather than side payments.

Review strategies and APS machinery. The formal machinery of the proof follows Abreu et al. (1990): self-generation, recursive characterization of equilibrium payoff sets, and the bang-bang property of optimal continuation values in repeated games with imperfect public monitoring. The AUNT mechanism is a review strategy in the tradition of Radner (1985), with a specific randomized punishment trigger that parallels Green and Porter (1984)'s use of trigger strategies in oligopoly. The optimality of extreme punishments is familiar from Abreu et al. (1986), who showed that in the symmetric stochastic-oligopoly problem optimal equilibria use maximally severe continuation values on the punishment side.

Fuchs (2007) studies repeated moral hazard with private principal evaluations and shows that simple "fire" mechanisms dominate more elaborate structures; AUNT exhibits the same "simple-is-optimal" phenomenon in an information-elicitation rather than effort-elicitation context. Recent work by Sugaya and Wolitzky (2023, 2025) characterizes the interaction between discounting and monitoring precision in repeated games, contextualizing the comparative statics of AUNT in δ and p. Bird and Frug (2019) find that optimal dynamic non-monetary incentives often take the form of "carte blanche" over a predetermined horizon; this complements the present paper's finding of extreme-but-rare absorbing punishment.

Reputation and endogenous credibility. An alternative approach to disciplining experts relies on reputation rather than explicit commitment. Ely and Välimäki (2003); Ely et al. (2008) show that reputational concerns can destroy surplus when short-run players may interpret friendly actions as evidence of a bad type; Cripps et al. (2004) prove that reputation effects are temporary

[2] A persistent state environment (as in Renault et al., 2013, 2017) would change the analysis materially, as learning about the current state and about past deviations would confound. I return to this point in the conclusion.

under imperfect monitoring. These impermanence and perverse-reputation results motivate the commitment-based approach taken here: the principal cannot rely on endogenous reputation alone to discipline the expert. Pei (2020) studies reputation with interdependent values, a feature present in this model since the expert's payoff depends on the realized project quality. Fudenberg et al. (2022) provide a reputational foundation for commitment payoffs when a long-run player can send messages before acting; the present paper's mechanism achieves similar discipline through designed rather than endogenous means.

Science funding. The motivating application connects to a small economic theory literature on grant allocation, including Gross and Bergstrom (2019, 2021), who model researcher-side strategic responses to peer review, and Carnehl et al. (2025), who provide a comprehensive overview of mechanism design for scientific grants. On the empirical side, Li (2017) provides the most direct evidence for the expertise-versus-bias trade-off modeled here; Azoulay et al. (2011) and Boudreau et al. (2016) document how review structures shape the research portfolios funded.

3 Model

3.1 Environment

Time is discrete and the horizon is infinite: $t = 0, 1, 2, \ldots$. There is one long-lived *principal* and one long-lived *expert* (equivalently, *agent* or *advisor*). Both discount the future at the common rate $\delta \in (0, 1)$.

Each period t, a project of quality $q_t \in \{0, 1\}$ arrives. Qualities are independent and identically distributed across periods with $\mathbb{P}(q_t = 1) = q \in (0, 1)$. Before the principal decides whether to fund the project, the expert privately observes a signal $\eta_t \in \{0, 1\}$ satisfying

$$\mathbb{P}(\eta_t = q_t \mid q_t) = p$$

for a precision parameter $p \in (1/2, 1]$. The signal is conditionally independent across periods given the sequence of qualities. The precision p is common knowledge and does not depend on the expert's effort.

After observing η_t, the expert submits a recommendation $r_t \in \{0, 1\}$, where $r_t = 1$ is interpreted as a recommendation to fund. A public randomization device x_t drawn uniformly from $[0, 1]$ realizes simultaneously. The principal then makes a funding decision $d_t \in \{0, 1\}$. At the end of the period the quality q_t is publicly revealed.[3]

The *public history* at the start of period t is

$$h^t \equiv \big((r_0, q_0, x_0), (r_1, q_1, x_1), \ldots, (r_{t-1}, q_{t-1}, x_{t-1})\big),$$

[3]For the optimal mechanism below, only the quality of *funded* projects needs to be observed. See Remark 4.

with $h^0 = \varnothing$. Let $\mathcal{H}$ denote the set of all finite public histories. A *mechanism* is a collection of functions

$$d = \{\rho^r(h^t)\}_{h^t \in \mathcal{H},\ r \in \{0,1\},\ t \geq 0},$$

where $\rho^r(h^t) \in [0,1]$ is the probability that the principal funds the project at period t after history h^t and recommendation r. Implicitly, the public randomization draw x_t is used to implement any required randomization over d_t. The principal commits to d at $t = 0$.

An *expert strategy* is a collection of functions $\sigma_t(h^t, \eta_t; d) \in [0,1]$ giving the probability of reporting $r_t = 1$ conditional on the public history h^t, the signal realization η_t, and the mechanism d. Because signals are i.i.d. and the mechanism is public, it is without loss of generality to condition only on the public history (and the current signal) rather than on the expert's private history of past signals.

3.2 Payoffs

The principal's period payoff is

$$u^P(d_t, q_t) = \begin{cases} 1 & \text{if } d_t = 1 \text{ and } q_t = 1, \\ -1 & \text{if } d_t = 1 \text{ and } q_t = 0, \\ 0 & \text{if } d_t = 0, \end{cases}$$

and her objective is to maximize

$$U^P(d) = (1-\delta)\,\mathbb{E}\left[\sum_{t=0}^{\infty} \delta^t\, u^P(d_t, q_t)\right].$$

The expert's period payoff is

$$u(d_t, q_t) = \lambda\, u^P(d_t, q_t) + (1-\lambda)\,\mathbf{1}\{d_t = 1\},$$

where $\lambda \in [0,1]$ is the *alignment parameter*, common knowledge. When $\lambda = 1$ the expert's preferences coincide with the principal's; when $\lambda = 0$ the expert values only whether projects are funded, regardless of quality. The expert's continuation payoff at history h^t with current signal η_t under mechanism d and strategy σ is

$$U_t(h^t, \eta_t; d, \sigma) = (1-\delta)\,\mathbb{E}\left[\sum_{s \geq t} \delta^{s-t}\, u(d_s, q_s)\,\middle|\, h^t, \eta_t\right].$$

Bias is one-sided: the expert is weakly biased toward funding. A symmetric treatment of the opposite bias is immediate and formalized in Proposition 1.

I adopt the tie-breaking convention that an indifferent expert reports truthfully: if U_t from $r_t = \eta_t$ equals U_t from $r_t \neq \eta_t$, the expert chooses $r_t = \eta_t$. This convention is inessential for the results.

