Screening Feedback for Language Models with Costly Verification

Anonymous authors
Paper under double-blind review

Abstract

Language model training and alignment rely on high-quality human feedback, yet platforms must incentivize valuable contributions while limiting harmful feedback from non-experts. We study a screening environment in which a platform commits to a uniform policy (ρ, R, P)—a verification rate, a reward for submitting feedback, and a sanction imposed when verified feedback is harmful—and heterogeneous users decide whether to participate. High-type users are more likely to produce helpful feedback, while low-type users are more likely to generate harmful feedback, and user types may differ in their effective exposure to sanctions.

We characterize platform-optimal verification, reward, and sanction policies under costly verification in robust pure-participation regimes. A key boundary condition, $\phi_H(1-\eta_H) = \phi_L(1-\eta_L)$, separates parameter regions in which incentives implement normal separation, where high types participate and low types abstain, from regions exhibiting reverse screening, where low types participate while high types are deterred. Verification is the primary policy margin: it improves the value of screened feedback and enables sanctions, but it is costly because aggregate verification costs are convex in the mass of verified feedback. Rewards are pinned down by participation constraints, while sanctions are useful only when their expected collections exceed the induced increase in reward compensation and are limited by enforcement and reputational costs.

We further show that, under optimal verification, platform profit need not increase monotonically with population quality. In our numerical illustration, this non-monotonicity appears as an inverted-U pattern, with profit peaking at intermediate shares of high-type users. The mechanism is that a larger high-type population changes the platform's optimal verification intensity, and the resulting adjustment in verification benefits and costs can make additional high-type participation less profitable at the margin. Finally, we provide an illustrative simulation using a bigram language model as a transparent calibration exercise to generate plausible magnitudes for (η_H, η_L) and to visualize the model's comparative statics.

Keywords: Costly verification, Incentives, User feedback, Quality control, Platform economics, Language models

1 Introduction

Human feedback is a central input to modern language model training and alignment pipelines (Ouyang et al., 2022). Yet platforms that solicit such feedback face a basic quality-control problem: they must attract contributors who are likely to provide useful feedback while limiting participation by contributors whose feedback is more likely to be harmful. This problem is especially important when user populations are heterogeneous, with expert contributors coexisting alongside a larger pool of lower-expertise users.

This setting connects to a broad literature on crowdsourcing and collective intelligence, which emphasizes that contributor quality cannot be reduced to expertise alone. Diverse groups may outperform homogeneous groups of high-ability individuals when contributors provide complementary perspectives or problem-solving heuristics (Hong & Page, 2004; Aggarwal et al., 2015), and incentives play an important role in shaping

collective performance (Mann & Helbing, 2017). Our focus is different but complementary. In language-model feedback pipelines, individual contributions may directly affect model updates, harmful feedback may degrade training, and feedback quality is often difficult to observe ex ante and costly to verify ex post.

We study this trade-off in a simple screening environment with costly verification. A platform commits to a uniform policy (ρ, R, P) applied to all users: a verification rate ρ, a reward R for submitting feedback, and a sanction P imposed when verified feedback is harmful. The sanction should be interpreted broadly. It may correspond to a withheld reward, a forfeited deposit, a clawback of future payments, reduced access to future feedback tasks, reputation loss, or suspension from an account-based feedback program. Users differ in their probability of generating helpful feedback and in their effective exposure to sanctions. After observing the platform's policy, users decide whether to participate.

Although we occasionally use the term "mechanism" in the broad platform-design sense, our model does *not* study a standard direct-revelation mechanism in which users report their types and the platform implements type-contingent allocations. Instead, we focus on the practically salient case in which expertise is difficult to certify and the platform's levers are verification, rewards, and enforcement applied uniformly through product design and terms of service. The setting is therefore best viewed as a screening-and-enforcement problem with costly verification rather than a classical mechanism design problem.

Our first contribution is to characterize when screening is feasible. A key boundary condition, $\phi_H(1-\eta_H) = \phi_L(1-\eta_L)$, partitions the parameter space into two robust pure-participation regimes. When $\phi_H(1-\eta_H) < \phi_L(1-\eta_L)$, the platform can implement normal separation: high-type users participate and low-type users abstain. When the inequality is reversed, the platform enters a reverse-screening region in which low-type users participate while high-type users are deterred. This reverse-screening outcome identifies a failure mode of sanction-based governance: enforcement intended to deter harmful feedback may instead discourage high-quality contributors if those contributors face sufficiently high risk-adjusted expected sanction exposure.

Our second contribution is to characterize the platform-optimal policy within these regimes. Verification is the primary strategic instrument. It improves the value of screened feedback and enables sanctions, but it is costly because aggregate verification costs are convex in the mass of verified feedback. Rewards are pinned down by participation constraints. Sanctions, in turn, are useful only when expected sanction collections exceed the induced increase in reward compensation needed to keep the participating type willing to provide feedback, and they are limited by enforcement and reputational costs. Thus, screening performance is determined jointly by verification intensity, reward compensation, and deterrence incentives.

Because the platform commits to (ρ, R, P) before users choose whether to participate, non-degenerate mixed participation is a knife-edge phenomenon. It requires exact user indifference and is generically eliminated by arbitrarily small platform deviations that break indifference. Accordingly, we interpret the equality $\phi_H(1-\eta_H) = \phi_L(1-\eta_L)$ as the boundary between the two robust pure-participation regimes, rather than as a stable region with interior mixing.

Our third contribution is to show that platform profit need not increase monotonically with population quality. Under optimal verification, a higher share of high-type users changes both the scale of participation and the platform's optimal verification intensity. Because verification costs are convex in the verified mass, the marginal value of additional high-type participation may be outweighed by participation costs and the adjustment in verification benefits and costs. As a result, optimal profit can decline over some range of population quality. In our numerical illustration, this non-monotonicity appears as an inverted-U pattern, with profit peaking at an intermediate share of high-type users.