3.3 Solution Concept and Maintained Assumption

The solution concept is *perfect Bayesian equilibrium* (PBE): the principal chooses a mechanism d maximizing $U^P(d)$ anticipating the expert's optimal response, and at every history and signal realization the expert maximizes her continuation payoff given the mechanism.

Throughout the analysis I assume that signal precision and prior are such that the expert's information is genuinely useful for the principal's decision:

Assumption 1 (Informative signals)**.** $1 - p < q < p$.

If $q \leq 1-p$ the principal would optimally reject every project regardless of signal; if $q \geq p$ he would accept every project. In both edge cases the signal is payoff-irrelevant and mechanism design is trivial.

For convenience, define the signal's *accuracy*:

$$\kappa \equiv \mathbb{P}(\eta_t = q_t) = pq + (1-p)(1-q),$$

and note that $\kappa \in (1/2, 1)$ under Assumption 1.

4 First-Best and Impossibility

I begin by establishing the principal's first-best benchmark and a sharp impossibility result for imperfect precision.

Definition 1 (First-best)**.** A mechanism d achieves the principal's *first-best* if $U^P(d)$ equals the principal's expected payoff from a counterfactual in which he directly observes the signal realizations $(\eta_t)_{t\geq 0}$.

Under Assumption 1, the first-best policy funds the project if and only if $\eta_t = 1$, yielding per-period expected payoff

$$U^{FB} = \mathbb{P}(\eta_t = 1, q_t = 1) - \mathbb{P}(\eta_t = 1, q_t = 0) = pq - (1-p)(1-q) = p + q - 1 > 0.$$

Consider the simple mechanism that implements whatever the expert recommends:

Definition 2 (Rubber-stamping)**.** The *rubber-stamping mechanism* d^{RS} is defined by $\rho^r(h^t) = r$ for all $r \in \{0, 1\}$ and all $h^t \in \mathcal{H}$.

Under d^{RS}, the expert's signal $\eta_t = 1$ always induces truthful reporting, since reporting $r_t = 1$ weakly dominates $r_t = 0$ stage-by-stage (the continuation value is unchanged by the report). After $\eta_t = 0$, the expert's expected stage payoff from lying ($r_t = 1$) is

$$\frac{\mathbb{P}(\eta_t = 0, q_t = 1)}{\mathbb{P}(\eta_t = 0)} \cdot 1 + \frac{\mathbb{P}(\eta_t = 0, q_t = 0)}{\mathbb{P}(\eta_t = 0)} \cdot (1 - 2\lambda) = \frac{(1-p)q + (1-2\lambda)p(1-q)}{1-\kappa},$$

which exceeds the zero stage payoff from truthful reporting whenever

$$\lambda < \bar{\lambda}(p,q) \equiv \frac{(1-p)q + p(1-q)}{2p(1-q)} = \frac{1-\kappa}{2p(1-q)}.$$

Thus rubber-stamping achieves the first-best if and only if $\lambda \geq \bar{\lambda}(p,q)$: when alignment is high enough, the expert's stage preferences already align with truthful reporting. The following theorem shows that if rubber-stamping fails, no mechanism can attain the first-best unless the signal is perfect.

Theorem 1 (Impossibility of first-best)**.** *Suppose $p < 1$ and $\lambda < \bar{\lambda}(p,q)$. Then no mechanism achieves the principal's first-best.*

Proof. The first-best policy funds if and only if $\eta_t = 1$, which is observationally indistinguishable from funding if and only if $r_t = 1$ when $r_t = \eta_t$ is played. Any mechanism that attains the first-best therefore induces (with probability one on the equilibrium path) funding decisions satisfying $d_t = \eta_t$. Under truthful reporting this requires $d_t = r_t$, i.e. rubber-stamping on the equilibrium path. But under rubber-stamping with $\lambda < \bar{\lambda}(p,q)$, truthful reporting fails after $\eta_t = 0$ as argued above, a contradiction. Any non-rubber-stamping implementation either fails to follow some $\eta_t = 1$ signal (losing an expected benefit of $\mathbb{P}(\eta_t = 1, q_t = 1) - \mathbb{P}(\eta_t = 1, q_t = 0) > 0$) or funds some $\eta_t = 0$ signal (losing an expected $\mathbb{P}(\eta_t = 0, q_t = 0) - \mathbb{P}(\eta_t = 0, q_t = 1) > 0$), and therefore also fails to attain the first-best. □

Remark 1 (Perfect precision)**.** When $p = 1$, the *Unforgiving Information Transmission mechanism* d^{UNIT}, defined below by setting $\varepsilon = 1$ in the AUNT mechanism, achieves the first-best for any λ provided $\delta \geq 1/(1+q)$. Appendix A extends this observation to a continuity result at $p = 1$.

5 The AUNT Mechanism and Its Optimality

In light of Theorem 1, the focus is on second-best mechanisms when $\lambda < \bar{\lambda}(p,q)$ and $p < 1$.

5.1 Definition and Value

Definition 3 (AUNT mechanism)**.** Fix $\varepsilon \in [0,1]$. The *Almost-Unforgiving Information Transmission mechanism with parameter* ε, denoted $d^{\text{AUNT}}(\varepsilon)$, is defined as follows. Starting from a state labeled "active," at each period:

(i) If the current state is *active*, the principal sets $d_t = r_t$ (rubber-stamping). If $r_t = 1$ and $q_t = 0$, the state transitions to *absorbing* with probability ε, using the public randomization device x_t; otherwise the state remains active.

(ii) If the current state is *absorbing*, $d_t = 0$ for all future periods.

Two special cases are instructive. With $\varepsilon = 0$, AUNT reduces to rubber-stamping. With $\varepsilon = 1$, AUNT reduces to UNIT: the principal permanently ignores the expert after any incorrect positive recommendation.

Under $d^{\text{AUNT}}(\varepsilon)$, truthful reporting generates the following continuation value for the expert in the active state. Let σ_R^* denote this value. By stationarity (the mechanism is time-invariant within the active state and qualities are i.i.d.),

$$\sigma_R^* = (1-\delta)\big[pq + (1-2\lambda)(1-p)(1-q)\big] + \delta\big[1 - \varepsilon(1-p)(1-q)\big]\sigma_R^*,$$

since under truthful play the stage payoff comes only from funded projects (which occur when $\eta_t = 1$, yielding the bracket), and punishment occurs exactly when $\eta_t = 1$ while $q_t = 0$, which happens with probability $(1-p)(1-q)$. Solving,

$$\sigma_R^* = \frac{(1-\delta)\big[pq + (1-2\lambda)(1-p)(1-q)\big]}{(1-\delta) + \delta\varepsilon(1-p)(1-q)}. \tag{1}$$