To connect the theory to a transparent numerical example, we include an illustrative simulation based on a bigram language model trained on the Brown Corpus (Schubert & Tong, 2003). The exercise is not intended to empirically validate the model or make quantitative claims about modern large language model training. Instead, it provides an interpretable calibration for the key helpfulness parameters (η_H, η_L) and visualizes the model's comparative statics in a controlled setting.

This work is closely connected to research on crowdsourcing and collective intelligence, screening and enforcement under costly verification, platform governance, and incentive provision for quality contributions

(Hong & Page, 2004; Aggarwal et al., 2015; Mann & Helbing, 2017; Townsend, 1979; Gale & Hellwig, 1985; Kofman & Lawarrée, 1993). We review and position these connections in Appendix A.

The remainder of the paper is organized as follows. Section 2 presents the model. Section 3 characterizes the optimal policy across regimes. Section 4 provides illustrative simulations. Section 5 concludes. Appendix A reviews related literature, and Appendix B contains proofs.

2 Feedback Screening with Heterogeneous Users

We consider a game between a platform (a language-model provider) and a continuum of users indexed by $i \in [0, 1]$. Users can optionally provide feedback that may improve or harm the platform's model. The key friction is that assessing feedback quality is feasible but costly, so the platform chooses verification intensity and uses rewards and sanctions to screen participation.

The model is represented as a function $f : T^* \to \Delta(T)$ mapping an input token sequence to a distribution over the next token, following Duetting et al. (2024). Here T is the vocabulary, $\Delta(T)$ is the set of distributions over T, and $T^* = T \cup T^2 \cup \cdots \cup T^K$ is the set of token sequences up to context length K. Generation is autoregressive: starting with prompt $s_0 \in T^*$, the model samples $\tau_1 \sim f(s_0)$, forms $s_1 = s_0 \oplus \tau_1$, and iterates $\tau_k \sim f(s_{k-1})$, $s_k = s_{k-1} \oplus \tau_k$, until termination. The model f is parameterized by architecture M and weights W, and platform quality depends on training and fine-tuning procedures that incorporate user feedback.

Each user i has a private type $t_i \in \{H, L\}$, where H denotes a high type (expert) and L denotes a low type (non-expert). Types are i.i.d. with $\mathbb{P}(t_i = H) = \lambda$ and $\mathbb{P}(t_i = L) = 1 - \lambda$.

If a user provides feedback, it can be helpful or harmful. We classify a feedback instance as helpful if incorporating it would increase expected log-likelihood on a held-out validation distribution. Let f^0 denote the baseline model before incorporating user i's feedback, and let $f^{(i)}$ denote the counterfactual model obtained by updating f^0 on the feedback instance (s_i, τ_i) according to the platform's update rule. Formally,

$$\tilde{\sigma}(s_i, \tau_i) := \mathbb{E}_{(s',\tau')\sim\mathcal{V}} \left[\log f^{(i)}(\tau' \mid s') - \log f^0(\tau' \mid s') \right], \qquad r_i := \mathbf{1}\{\tilde{\sigma}(s_i, \tau_i) > 0\},$$

where $\mathcal{V}$ is a held-out validation distribution over prompt–continuation pairs.

This likelihood-improvement criterion is motivated by recent likelihood-based views of alignment from preference feedback. In particular, Tang and Feng (Tang & Feng, 2025) propose Ranked Choice Preference Optimization (RCPO), which frames preference alignment as maximum likelihood estimation of ranked choice models. In RCPO, the training objective is a ranked choice log-likelihood $\sum_i \log g(\mu_i^k, S_i, \{r_{\pi_\theta}(x_i, y)\}_{y \in S_i})$, where the reward is derived from policy log-probability ratios in the DPO family, e.g.,

$$r_{\pi_\theta}(x, y) = \beta \log \frac{\pi_\theta(y \mid x)}{\pi_{\text{ref}}(y \mid x)} + \beta \log Z(x),$$

so that updates are naturally driven by log-probability differences. Under this perspective, using held-out expected log-likelihood improvement $\tilde{\sigma}(s_i, \tau_i)$ as a proxy for whether incorporating a feedback instance improves model fit provides a likelihood-consistent notion of "helpfulness".

We define the type-dependent helpfulness probability

$$\eta_t := \mathbb{P}(r_i = 1 \mid t_i = t) = \mathbb{P}\big(\tilde{\sigma}(s_i, \tau_i) > 0 \mid t_i = t\big),$$

and assume $\eta_H > \eta_L$.

The penalty P should be interpreted as an enforceable platform sanction rather than an unrestricted monetary fine. In practice, it may correspond to withheld rewards, forfeited deposits, reduced access to future feedback tasks, reputation losses, or suspension from an account-based feedback program. Thus, the model is most applicable to settings with repeated interaction or platform control over future payments and access; when such enforcement is weak, the effective penalty is lower.

The interaction is a sequential (Stackelberg) game:

1. The platform commits to a uniform policy (ρ, R, P), where $\rho \in [0,1]$ is the verification rate, $R \geq 0$ is the reward for submitting feedback, and $P \geq 0$ is a penalty (sanction) applied when verified feedback is harmful.
2. Users observe (ρ, R, P) and each user chooses an action $a_i \in \{F, N\}$, where F denotes providing feedback (incurring cost $c > 0$) and N denotes not providing feedback.
3. If $a_i = F$, the user submits a feedback string $\tau_i \in T^*$, yielding a feedback instance (s_i, τ_i). The platform verifies the instance with probability ρ: let $v_i \sim \text{Bernoulli}(\rho)$. If $v_i = 1$, the platform observes whether the feedback is harmful ($r_i = 0$) and applies the penalty accordingly.

Verification is costly at the aggregate level. Let $k > 0$ denote the verification-cost parameter; as specified in the platform objective below, total verification cost is quadratic in the mass of verified feedback, $\frac{k}{2}(\theta^{v,H} + \theta^{v,L})^2$, capturing diseconomies of scale in review. Transfers to user i are

$$m_i = \begin{cases} R & \text{if } a_i = F \text{ and } (v_i = 0 \text{ or } r_i = 1), \\ R - P & \text{if } a_i = F \text{ and } v_i = 1 \text{ and } r_i = 0, \\ 0 & \text{if } a_i = N. \end{cases}$$

Equivalently, $m_i = \mathbf{1}\{a_i = F\}\,[R - P \cdot \mathbf{1}\{v_i = 1, r_i = 0\}]$. In practice, P need not be a monetary fine: it can be implemented through escrow/deposits that are forfeited upon verified harmful feedback, reward clawbacks or withholding of future payouts, or account-level sanctions (e.g., suspension from the feedback program) with an equivalent monetized cost. Accordingly, we interpret P as a reduced-form measure of expected sanction severity conditional on verified harmful feedback, while allowing for enforcement and reputational constraints captured by the platform's penalty cost term below.

For unverified feedback ($v_i = 0$), the platform does not observe whether the instance is helpful or harmful. We summarize its effective quality by reduced-form probabilities η_u^+ and η_u^-, with $\eta_u^+, \eta_u^- \in [0,1]$ and $\eta_u^+ + \eta_u^- = 1$. The net term $\eta_u^+ - \eta_u^-$ should be interpreted as the average effective impact of unverified feedback, possibly after filtering, aggregation, or downweighting; it may vary with the participating pool but is treated as a regime-specific primitive in the policy characterizations below.

User decision-making is modeled through a reduced-form participation condition with type-specific sensitivity to sanctions. We treat $\phi_t > 0$ as an effective sanction-exposure coefficient summarizing how type-t users convert formal penalties into perceived participation costs. The mean–variance formulation in Appendix B.1 provides one illustrative microfoundation for this heterogeneity, showing how local risk aversion can amplify expected penalties and generate $\phi_t \geq 1$. The core screening analysis below does not impose this restriction and relies only on the reduced-form expected penalty term $\phi_t P \rho(1 - \eta_t)$.

Let $\bar{U} \equiv U(s_{i,k})$ denote the deterministic base utility from receiving the model response, independent of type. Under the local approximation, type-t utility is

$$u_t(a_i) = \begin{cases} \bar{U} + R - c - \phi_t P q_t & \text{if } a_i = F, \\ \bar{U} & \text{if } a_i = N, \end{cases}$$

where $q_t \equiv \rho(1 - \eta_t)$ is the probability that a type-t user faces a penalty when providing feedback, and $\phi_t \equiv 1 + \psi_t \bar{P}(1 - \bar{q}_t)$ is the type-specific risk-adjusted penalty coefficient defined around a reference point $(\bar{P}, \bar{q}_t)$. Thus, a type-t user provides feedback if and only if

$$R - c \geq \phi_t P q_t = \phi_t P \rho(1 - \eta_t).$$

Let $\theta^{F,H}$ and $\theta^{F,L}$ denote the mass of feedback supplied by high and low types. Verified and unverified masses are

$$\theta^{v,H} = \rho\,\theta^{F,H}, \quad \theta^{v,L} = \rho\,\theta^{F,L}, \quad \theta^u = (1 - \rho)(\theta^{F,H} + \theta^{F,L}).$$

Model quality improvement is

$$\Delta Q = \gamma_v\big(\eta_H \theta^{v,H} + \eta_L \theta^{v,L}\big) + \gamma_u(\eta_u^+ - \eta_u^-)\theta^u,$$

where $\gamma_v > \gamma_u > 0$ weight verified and unverified feedback.

The platform chooses (ρ, R, P) to maximize profit, trading off quality improvements, verification costs, transfers, and enforcement/reputational costs. We assume a quadratic verification cost $\frac{k}{2}(\theta^{v,H} + \theta^{v,L})^2$ to capture diseconomies of scale in review, and a quadratic policy-level reputational/enforcement cost $\alpha P^2(\theta^{F,H} + \theta^{F,L})$. The latter is interpreted as the cost of maintaining and communicating a sanction schedule of severity P for the participating user base, rather than as a cost incurred only when sanctions are actually imposed. The platform's objective is

$$\Pi(\rho, R, P) = \underbrace{\gamma_v[\eta_H\theta^{v,H} + \eta_L\theta^{v,L}] + \gamma_u(\eta_u^+ - \eta_u^-)\theta^u}_{\text{Quality improvement}} - \underbrace{\frac{k}{2}(\theta^{v,H} + \theta^{v,L})^2}_{\text{Verification costs}}$$
$$- \underbrace{R(\theta^{F,H} + \theta^{F,L})}_{\text{Reward payments}} + \underbrace{P\rho\big[(1-\eta_H)\theta^{F,H} + (1-\eta_L)\theta^{F,L}\big]}_{\text{Penalty collections}} - \underbrace{\alpha P^2(\theta^{F,H} + \theta^{F,L})}_{\text{Reputational/enforcement costs}} . \quad (1)$$

We focus on subgame-perfect equilibrium outcomes of this sequential game given the platform's commitment to (ρ, R, P).

3 Results

3.1 Perfect Separation

Theorem 3.1 (Perfect Separation Condition)**.** *A perfect-separation equilibrium, in which high-type users provide feedback $\theta^{F,H} = \lambda$ and low-type users abstain $\theta^{F,L} = 0$, exists if and only if*

$$\phi_H(1 - \eta_H) < \phi_L(1 - \eta_L). \quad (2)$$

Equivalently,

$$\frac{\phi_H}{\phi_L} < \frac{1 - \eta_L}{1 - \eta_H}. \quad (3)$$

Moreover, in any platform-optimal policy implementing perfect separation, the high-type participation constraint binds:

$$R = c + \phi_H P\rho(1 - \eta_H). \quad (4)$$

Proof. See Appendix B.2. □

Economic implications. The condition says that perfect separation is feasible only when high-type users face a lower risk-adjusted expected sanction from participation than low-type users. Since low types are more likely to generate harmful feedback, this requirement is easier to satisfy when their higher error rate is reinforced by greater effective penalty sensitivity. Conversely, if high types are sufficiently exposed to sanctions, the same penalty-based policy intended to deter low types may also discourage the contributors the platform wants to attract.