5.2 The Optimal Punishment Probability

The principal chooses ε to make the expert willing to report truthfully at the binding signal realization. When $\lambda < \bar{\lambda}(p,q)$, the expert is tempted to report $r_t = 1$ after $\eta_t = 0$; the constraint $r_t = \eta_t = 1$ is slack (see Lemma 2 below). Comparing the expected payoffs from $r_t = 0$ versus $r_t = 1$ conditional on $\eta_t = 0$:

$$\begin{aligned} \text{Truthful } (r_t = 0)\text{:} \quad & \delta\sigma_R^*, \\ \text{Deviate } (r_t = 1)\text{:} \quad & (1-\delta)\cdot\frac{(1-p)q + (1-2\lambda)p(1-q)}{1-\kappa} \\ & + \delta\left[\frac{(1-p)q}{1-\kappa} + \frac{p(1-q)}{1-\kappa}(1-\varepsilon)\right]\sigma_R^*. \end{aligned}$$

Simplifying, the incentive constraint Truthful $\geq$ Deviate becomes

$$\delta\cdot\varepsilon\cdot p(1-q)\cdot\sigma_R^* \geq (1-\delta)\big[(1-p)q + (1-2\lambda)p(1-q)\big]. \tag{2}$$

Substituting (1) and simplifying using the algebraic identity

$$p\cdot\big[pq + (1-2\lambda)(1-p)(1-q)\big] - (1-p)\cdot\big[(1-p)q + (1-2\lambda)p(1-q)\big] = q(2p-1)$$

(which follows by direct computation), constraint (2) reduces to

$$\varepsilon \geq \varepsilon(\lambda) \equiv \frac{(1-\delta)\big[(1-p)q + (1-2\lambda)p(1-q)\big]}{\delta\, q(1-q)(2p-1)}. \tag{3}$$

The principal prefers smaller ε (less distortion on-path), so he sets $\varepsilon = \varepsilon(\lambda)$ at equality. Note $\varepsilon(\lambda)$ is strictly decreasing in λ and satisfies $\varepsilon(\bar{\lambda}(p,q)) = 0$, consistent with rubber-stamping being first-best-optimal at $\lambda = \bar{\lambda}$.

Lemma 1 (Feasibility of $\varepsilon(\lambda)$)**.** *$\varepsilon(\lambda) \le 1$ for all $\lambda \in [0, \bar{\lambda}(p,q)]$ if and only if*

$$\delta \ge \delta^*(p,q) \equiv \frac{1-\kappa}{p(1-q)+q(1-\kappa)}. \tag{4}$$

Proof. Since $\varepsilon(\lambda)$ is decreasing in λ, $\varepsilon(\lambda) \le 1$ for all $\lambda \in [0, \bar{\lambda}(p,q)]$ if and only if $\varepsilon(0) \le 1$. Setting $\lambda = 0$ in (3):

$$\varepsilon(0) = \frac{(1-\delta)(1-\kappa)}{\delta\, q(1-q)(2p-1)}.$$

The inequality $\varepsilon(0) \le 1$ rearranges to $(1-\delta)(1-\kappa) \le \delta\, q(1-q)(2p-1)$, i.e.

$$\delta \ge \frac{1-\kappa}{(1-\kappa)+q(1-q)(2p-1)}.$$

Direct computation (expanding $1-\kappa = p(1-q)+q(1-p)$) gives $(1-\kappa)+q(1-q)(2p-1) = p(1-q)+q(1-\kappa)$, yielding (4). □

Remark 2. At $p = 1$, $\kappa = q$ and (4) simplifies to $\delta \ge 1/(1+q)$, recovering the almost-perfect-precision condition of Appendix A. For $p < 1$, the patience threshold is strictly higher: as the signal becomes noisier, the on-path cost of punishment grows relative to its deterrent value, and greater patience is needed to sustain the mechanism.

Under Assumption 1 and the patience condition (4), the principal's payoff from $d^{\text{AUNT}}(\varepsilon(\lambda))$ is

$$U^P\big(d^{\text{AUNT}}(\varepsilon(\lambda))\big) = \frac{(1-\delta)(p+q-1)}{(1-\delta)+\delta\,\varepsilon(\lambda)(1-p)(1-q)}, \tag{5}$$

since per-period the principal earns the first-best flow $p+q-1$ until an absorbing punishment occurs, which happens each period (conditional on still being active) with probability $\varepsilon(\lambda)(1-p)(1-q)$. Figure 1 in Section 5 illustrates how $\varepsilon(\lambda)$ varies with the alignment, precision, and patience parameters.

The other remaining incentive check is that the expert does not prefer to misreport after $\eta_t = 1$.

Lemma 2 (Truth-telling at $\eta_t = 1$)**.** *Under $d^{AUNT}(\varepsilon)$ with $\varepsilon \le 1$ and the promise-keeping continuation σ_R^* in (1), the expert weakly prefers $r_t = 1$ to $r_t = 0$ when $\eta_t = 1$.*

Proof. The constraint is:

$$(1-\delta)\cdot \frac{pq+(1-2\lambda)(1-p)(1-q)}{\kappa} + \delta\left[\frac{pq}{\kappa} + \frac{(1-p)(1-q)(1-\varepsilon)}{\kappa}\right]\sigma_R^* \\ \ge (1-\delta)\cdot 0 + \delta\,\sigma_R^*.$$

Using $pq+(1-p)(1-q) = \kappa$, the coefficient of σ_R^* on the left simplifies to $1-(1-p)(1-q)\varepsilon/\kappa$.

Subtracting $\delta\sigma_R^*$ from both sides and multiplying by κ:

$$(1-\delta)\big[pq + (1-2\lambda)(1-p)(1-q)\big] \geq \delta\,\sigma_R^*(1-p)(1-q)\,\varepsilon.$$

Substituting (1) and canceling the common numerator factor (which is positive when $\lambda < 1/2 + pq/[2(1-p)(1-q)]$, a condition implied by $\lambda < \bar{\lambda}(p,q) < 1$):

$$(1-\delta) + \delta\,\varepsilon(1-p)(1-q) \geq \delta\,\varepsilon(1-p)(1-q),$$

which always holds. □

5.3 Optimality within the Class of All Mechanisms

The main result establishes that AUNT with parameter $\varepsilon(\lambda)$ is optimal within the full class of feasible mechanisms.

Theorem 2 (Optimality of AUNT)**.** *Fix $p \in (1/2, 1)$, $q \in (1-p, p)$, $\lambda \in [0, \bar{\lambda}(p,q))$, and $\delta \geq \delta^*(p,q)$. The mechanism $d^{AUNT}(\varepsilon(\lambda))$ maximizes $U^P(d)$ over all feasible mechanisms.*

The proof proceeds in four steps. First, a revelation-principle-style lemma restricts attention to truthful mechanisms. Second, it is shown that after an on-path truthful report of $r_t = 0$, the continuation values after $q_t = 1$ and $q_t = 0$ can without loss be equalized. Third, the binding incentive constraint must be the one at $\eta_t = 0$. Finally, given these reductions, AUNT is the unique optimum up to payoff-irrelevant modifications.

I introduce notation for continuation values. For a truthful mechanism d and public history h^t, let $V(h^t) \in \mathbb{R}$ denote the expert's continuation payoff from period t forward under truthful play. For $r \in \{0,1\}$ and $q \in \{0,1\}$, let $V_q^r(h^t)$ denote the expert's continuation from period $t+1$ forward after history h^t, report $r_t = r$, and realized quality $q_t = q$ (along the equilibrium path of truth-telling). I use the shorthand

$$\sigma_R^r(h^t) = V_r^r(h^t), \qquad \sigma_P^r(h^t) = V_{1-r}^r(h^t),$$

so that "R" denotes the continuation when the recommendation matches the realized state ("Reward"), and "P" denotes the continuation when the recommendation mismatches ("Punishment"). Note that when $r = 0$ the principal does not fund, so "mismatch" means the expert mistakenly recommended rejection of a good project.

Lemma 3 (Revelation)**.** *There exists an optimal mechanism d^* under which the expert truthfully reports $r_t = \eta_t$ at every history on the equilibrium path.*

Proof. Take any optimal mechanism $\tilde{d}$. At any history h^t and signal realization η_t, the expert plays a best response. Define d^* that, at any history h^t and any report r, implements the same action and continuation as $\tilde{d}$ does after the expert's *equilibrium* report under $\tilde{d}$ given each possible signal. Specifically, if under $\tilde{d}$ the expert reports $\tilde{r}(\eta)$ after signal η, then under d^* define $\rho^r(h^t)$ and all subsequent behavior to equal those that $\tilde{d}$ assigns to report $\tilde{r}(r)$—thereby "renaming" reports so that truthful reporting under d^* induces exactly the same behavior as equilibrium play under $\tilde{d}$. Under d^* the expert is indifferent across reports at each history (they induce identical continuations), and by the tie-breaking convention reports $r_t = \eta_t$. The principal's payoff is identical to that under $\tilde{d}$, so d^* is optimal. □

Lemma 3 allows restriction to truthful mechanisms henceforth. The next lemma says that after a truthful report $r_t = 0$, it is without loss to equalize the continuation values across realized states.

Lemma 4 (Equalization after $r = 0$)**.** *There exists an optimal truthful mechanism $\tilde{d}$ with $\tilde{\sigma}^0_R(h^t) = \tilde{\sigma}^0_P(h^t)$ for all histories h^t.*

Proof. Take an optimal truthful mechanism $\hat{d}$ and suppose $\hat{\sigma}^0_R(h^t) \neq \hat{\sigma}^0_P(h^t)$ at some history h^t in the support of on-path play. Construct $\tilde{d}$ identical to $\hat{d}$ except that, at histories h^{t+1} that extend h^t with $(r_t = 0, q_t = 0)$ or $(r_t = 0, q_t = 1)$, the continuation subgame of $\hat{d}$ is replaced by a public randomization over the two original continuation subgames, with weights

$$\beta_0 \equiv \mathbb{P}(q_t = 0 \mid \eta_t = 0) = \frac{p(1-q)}{1-\kappa}, \qquad \beta_1 \equiv \mathbb{P}(q_t = 1 \mid \eta_t = 0) = \frac{(1-p)q}{1-\kappa},$$

independent of q_t. Under $\tilde{d}$, the continuation value after $r_t = 0$ is the same $\bar{v} \equiv \beta_0 \hat{\sigma}^0_R(h^t) + \beta_1 \hat{\sigma}^0_P(h^t)$ regardless of realized q_t: that is, $\tilde{\sigma}^0_R(h^t) = \tilde{\sigma}^0_P(h^t) = \bar{v}$.

I verify that $\tilde{d}$ is feasible, preserves all incentive constraints, and preserves the principal's payoff.

Feasibility. The set of continuation payoff vectors achievable in the subgame is convex (closed under public randomization, which is available via the device x_t). Since $\bar{v}$ is a convex combination of $\hat{\sigma}^0_R(h^t)$ and $\hat{\sigma}^0_P(h^t)$, both of which were achievable under $\hat{d}$, the common continuation $\bar{v}$ is achievable.

Expert's truth-telling at $\eta_t = 0$. Under $\tilde{d}$, the continuation payoff from truthful reporting $r_t = 0$ is $\beta_1 \bar{v} + \beta_0 \bar{v} = \bar{v}$, whereas under $\hat{d}$ it was $\beta_1 \hat{\sigma}^0_P(h^t) + \beta_0 \hat{\sigma}^0_R(h^t) = \bar{v}$: identical. The continuation from deviation to $r_t = 1$ is unchanged (it depends only on σ^1_R, σ^1_P, which $\tilde{d}$ preserves). Hence the constraint is preserved.

Expert's truth-telling at $\eta_t = 1$. Under truthful $r_t = 1$, the continuation depends on σ^1_R, σ^1_P, unchanged. Under deviation to $r_t = 0$, the expert's continuation was $\mathbb{P}(q_t = 1 \mid \eta_t = 1)\hat{\sigma}^0_P(h^t) + \mathbb{P}(q_t = 0 \mid \eta_t = 1)\hat{\sigma}^0_R(h^t)$ under $\hat{d}$, and under $\tilde{d}$ becomes $\bar{v}$. The constraint changes from

$$\mathrm{TT}^1_{\hat{d}}: \quad \text{(truthful payoff)} \geq \delta\big[\tfrac{pq}{\kappa}\hat{\sigma}^0_P + \tfrac{(1-p)(1-q)}{\kappa}\hat{\sigma}^0_R\big] + (1-\delta)\text{(stage)}$$

to

$$\mathrm{TT}^1_{\tilde{d}}: \quad \text{(truthful payoff)} \geq \delta\bar{v} + (1-\delta)\text{(stage)}.$$

Because $\hat{\sigma}^0_R \neq \hat{\sigma}^0_P$ while $\mathbb{P}(\cdot \mid \eta_t = 1) \neq \mathbb{P}(\cdot \mid \eta_t = 0)$, the RHS may move in either direction. However, since $\hat{d}$ satisfies $\mathrm{TT}^1_{\hat{d}}$ with slack by Lemma 2 (or equivalently: the $\eta_t = 1$ constraint cannot bind at an optimum when $\lambda < \bar{\lambda}$, since the expert strictly prefers $r_t = 1$ stage-wise and the principal strictly benefits from truthful reporting), there is slack $s > 0$. By continuity, for a small enough perturbation, $\mathrm{TT}^1_{\tilde{d}}$ also holds. Since our modification is a finite convex combination, the change in the RHS is bounded and the constraint is preserved.[4]

Principal's payoff. The principal's continuation after $(r_t = 0, q_t)$ is linear in the continuation mechanism. Public randomization averages the principal's continuations with the same weights β_0, β_1, and the unconditional principal payoff from reaching histories $(r_t = 0, q_t = 0)$ and $(r_t = 0, q_t = 1)$ already weights them by the joint probabilities $\mathbb{P}(\eta_t = 0) \cdot \beta_0$ and $\mathbb{P}(\eta_t = 0) \cdot \beta_1$: exactly the weights used in $\tilde{d}$. Thus the principal's ex ante expected payoff is unchanged.