The binding reward condition describes the least-cost way to implement separation. In a platform-optimal separating policy, the reward is set just high enough to compensate high types for their participation cost and expected risk-adjusted sanction exposure:

$$R = c + \phi_H P\rho(1 - \eta_H). \quad (5)$$

Any higher reward would increase transfer costs without improving sorting.

Verification and sanctions jointly generate the type-dependent expected costs needed for screening. If either $P = 0$ or $\rho = 0$, this screening channel disappears: both types then face the same reward-only participation margin. Thus, successful separation depends not only on differences in feedback quality, but also on how different user types respond to enforcement risk.

A technical qualification applies to the boundary cases in the optimal-policy characterizations below. Strict screening requires positive sanction-based screening intensity, $P\rho > 0$; if either $P = 0$ or $\rho = 0$, the two types face the same reward-only participation margin and strict type separation cannot be implemented through incentives. Accordingly, when a closed-form solution below yields $P^* = 0$ or $\rho^* = 0$, we interpret it as a boundary solution of the relaxed policy problem, or equivalently as a limiting case approached by policies with arbitrarily small positive $P\rho$, rather than as an exact strict-separation implementation.

Theorem 3.2 (Optimal Policy under Perfect Separation)**.** *Under perfect separation conditions, the platform's optimal mechanism is characterized by the following two cases:*

Case 1: $\phi_H < 1$***,*** *when the condition*

$$k\lambda > \frac{(1-\eta_H)^2(1-\phi_H)^2}{2\alpha} \tag{6}$$

is satisfied, the optimal mechanism is:

$$\rho^* = \min\left\{1, \max\left\{0, \frac{\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-)}{k\lambda - \frac{(1-\eta_H)^2(1-\phi_H)^2}{2\alpha}}\right\}\right\} \tag{7}$$

$$P^* = \frac{\rho^*(1-\eta_H)(1-\phi_H)}{2\alpha} \tag{8}$$

$$R^* = c + \phi_H P^* \rho^* (1-\eta_H) \tag{9}$$

Case 2: *When $\phi_H \geq 1$, or when $\phi_H < 1$ but condition equation 6 is violated, the optimal mechanism is characterized as follows:*

(i) If $\phi_H \geq 1$***:***

$$\rho^* = \min\left\{1, \max\left\{0, \frac{\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-)}{k\lambda}\right\}\right\} \tag{10}$$

$$P^* = 0 \tag{11}$$

$$R^* = c \tag{12}$$

(ii) If $\phi_H < 1$ ***and equation 6 is violated:***

$$\rho^* \in \{0, 1\}, \tag{13}$$

with the boundary choice given by

$$\rho^* = \begin{cases} 1, & \text{if } \gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-) - \dfrac{k\lambda}{2} + \dfrac{(1-\eta_H)^2(1-\phi_H)^2}{4\alpha} \geq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{14}$$

and

$$P^* = \frac{\rho^*(1-\eta_H)(1-\phi_H)}{2\alpha}, \tag{15}$$

$$R^* = c + \phi_H P^* \rho^* (1-\eta_H). \tag{16}$$

Economic implications. The theorem shows that, conditional on perfect separation, verification is the platform's main margin of adjustment. In Case 1, penalties are useful because the high type's effective penalty coefficient is below one. With the high-type participation constraint binding, increasing P raises the reward required to keep high types participating by $\phi_H\rho(1-\eta_H)$ per unit of participating high-type mass. However, it also raises expected penalty collections by $\rho(1-\eta_H)$. Hence the net marginal benefit of increasing the penalty before reputational/enforcement costs is $\rho(1-\eta_H)(1-\phi_H)$, which is positive when $\phi_H < 1$. The

optimal penalty therefore rises with verification intensity, while the reward is pinned down by the binding high-type participation constraint.

The denominator in equation 7 captures the net curvature of verification. Verification is costly, but it also increases the effectiveness of penalties by raising the probability that harmful feedback is detected and sanctioned. When condition equation 6 holds, this trade-off is concave and yields an interior verification rule, truncated to the feasible interval. When high types are sufficiently penalty-sensitive, penalties are no longer profitable because the induced reward increase weakly exceeds the expected penalty collections, so the platform sets $P^* = 0$ and relies only on verification. If condition equation 6 fails, the verification objective becomes non-concave, and the platform optimally chooses a boundary policy with either no verification or full verification.

3.2 Mixed Strategies and Non-Robust Indifference

Our model is a sequential commitment game: the platform commits to a policy (ρ, R, P) and users then decide whether to provide feedback. In such commitment environments, non-degenerate mixed participation, i.e., $p_t \in (0, 1)$ for some type $t \in \{H, L\}$, can arise only when that type is exactly indifferent between providing and not providing feedback. This indifference is a knife-edge property of the policy parameters and is eliminated by arbitrarily small perturbations of the reward, penalty, or verification rate.

We therefore do not treat mixed or semi-separating participation as robust screening regimes. This does not rule out the possibility that, under a particular tie-breaking rule or equilibrium-selection convention, an interior participation probability could be selected at an indifference point. Rather, our focus is on robust pure-participation regimes whose implementation does not rely on exact indifference. Accordingly, the mixed-participation condition is interpreted as a boundary between the two robust screening regimes: perfect separation and reverse screening.

Let a type-t user's expected payoff gain from providing feedback, relative to not providing feedback, be

$$\Delta_t(\rho, R, P) \;\equiv\; (R - c) \;-\; \phi_t \, P \, \rho \, (1 - \eta_t), \qquad t \in \{H, L\}. \tag{17}$$

Then type t chooses F if $\Delta_t > 0$, chooses N if $\Delta_t < 0$, and is indifferent if $\Delta_t = 0$.