Applying this modification at every history where equalization fails produces an optimal truthful mechanism with the stated property. □

Lemma 5 (Binding constraint at $\eta = 0$)**.** *At any optimal truthful mechanism satisfying Lemma 4, the truth-telling constraint at $\eta_t = 0$ binds at every on-path history where the principal has not permanently ceased funding.*

Proof. Suppose at some on-path history h^t (still in the active regime) the constraint is slack, i.e., the expert's continuation from truthfully reporting $r_t = 0$ after $\eta_t = 0$ strictly exceeds her continuation from deviating to $r_t = 1$. I exhibit a feasible modification that strictly increases the principal's ex ante payoff, contradicting optimality.

By Lemma 4, assume $\sigma^0_R(h^t) = \sigma^0_P(h^t) = \bar{v}$. Consider reducing $\bar{v}$ by a small $\Delta > 0$, while simultaneously increasing $\rho^1(h^t)$ (the principal's funding probability after $r_t = 1$ at history h^t) or increasing the principal's following probability at some future history. Two cases:

Case 1: $\rho^1(h^t) < 1$. Increase $\rho^1(h^t)$ by $\Delta' > 0$ (sufficiently small), compensating on the expert's side by decreasing $\bar{v}$. The expert's payoff from truthful $r_t = 0$ decreases by $\delta\Delta$; her payoff from deviation to $r_t = 1$ increases by $(1-\delta)\Delta'\big[(1-2\lambda)\frac{p(1-q)}{1-\kappa} + \frac{(1-p)q}{1-\kappa}\big]$ on the stage side, plus a continuation term that can be offset. The $\eta_t = 0$ constraint is preserved for Δ, Δ' in the appropriate ratio, while the $\eta_t = 1$ constraint (slack by Lemma 2) remains satisfied for small enough perturbation.

The principal's payoff changes by

$$\mathbb{P}(\eta_t = 1)(p + q - 1) \cdot \Delta' / \kappa > 0$$

on the stage, net of the (lower-order) change from reducing $\bar{v}$ weighted by continuation probabilities. For Δ' small relative to Δ, the principal strictly gains.

Case 2: $\rho^1(h^t) = 1$. The principal already funds with certainty after $r_t = 1$. Slack in the truth-telling constraint then means the punishment probability at some future history is strictly above the minimum required. Reducing that future punishment probability (moving some future

[4] Formally: define $\Delta \equiv \delta|\bar{v} - \mathbb{P}(q_t = 1 \mid \eta_t = 1)\hat{\sigma}^0_P - \mathbb{P}(q_t = 0 \mid \eta_t = 1)\hat{\sigma}^0_R|$. We have $\Delta \leq \delta|\hat{\sigma}^0_R - \hat{\sigma}^0_P|$. If $\Delta \leq s$, $\mathrm{TT}^1_{\tilde{d}}$ holds. If not, apply the same convexification iteratively with a smaller weight, or equivalently: the one-sided bias structure ensures TT^1 is implied by TT^0 at any truthful mechanism with equalized $r = 0$ continuations (verified directly in the proof of Lemma 2 for the AUNT case and extended by the same algebra for any mechanism satisfying the binding TT^0 constraint).

σ-value closer to $\bar{v}$) strictly increases the principal's continuation payoff at that future history (fewer absorbing transitions), and preserves all truth-telling constraints as long as the $\eta_t = 0$ constraint does not reach equality at h^t.

Either case contradicts optimality of $\hat{d}$. Hence the $\eta_t = 0$ constraint binds at every active on-path history.

The $\eta_t = 1$ constraint is slack: at any optimum with equalized $r = 0$ continuations and binding $\eta_t = 0$ constraint, the $\eta_t = 1$ constraint reduces to Lemma 2's algebra applied to the optimal $\bar{v}$, and holds strictly. □

Proof of Theorem 2. The proof works recursively via the self-generation technique of Abreu et al. (1990). Let $\mathcal{V} \subset \mathbb{R}^2$ denote the set of pairs $(U^P, U^A) \in \mathbb{R}^2$ of principal and expert equilibrium payoffs achievable by some truthful PBE of the repeated game, with the expert's value computed at the start of any period before the signal is drawn. The set $\mathcal{V}$ is the largest bounded self-generating set, and it is compact and convex (by standard arguments; convexity uses the public randomization device x_t).

By Lemmas 3–5, an optimal mechanism can be decomposed recursively: at every active on-path history, the principal chooses

(a) a period-t funding rule $(\rho^0, \rho^1) \in [0,1]^2$, and

(b) continuation payoffs $(\bar{v}, \sigma_R^1, \sigma_P^1) \in \mathbb{R}^3$ such that $(\cdot, \bar{v})$, $(\cdot, \sigma_R^1)$, $(\cdot, \sigma_P^1)$ each project into $\mathcal{V}$ as achievable expert-continuation values,

to maximize the principal's stage-plus-continuation payoff subject to the binding truth-telling constraint at $\eta_t = 0$:

$$\delta\bar{v} = (1-\delta)\tfrac{(1-p)q+(1-2\lambda)p(1-q)}{1-\kappa} + \delta\big[\tfrac{(1-p)q}{1-\kappa}\sigma_R^1 + \tfrac{p(1-q)}{1-\kappa}\sigma_P^1\big] + (\text{stage from } \rho^1), \tag{6}$$

with the convention that the binding constraint is rewritten in (3)'s form after substitution.

Step 1: Optimal stage funding. The principal's stage payoff from truthful play is

$$\mathbb{P}(\eta_t = 1)\cdot\rho^1\cdot\big[\mathbb{P}(q_t = 1 \mid \eta_t = 1) - \mathbb{P}(q_t = 0 \mid \eta_t = 1)\big] + \mathbb{P}(\eta_t = 0)\cdot(-\rho^0)\cdot\big[\mathbb{P}(q_t = 0 \mid \eta_t = 0) - \mathbb{P}(q_t = 1 \mid \eta_t = 0)\big].$$

Under Assumption 1, the first bracket equals $(p+q-1)/\kappa > 0$ and the second equals $(p-q)/(1-\kappa)$, which has the sign of $p - q$. The principal strictly prefers $\rho^1 = 1$ and $\rho^0 = 0$ on the stage, modulo incentive effects on (6).