Lemma 3.3 (Interior mixing requires knife-edge indifference)**.** *Fix any policy (ρ, R, P). If $\Delta_t(\rho, R, P) \neq 0$, then type t has a unique pure best response, so any best-response participation probability satisfies $p_t \in \{0, 1\}$. If $\Delta_t(\rho, R, P) = 0$, then type t is indifferent and any $p_t \in [0, 1]$ can be supported by an appropriate tie-breaking or mixing convention.*

Moreover, such interior mixing is not robust to small policy perturbations. For any $\varepsilon > 0$, setting $R' = R + \varepsilon$ yields $\Delta_t(\rho, R', P) > 0$, while setting $R' = R - \varepsilon$ yields $\Delta_t(\rho, R', P) < 0$. Thus an arbitrarily small reward perturbation selects a pure best response. Consequently, mixed or semi-separating participation can arise only at indifference points and is not a robust screening regime under the platform's commitment policy.

Proof. The payoff gain from participation for type t is

$$\Delta_t(\rho, R, P) = (R - c) - \phi_t P \rho (1 - \eta_t). \tag{18}$$

The user's Stage-2 objective is linear in the participation decision. If $\Delta_t(\rho, R, P) > 0$, participation is the unique best response; if $\Delta_t(\rho, R, P) < 0$, non-participation is the unique best response. Hence strict payoff differences imply $p_t \in \{0, 1\}$.

If $\Delta_t(\rho, R, P) = 0$, the user is indifferent between participation and non-participation, so any mixing probability can be supported by a tie-breaking or mixing convention. However, this indifference is knife-edge. For any $\varepsilon > 0$,

$$\Delta_t(\rho, R + \varepsilon, P) = \varepsilon > 0, \qquad \Delta_t(\rho, R - \varepsilon, P) = -\varepsilon < 0. \tag{19}$$

Thus an arbitrarily small reward perturbation breaks indifference and induces a pure best response. The same logic applies to perturbations of P or ρ whenever these instruments affect the expected penalty term. Therefore, interior mixing requires exact indifference and is not robust to small policy perturbations. □

Thus the mixed-participation condition is a boundary, not a robust screening regime. The equality

$$\phi_H(1-\eta_H) = \phi_L(1-\eta_L) \tag{20}$$

implies that the two types have the same expected risk-adjusted penalty exposure per unit $P\rho$. In particular, any policy satisfying the high-type participation indifference condition

$$R = c + \phi_H P\rho(1-\eta_H) \tag{21}$$

also makes the low type indifferent, and vice versa. Hence strict type sorting through (R, P, ρ) is impossible on equation 20.

Lemma 3.3 establishes a robustness, rather than dominance, claim. Interior participation probabilities can be supported only when the relevant type is exactly indifferent. Because such indifference is destroyed by arbitrarily small policy perturbations, we treat mixed and semi-separating outcomes as knife-edge cases and focus on robust pure-participation regimes.

Accordingly, we do not treat mixed or semi-separating, one-type-mixing outcomes as separate equilibrium regimes in the sequential model. Instead, equation 20 is interpreted as the knife-edge boundary separating the two robust screening regimes:

- **Perfect separation** when $\phi_H(1-\eta_H) < \phi_L(1-\eta_L)$ (Theorem 3.1);
- **Reverse screening** when $\phi_H(1-\eta_H) > \phi_L(1-\eta_L)$ (Theorem 3.4).

3.3 Reverse Screening

Penalty-based screening can also generate the opposite of the intended sorting pattern. If high-quality users have sufficiently high risk-adjusted exposure to sanctions, then policies that make low types willing to participate may deter high types instead. We call this outcome reverse screening.

Theorem 3.4 (Reverse Screening Condition)**.** *Strict reverse screening, with $(p_H, p_L) = (0, 1)$, that is, low-type users participate while high-type users are strictly deterred, exists if and only if*

$$\phi_L(1-\eta_L) < \phi_H(1-\eta_H). \tag{22}$$

Equivalently,

$$\frac{\phi_H}{\phi_L} > \frac{1-\eta_L}{1-\eta_H}. \tag{23}$$

Moreover, in any optimal reverse-screening mechanism, the low-type participation constraint binds, so the reward satisfies

$$R = c + \phi_L P\rho(1-\eta_L). \tag{24}$$

Proof. The proof can be found in Appendix B.4. □

Economic implications. The condition identifies reverse screening as the mirror image of perfect separation. Reverse screening arises when high-type users face a higher risk-adjusted expected sanction from participation than low-type users. In that case, a reward level that is sufficient to attract low types may still be too low to compensate high types for their expected sanction exposure. Thus, penalty-based governance can select the wrong contributors when high-quality users are disproportionately sensitive to sanctions or exposed to enforcement risk.

In any optimal reverse-screening policy, the platform sets the reward just high enough to induce low-type participation. The binding condition

$$R = c + \phi_L P\rho(1-\eta_L) \tag{25}$$

reflects the fact that any higher reward would only increase transfer costs while preserving the same reverse-screening outcome.

The same boundary qualification applies in the reverse-screening regime: exact strict reverse screening requires $P\rho > 0$. Boundary solutions with $P^* = 0$ or $\rho^* = 0$ should therefore be read as relaxed or limiting policy candidates rather than exact strict implementations.

Theorem 3.5 (Optimal Policy under Reverse Screening)**.** *Consider the reverse-screening regime in which high-type users abstain and low-type users participate,* $(p_H, p_L) = (0, 1)$*. Strict implementation requires*

$$\phi_L(1-\eta_L) < \phi_H(1-\eta_H), \tag{26}$$

while the equality case is a knife-edge case requiring tie-breaking against high-type participation. Then the platform's optimal policy is characterized by the following cases.