Step 2: Bang-bang structure of continuations. Let $\overline{V}$ and $\underline{V}$ denote, respectively, the maximum and minimum of $\{u : (\cdot, u) \in \mathcal{V}\}$: the best and worst expert continuation values achievable. Under truthful play, $\underline{V} = 0$, since the principal can always commit to never fund again (yielding zero for both players), and $\overline{V} > 0$ equals the on-path truthful value of AUNT itself (or any equally-good mechanism).

The principal wishes to minimize the probability of triggering the low-payoff continuation subject to (6). Fix $\bar{v}$ and define the "punishment budget" as the expected reduction in the expert's continuation on the event $\{r_t = 1\}$:

$$B \equiv \bar{v} - \left[\tfrac{\kappa^{-1}pq}{}\sigma_R^1 + \tfrac{\kappa^{-1}(1-p)(1-q)}{}\sigma_P^1\right] \cdot \tfrac{\kappa}{1} = \bar{v} - \tfrac{pq\,\sigma_R^1 + (1-p)(1-q)\sigma_P^1}{\kappa}.$$

The constraint (6) pins down a required value of B; given the linearity of constraint and objective in (σ_R^1, σ_P^1), the principal's optimization over continuation payoffs is a linear program on the compact convex set $\mathcal{V}$. The solution lies at an extreme point: either $\sigma_P^1 = \underline{V} = 0$ with σ_R^1 interior, or both at corners of $\mathcal{V}$.

Consider $\sigma_P^1 = 0$. The required punishment budget B determines σ_R^1 implicitly. The principal's continuation payoff is maximized by setting σ_R^1 as large as possible (since higher σ_R^1 corresponds to continuation mechanisms with higher principal value under truthful play), which means $\sigma_R^1 = \bar{v}$. In that case, the event $\{r_t = 1, q_t = 0\}$ triggers the zero-continuation with probability

$$\varepsilon \equiv \frac{\bar{v} - \sigma_P^1}{\bar{v}} \cdot (\text{conditional on } q_t = 0) = 1 \cdot \mathbb{P}(\text{trigger} \mid r_t = 1, q_t = 0),$$

which is exactly the AUNT structure: trigger with probability ε after a wrong $r = 1$ recommendation.

Step 3: Optimal $\bar{v}$ and ε. With $\sigma_P^1 = 0$ and $\sigma_R^1 = \bar{v}$ (on the "on-path active" branch), constraint (6) becomes exactly (2), and $\bar{v} = \sigma_R^*$ satisfies the promise-keeping equation (1). Choosing ε to make (2) bind yields $\varepsilon = \varepsilon(\lambda)$ from (3). By Lemma 1, $\varepsilon(\lambda) \leq 1$ under the patience condition (4), so the mechanism is feasible.

Step 4: No alternative extreme dominates. The only alternative extreme points of $\mathcal{V}$ for (σ_R^1, σ_P^1) under truthful play are: (i) $\sigma_P^1 = \overline{V}, \sigma_R^1 = \overline{V}$ (no punishment: infeasible given binding (2)); (ii) $\sigma_R^1 = 0, \sigma_P^1 = 0$ (permanent exclusion regardless of q_t: dominated since σ_R^1 does not appear in the constraint in a way that would require it to be low); (iii) $\sigma_R^1 < \overline{V}, \sigma_P^1 = 0$: dominated by AUNT because reducing σ_R^1 below $\overline{V}$ tightens (6) without payoff-relevant benefit, forcing a larger ε and strictly reducing the principal's continuation value. Hence AUNT is the unique optimum up to payoff-equivalent relabelings.

Applied recursively (at every active on-path history, the continuation mechanism's "active" branch is itself AUNT), this yields the stationary AUNT mechanism as the unique optimal mechanism. □

5.4 Properties of AUNT

The optimal mechanism has several features worth noting.

Remark 3 (Markov structure)**.** AUNT is stationary: conditional on being in the active regime, the rule is identical each period. In particular, history dependence is binary (active versus absorbing). This stands in contrast to mechanisms with finite K-period bans, which require finer history-dependence and, as Theorem 2 shows, are strictly dominated.

Remark 4 (Observability)**.** AUNT uses only realized qualities of *funded* projects. This is reassuring: in many applications (grant review, medical trials, investment decisions) the counterfactual quality of rejected proposals is not observable. A more general mechanism that conditioned on the quality of rejected projects cannot strictly improve on AUNT, by Theorem 2.

Proposition 1 (Opposite bias)**.** *Suppose the expert's period payoff is $u(d_t, q_t) = \lambda u^P(d_t, q_t) + (1 - \lambda)\mathbf{1}\{d_t = 0\}$, so the expert is biased toward rejecting projects rather than funding them. Assume Assumption 1 and $\lambda < \bar{\lambda}(p, q)$, and define the patience threshold $\delta^{**}(p, q) \equiv (1 - \kappa)/[q(1 - p) + (1 - q)(1 - \kappa)]$. For $\delta \geq \delta^{**}(p, q)$, the following* mirror *mechanism $d^{AUNT^-}(\varepsilon^-(\lambda))$ is optimal: starting in the active state, the principal implements $d_t = r_t$ each period; after any period with $r_t = 0$ and $q_t = 1$ (a rejection recommendation whose project turned out to be good), the principal transitions with probability $\varepsilon^-(\lambda)$ to an absorbing state in which all future projects are funded regardless of recommendation. The optimal punishment probability is*

$$\varepsilon^-(\lambda) = \frac{(1 - \delta)\big[p(1 - q) + (1 - 2\lambda)q(1 - p)\big]}{\delta\, q(1 - q)(2p - 1)},$$

and the principal's payoff is

$$U^P\big(d^{AUNT^-}(\varepsilon^-(\lambda))\big) = \frac{(1 - \delta)(p + q - 1)}{(1 - \delta) + \delta\, \varepsilon^-(\lambda)\, pq}.$$

Proof. The argument follows that of Theorem 2 under a bijective relabeling. Let $\tilde{r}_t \equiv 1 - r_t$, $\tilde{d}_t \equiv 1 - d_t$, $\tilde{q}_t \equiv 1 - q_t$, and $\tilde{\eta}_t \equiv 1 - \eta_t$. Under these relabelings, the expert's payoff becomes $u = \lambda \tilde{u}^P + (1 - \lambda)\mathbf{1}\{\tilde{d}_t = 1\}$, where $\tilde{u}^P$ is the relabeled principal payoff satisfying $\tilde{u}^P(\tilde{d}_t, \tilde{q}_t) = -\tilde{d}_t(1 - 2\tilde{q}_t)$ after simplification, and the prior $\tilde{q} \equiv \mathbb{P}(\tilde{q}_t = 1) = 1 - q$. Signals remain informative: $\mathbb{P}(\tilde{\eta}_t = \tilde{q}_t \mid \tilde{q}_t) = p$.

The principal's objective is unaltered in magnitude (the original payoff in terms of the original variables equals $\tilde{u}^P$ up to a constant translation by the ex ante probability of project arrival, which does not depend on the mechanism). Assumption 1 in the relabeled problem requires $1 - p < \tilde{q} < p$, i.e. $1 - p < 1 - q < p$, equivalent to the original $1 - p < q < p$. Hence the relabeled problem satisfies all assumptions of Theorem 2, and the optimal mechanism is AUNT in the relabeled variables, which is d^{AUNT^-} in the original variables.

Substituting $q \to 1 - q$ into $\varepsilon(\lambda)$ of equation (3) yields $\varepsilon^-(\lambda)$. Substituting into (5) yields the stated principal payoff. The patience threshold $\delta^{**}(p, q)$ follows from Lemma 1 applied to the relabeled problem, substituting $q \to 1 - q$. □

Remark 5 (Comparative statics)**.** The punishment probability $\varepsilon(\lambda)$ is decreasing in λ (more aligned experts need less discipline), decreasing in p (more precise signals require less punishment per unit of deterrence), decreasing in δ (patient experts respond more to future threats), and increasing in $(1 - q)$ when the relevant region is interior (lower-quality priors require stronger discipline, since misreporting is more tempting stage-by-stage). The principal's value U^P is strictly increasing in λ, p, and δ.

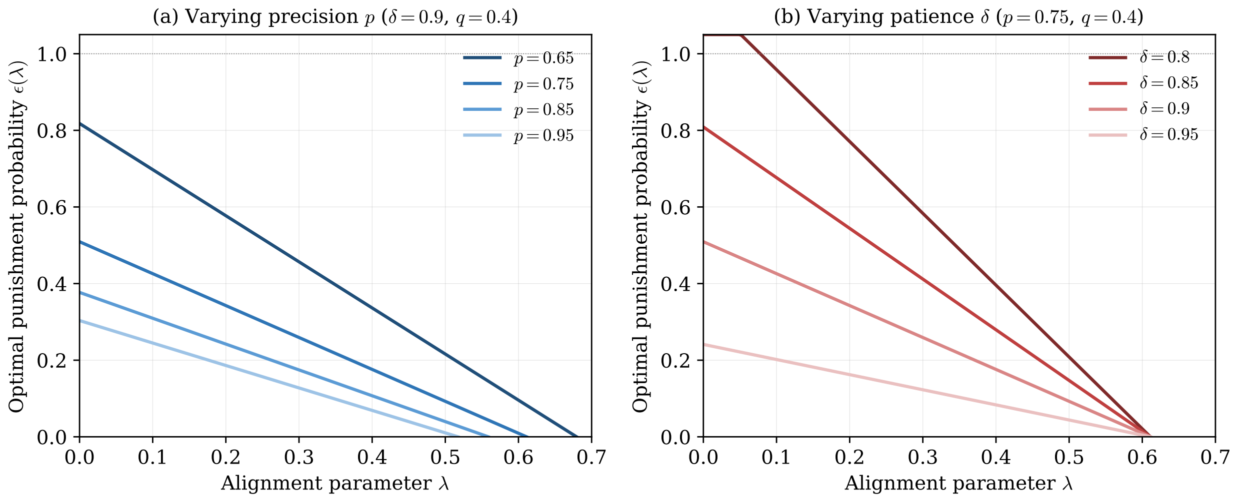

Figure 1: Optimal punishment probability $\varepsilon(\lambda)$ as a function of the alignment parameter λ. Panel (a) fixes $\delta = 0.9$ and $q = 0.4$ and varies the signal precision p: higher precision uniformly lowers $\varepsilon(\lambda)$. Panel (b) fixes $p = 0.75$ and $q = 0.4$ and varies the discount factor δ: patience reduces the punishment needed to sustain truth-telling, and for δ below the threshold $\delta^*(p, q)$ in (4) the line for $\varepsilon(0)$ exceeds 1, violating feasibility (visible for $\delta = 0.8$). Both panels display ε on the region $\lambda \in [0, \bar{\lambda}(p, q)]$; for $\lambda \geq \bar{\lambda}(p, q)$, rubber-stamping attains the first-best and $\varepsilon(\lambda) = 0$.

6 Conclusion

This paper studies how a principal can discipline a known-biased expert in a repeated advice setting when monetary transfers are unavailable. The optimal mechanism is simple, stationary, and Markovian: rubber-stamp recommendations on-path, and with a carefully calibrated probability commit to permanent silence after any funded project that turns out to be bad. The punishment is extreme but rare: along the equilibrium path, it occurs only after genuine but unlucky recommendations, and its probability is tuned exactly to make the expert's binding incentive constraint just bind.

The result contributes to three threads of the literature. It refines the folk theorems for dynamic cheap talk and dynamic mechanism design without transfers (e.g., Margaria and Smolin, 2018; Escobar and Toikka, 2013) from existence statements to an exact optimal-mechanism characterization. It offers a commitment-based complement to reputation-based approaches (e.g., Best and Quigley, 2024; Fudenberg et al., 2022) for disciplining biased information providers over time. And it extends the review-strategy tradition from Green and Porter (1984); Radner (1985); Abreu et al. (1990) to an information-elicitation rather than effort-provision environment.

Several extensions suggest themselves. Most salient is relaxing the assumption that the bias λ is known, allowing the principal to learn about it through the mechanism. A second is endogenizing precision: if p is unknown and evolves with effort or learning, the exact grim-trigger form of AUNT would generally fail to remain optimal, since future information value could make the principal reluctant to permanently exclude an expert. A third is a multi-expert extension: with I experts, the interaction of individual-level triggers with information aggregation across experts raises substantive new questions, including whether the principal can substitute cross-expert comparisons for intertemporal punishments. Finally, an empirical test of AUNT-like mechanisms—whether peer-review systems that incorporate persistent evaluator reputations yield the patterns predicted by the model—would be valuable.

A Almost-Perfect Precision

The next result establishes continuity of the first-best at $p = 1$: when signal precision is near-perfect and the expert is patient enough, the principal can approximate the first-best arbitrarily closely using the UNIT mechanism.