Case 1: $\phi_L < 1$*. If*

$$k(1-\lambda) > \frac{(1-\eta_L)^2(1-\phi_L)^2}{2\alpha}, \tag{27}$$

then the optimal mechanism is

$$\rho^* = \min\left\{1, \max\left\{0, \frac{\gamma_v\eta_L - \gamma_u(\eta_u^+ - \eta_u^-)}{k(1-\lambda) - \frac{(1-\eta_L)^2(1-\phi_L)^2}{2\alpha}}\right\}\right\}, \tag{28}$$

$$P^* = \frac{\rho^*(1-\eta_L)(1-\phi_L)}{2\alpha}, \tag{29}$$

$$R^* = c + \phi_L P^* \rho^* (1-\eta_L). \tag{30}$$

Case 2: $\phi_L \geq 1$***, or*** $\phi_L < 1$ ***with non-concavity.*** *If* $\phi_L \geq 1$*, then penalties are not profit-improving and the optimal mechanism is*

$$\rho^* = \min\left\{1, \max\left\{0, \frac{\gamma_v\eta_L - \gamma_u(\eta_u^+ - \eta_u^-)}{k(1-\lambda)}\right\}\right\}, \tag{31}$$

$$P^* = 0, \tag{32}$$

$$R^* = c. \tag{33}$$

If instead $\phi_L < 1$ *but condition equation 27 is violated, then the objective is weakly convex in verification intensity and the optimum occurs at a boundary:*

$$\rho^* \in \{0, 1\}. \tag{34}$$

The boundary choice is

$$\rho^* = \begin{cases} 1, & \text{if } \gamma_v\eta_L - \gamma_u(\eta_u^+ - \eta_u^-) - \dfrac{k(1-\lambda)}{2} + \dfrac{(1-\eta_L)^2(1-\phi_L)^2}{4\alpha} \geq 0, \\ 0, & \text{otherwise,} \end{cases} \tag{35}$$

with

$$P^* = \frac{\rho^*(1-\eta_L)(1-\phi_L)}{2\alpha}, \tag{36}$$

$$R^* = c + \phi_L P^* \rho^* (1-\eta_L). \tag{37}$$

Proof. The proof can be found in Appendix B.5. □

Economic implications. Conditional on reverse screening, the platform optimizes over a participating pool composed only of low-type users. The policy therefore mirrors the perfect-separation case, but with the relevant population mass given by $1-\lambda$ and feedback quality governed by η_L. Verification remains the main adjustment margin: it increases the value of screened feedback and enables penalties, but it is costly when applied to a larger participating mass of low-type users.

When $\phi_L < 1$, penalties can be profit-improving because the platform's expected penalty revenue and reward-saving effect outweigh the additional compensation needed to keep low types participating. The optimal penalty therefore increases with verification intensity. When $\phi_L \geq 1$, penalties are no longer profitable, so the platform sets $P^* = 0$ and relies only on verification. If the verification objective is non-concave, the platform does not choose an interior verification rate; it instead selects a boundary policy with either no verification or full verification.

4 Illustrative Simulations and Stylized Calibration

Our theoretical results characterize the platform-optimal verification–reward–penalty policy across screening regimes. This section provides a transparent numerical illustration of the model's primitives and comparative statics. Because the simulation environment is intentionally simple, it is not intended as an empirical validation of the model or as a quantitative claim about large language model training. Rather, it serves two purposes: (i) to produce plausible magnitudes for the key heterogeneity parameters (η_H, η_L) in a controlled setting, and (ii) to visualize how optimal verification and profit respond to changes in costs and population composition.

We conduct two complementary exercises. A micro-simulation uses a bigram language model trained on the Brown Corpus (Schubert & Tong, 2003) to obtain a stylized calibration of the helpfulness probabilities η_H and η_L. A macro-simulation then uses these calibrated magnitudes to illustrate the platform's optimal policy and the implied profit comparative statics across population shares and cost parameters.

4.1 Micro-Simulation for Stylized Calibration

To obtain transparent magnitudes for (η_H, η_L), we implement a micro-simulation using a bigram language model trained on the Brown Corpus. The bigram model is a minimal and interpretable testbed in which we can explicitly control the source of model errors and the quality of user feedback. The goal is not to represent the full complexity of modern language models, but to generate a simple environment in which high-type feedback is more likely to be helpful than low-type feedback, consistent with the model assumption $\eta_H > \eta_L$.

We estimate η_H and η_L using 10,000 Monte Carlo iterations.

4.1.1 Simulation Setup

We train a bigram language model f_{truth} on 2,000 sentences from the Brown Corpus with a vocabulary of 5,000 tokens. The model represents transition probabilities $P(w_{t+1} \mid w_t)$ estimated via maximum likelihood with Laplace smoothing ($\alpha = 0.01$). To create a model that can benefit from feedback, we generate a flawed model f_T by intentionally degrading 100 randomly selected prompts. For each flawed prompt w_t, we identify the correct continuation w^* (highest probability under f_{truth}) and an alternative w^- (randomly selected from the top five alternatives), then swap and exaggerate these probabilities:

$$f_T(w^* \mid w_t) = 0.5 \cdot f_{\text{truth}}(w^- \mid w_t), \qquad f_T(w^- \mid w_t) = 1.5 \cdot f_{\text{truth}}(w^* \mid w_t), \tag{38}$$

followed by renormalization.

We model two user types with different feedback accuracies: High-Type users provide correct feedback with probability 88%, while Low-Type users provide correct feedback with probability 38%. Upon receiving a feedback token $\hat{w}$ for prompt w_t, we update the model using an additive learning rule with rate $\delta = 0.2$, and we evaluate feedback helpfulness using expected log-likelihood improvement on a validation corpus:

$$\tilde{\sigma} = \mathbb{E}_{(w_t, w_{t+1}) \sim \text{Validation}} \left[\log f_{T+1}(w_{t+1} \mid w_t) - \log f_T(w_{t+1} \mid w_t) \right]. \tag{39}$$

Feedback is classified as helpful if $\tilde{\sigma} > 0$.

4.1.2 Results

Table 1 presents example interactions. The table includes prompts with [UNK] (unknown) tokens, representing words outside the training vocabulary or tokens mapped to the unknown category during preprocessing.

Table 1: Example user interactions showing differential feedback quality between High-Type (experts) and Low-Type (non-experts) users. [UNK] denotes unknown tokens not in the model's vocabulary.

Prompt	Truth	Model	H-Feed	H-Correct	H-$\tilde{\sigma}$	L-Feed	L-Correct	L-$\tilde{\sigma}$
out	of	in	of	Yes	+8.4e-4	of	Yes	+8.4e-4
motel	in	issues	in	Yes	+5.0e-4	issues	No	-2.0e-5
enough	to	.	to	Yes	+1.8e-4	.	No	-7.7e-6
who	[UNK]	are	[UNK]	Yes	+3.3e-4	[UNK]	Yes	+3.3e-4
--	the	and	the	Yes	+4.3e-4	and	No	-3.7e-4
pink	[UNK]	awake	[UNK]	Yes	+5.0e-4	awake	No	-1.6e-5
care	plan	.	.	No	-1.9e-5	.	No	-1.9e-5
f.	[UNK]	lee	[UNK]	Yes	+4.6e-4	lee	No	-3.1e-5

Table 2: Summary statistics from 10,000 Monte Carlo simulations.

Metric	High-Type	Low-Type
Helpfulness probability (η)	0.892	0.427
Mean $\tilde{\sigma}$	7.70×10^{-4}	2.05×10^{-4}
Median $\tilde{\sigma}$	2.58×10^{-4}	-6.86×10^{-6}
Std. dev. $\tilde{\sigma}$	4.24×10^{-3}	2.85×10^{-3}
Ratio η_H/η_L	2.09	
Difference $\eta_H - \eta_L$	0.465	

High-Type users provide correct feedback more frequently and more often generate positive $\tilde{\sigma}$, while Low-Type users more often generate negative $\tilde{\sigma}$.

Table 2 summarizes results from 10,000 Monte Carlo iterations. The estimated helpfulness probability for High-Type users is $\eta_H = 0.892$. For Low-Type users, we obtain $\eta_L = 0.427$, which is higher than the 38% feedback accuracy because incorrect feedback can occasionally improve validation performance by chance. The Low-Type median $\tilde{\sigma}$ is negative, indicating that more than half of non-expert feedback is harmful in this stylized environment.

4.2 Macro-Simulation: Optimal Policy and Comparative Statics

We next illustrate the platform's optimal policy and resulting profits as we vary key economic parameters and the population share of high types. The macro-simulation uses parameter magnitudes consistent with Section 4.1 and is intended to visualize regime boundaries and comparative statics rather than to provide quantitative predictions.

4.2.1 Parameter Configuration

We set the helpfulness probabilities to $\eta_H = 0.88$ and $\eta_L = 0.38$, preserving the large quality gap observed in the micro-simulation while using a slightly more conservative η_L to reflect additional noise and heterogeneity not captured by the bigram testbed. Platform benefits are $\gamma_v = 2.2$ for verified feedback. We set the net quality impact from unverified feedback, $\gamma_u(\eta_u^+ - \eta_u^-)$, to 0.4, reflecting that unverified feedback is on average helpful but less valuable than verified feedback.

The verification cost parameter is $k = 3.0$, and reputation/enforcement sensitivity is $\alpha = 0.8$. Penalty sensitivities are $\phi_H = 0.5$ and $\phi_L = 1.2$, capturing the possibility that lower-quality users are effectively more penalty-sensitive. These values are interpreted as reduced-form effective sanction-exposure parameters, rather than as coefficients restricted to the illustrative mean–variance specification in Appendix B.1. The baseline population share is $\lambda = 0.25$, and the participation cost is $c = 0.5$.

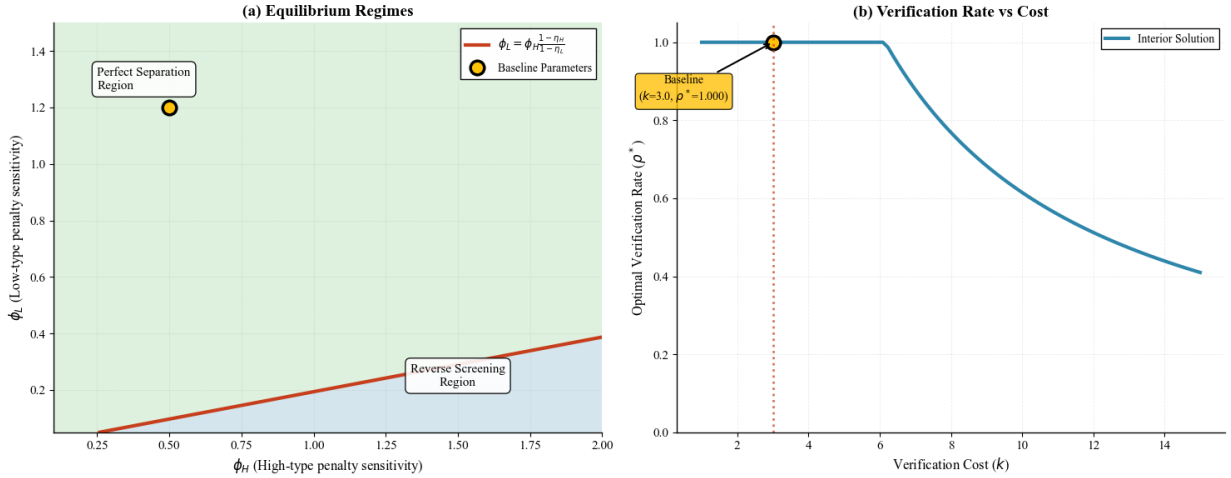

Figure 1: Equilibrium regimes and verification strategy. Panel (a) partitions the parameter space by the boundary condition $\phi_H(1-\eta_H) = \phi_L(1-\eta_L)$ into normal separation and reverse screening. Panel (b) illustrates the optimal verification rate ρ^* as verification cost k varies.