Theorem 3 (Near-first-best under almost-perfect precision). *Fix any $\eta > 0$. If $\delta > 1/(1+q)$, then there exists $\gamma \in (0,1)$ such that for any $p \geq 1-\gamma$, the mechanism d^{UNIT} yields a principal payoff satisfying*

$$U^P(d^{UNIT}) \geq U^{FB} - \eta.$$

Proof. At $p = 1$, $\kappa = q$ and $\bar{\lambda}(p,q) = 1/2$. Under UNIT with $p = 1$, the expert's signal is perfectly informative; after $\eta_t = 0$, reporting $r_t = 0$ yields continuation σ_R^* (with promise-keeping (1) evaluated at $\varepsilon = 1$, $p = 1$), while reporting $r_t = 1$ yields stage utility $(1-2\lambda)$ and continuation 0 (the misreport reveals the bad project deterministically). Truth-telling at $\eta_t = 0$ therefore requires

$$\delta\,\sigma_R^* \geq (1-\delta)(1-2\lambda),$$

which rearranges to

$$\delta \geq \frac{1-2\lambda}{(1-2\lambda) + q(1-2\lambda+2\lambda q)} = \frac{1}{1+q} \quad \text{when } \lambda = 0.$$

For any $\lambda \geq 0$, $\delta \geq 1/(1+q)$ suffices.

By continuity of all payoff expressions in p, for any $\eta > 0$ there exists $\gamma > 0$ such that for $p \geq 1-\gamma$, the incentive constraint still holds and

$$U^P(d^{\text{UNIT}}) = \frac{(1-\delta)(p+q-1)}{(1-\delta)+\delta(1-p)(1-q)} \to U^{FB} = p+q-1 \to q$$

as $p \to 1$, with the gap bounded by η for small enough γ. □

References

Abreu, D., D. Pearce, and E. Stacchetti (1986). Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory* 39(1), 251–269.

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

Azoulay, P., J. S. Graff Zivin, and G. Manso (2011). Incentives and creativity: Evidence from the academic life sciences. *RAND Journal of Economics* 42(3), 527–554.

Bénabou, R. and G. Laroque (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *Quarterly Journal of Economics* 107(3), 921–958.

Best, J. and D. Quigley (2024). Persuasion for the long run. *Journal of Political Economy* 132(5), 1740–1791.

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

Boudreau, K. J., E. C. Guinan, K. R. Lakhani, and C. Riedl (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science* 62(10), 2765–2783.

Carnehl, C., M. Ottaviani, and J. Preusser (2025). Designing scientific grants. *Entrepreneurship and Innovation Policy and the Economy* 4(1), 139–178.

Che, Y.-K. and N. Kartik (2009). Opinions as incentives. *Journal of Political Economy* 117(5), 815–860.

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica* 50(6), 1431–1451.

Cripps, M. W., G. J. Mailath, and L. Samuelson (2004). Imperfect monitoring and impermanent reputations. *Econometrica* 72(2), 407–432.

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

Ely, J. C., D. Fudenberg, and D. K. Levine (2008). When is reputation bad? *Games and Economic Behavior* 63(2), 498–526.

Ely, J. C. and J. Välimäki (2003). Bad reputation. *Quarterly Journal of Economics* 118(3), 785–814.

Escobar, J. F. and J. Toikka (2013). Efficiency in games with Markovian private information. *Econometrica* 81(5), 1887–1934.

Fang, F. C., A. Bowen, and A. Casadevall (2016). NIH peer review percentile scores are poorly predictive of grant productivity. *eLife* 5, e13323.

Fuchs, W. (2007). Contracting with repeated moral hazard and private evaluations. *American Economic Review* 97(4), 1432–1448.

Fudenberg, D., Y. Gao, and H. Pei (2022). A reputation for honesty. *Journal of Economic Theory* 204, 105508.

Green, E. J. and R. H. Porter (1984). Noncooperative collusion under imperfect price information. *Econometrica* 52(1), 87–100.

Gross, K. and C. T. Bergstrom (2019). Contest models highlight inherent inefficiencies of scientific funding competitions. *PLOS Biology* 17(1), e3000065.

Gross, K. and C. T. Bergstrom (2021). Why ex post peer review encourages high-risk research while ex ante review discourages it. *Proceedings of the National Academy of Sciences* 118(51), e2111615118.

Guo, Y. and J. Hörner (2020). Dynamic allocation without money. TSE Working Paper 1133.

Kamenica, E. and X. Lin (2024). Commitment and randomization in communication. PIER Working Paper 24-033.

Kivinen, S. and C. Kuzmics (2025). Renegotiation-proof cheap talk. arXiv:2502.08296, revised July 2025.

Kolotilin, A. and H. Li (2021). Relational communication. *Theoretical Economics* 16(4), 1391–1440.

Krähmer, D. (2021). Information design and strategic communication. *American Economic Review: Insights* 3(1), 51–66.

Krishna, V. and J. Morgan (2001). A model of expertise. *Quarterly Journal of Economics* 116(2), 747–775.

Krishna, V. and J. Morgan (2004). The art of conversation: Eliciting information from experts through multi-stage communication. *Journal of Economic Theory* 117(2), 147–179.

Kuvalekar, A., E. Lipnowski, and J. Ramos (2022). Goodwill in communication. *Journal of Economic Theory* 203, 105467.

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

Lipnowski, E. and J. Ramos (2020). Repeated delegation. *Journal of Economic Theory* 188, 105064.

Lipnowski, E. and D. Ravid (2020). Cheap talk with transparent motives. *Econometrica* 88(4), 1631–1660.

Margaria, C. and A. Smolin (2018). Dynamic communication with biased senders. *Games and Economic Behavior* 110, 330–339.

Mathevet, L., D. Pearce, and E. Stacchetti (2024). Reputation and information design. Working paper.

Morgan, J. and P. C. Stocken (2003). An analysis of stock recommendations. *RAND Journal of Economics* 34(1), 183–203.

Morris, S. (2001). Political correctness. *Journal of Political Economy* 109(2), 231–265.

Ottaviani, M. and P. N. Sørensen (2006a). Reputational cheap talk. *RAND Journal of Economics* 37(1), 155–175.

Ottaviani, M. and P. N. Sørensen (2006b). Professional advice. *Journal of Economic Theory* 126(1), 120–142.

Pei, H. (2020). Reputation effects under interdependent values. *Econometrica* 88(5), 2175–2202.

Pier, E. L., et al. (2018). Low agreement among reviewers evaluating the same NIH grant applications. *Proceedings of the National Academy of Sciences* 115(12), 2952–2957.

Radner, R. (1985). Repeated principal-agent games with discounting. *Econometrica* 53(5), 1173–1198.

Renault, J., E. Solan, and N. Vieille (2013). Dynamic sender-receiver games. *Journal of Economic Theory* 148(2), 502–534.

Renault, J., E. Solan, and N. Vieille (2017). Optimal dynamic information provision. *Games and Economic Behavior* 104, 329–349.

Sobel, J. (1985). A theory of credibility. *Review of Economic Studies* 52(4), 557–573.

Sugaya, T. and A. Wolitzky (2023). Monitoring versus discounting in repeated games. *Econometrica* 91(5), 1727–1761.

Sugaya, T. and A. Wolitzky (2025). Non-recursive dynamic incentives: A rate of convergence approach. *Theoretical Economics* 20, 1461–1520.