4.2.2 Equilibrium Regimes and Verification Strategy

Figure 1 illustrates two regime regions separated by $\phi_H(1-\eta_H) = \phi_L(1-\eta_L)$. In the normal-separation region, $\phi_H(1-\eta_H) < \phi_L(1-\eta_L)$, the platform can deter low types while maintaining high-type participation. In the reverse-screening region, $\phi_H(1-\eta_H) > \phi_L(1-\eta_L)$, policies that make low types willing to participate can deter high types.

Panel (b) illustrates how optimal verification responds to verification costs. At low costs, the platform chooses high verification intensity; as k increases, optimal verification becomes selective. The figure visualizes the threshold behavior implied by the theoretical characterizations.

4.2.3 Population Quality and the Inverted-U Profit Relationship

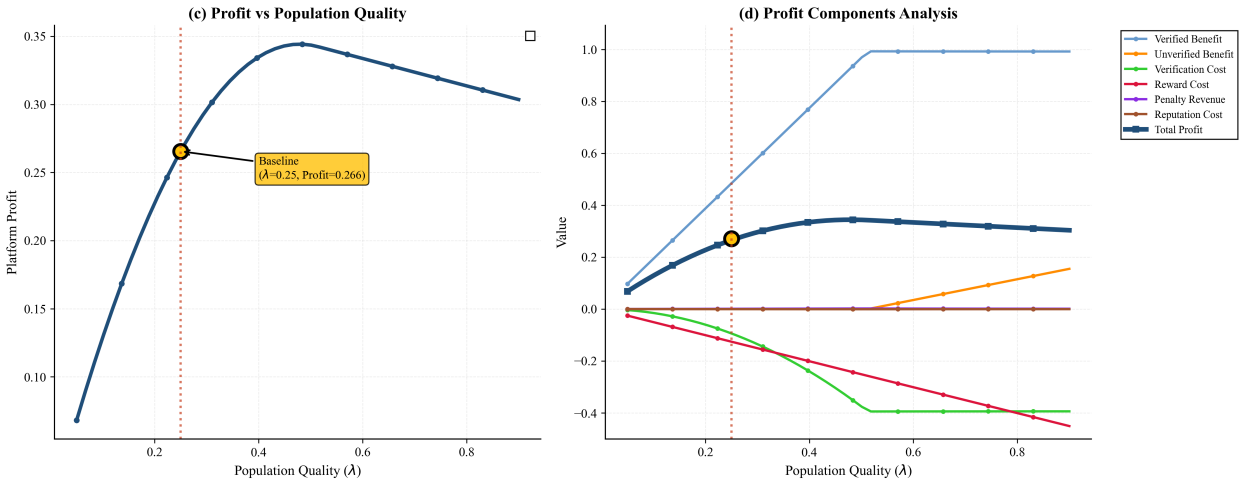

Figure 2: Population Quality and Profit Components Analysis. Panel (c) shows the relationship between population quality λ and platform profit. Panel (d) decomposes this relationship into constituent components. Note: reputation cost and penalty collections appear negligible due to the small optimal penalty levels under our parameter values, though they are non-zero.

Figure 2 illustrates an inverted-U relationship between population quality and platform profit: profit initially rises with λ, peaks around an interior point under this calibration, and then declines for high-quality populations. This pattern is consistent with the non-monotonicity result formalized below.

Proposition 1 (Non-Monotonicity of Profit in Population Quality)**.** *Maintain the perfect-separation equilibrium of Theorem 3.2. Suppose $\phi_H < 1$ and define*

$$N \equiv \gamma_v \eta_H - \gamma_u(\eta_u^+ - \eta_u^-), \qquad D \equiv \frac{(1-\eta_H)^2(1-\phi_H)^2}{2\alpha}. \tag{40}$$

Assume that the interior verification solution is feasible in a neighborhood of $\lambda = 1$, i.e.,

$$0 < N < k - D. \tag{41}$$

If the participation cost satisfies

$$c > \gamma_u(\eta_u^+ - \eta_u^-) - \frac{DN^2}{2(k-D)^2}, \tag{42}$$

then the platform's optimal profit $\Pi^*(\lambda)$ is not monotonically increasing in λ. In particular, there exists $\bar{\lambda} \in (0,1)$ such that

$$\left.\frac{d\Pi^*}{d\lambda}\right|_{\lambda=\bar{\lambda}} < 0. \tag{43}$$

Proof. See Appendix B.6. □

Panel (d) decomposes profit into its constituent components. The non-monotonicity is best interpreted as a scale effect under perfect separation. Since low-quality users abstain, an increase in λ expands the mass of high-type participants rather than improving the average quality of the participating pool. For small and moderate λ, this raises the value of verified feedback. However, because verification costs are convex in the verified mass $\rho\lambda$, a larger participating mass also makes verification increasingly costly at the margin.

5 Conclusion

This paper studies feedback screening under costly verification in language model training environments. We analyze a platform that commits to a uniform policy (ρ, R, P)—a verification rate, a reward for submitting feedback, and a sanction imposed when verified feedback is harmful—and a heterogeneous user population that differs in both feedback quality and effective exposure to sanctions.

Our main results characterize when screening is feasible and how verification, rewards, and sanctions interact in the platform-optimal policy. Successful screening depends on the relative strength of quality differences and sanction-exposure heterogeneity. The platform can implement normal separation, where high-type users participate and low-type users abstain, when

$$\frac{\phi_H}{\phi_L} < \frac{1-\eta_L}{1-\eta_H}.$$

When this inequality is reversed, the platform enters a reverse-screening region in which low-type users participate while high-type users are deterred. This regime highlights a failure mode of sanction-based governance: stronger enforcement need not improve screening outcomes if it imposes disproportionate expected participation costs on high-quality contributors.

Across regimes, verification is the primary strategic instrument. It improves the value of screened feedback and enables sanctions, but aggregate verification costs are convex in the mass of verified feedback. Rewards are pinned down by participation constraints. Sanctions are useful only when expected sanction collections exceed the induced increase in reward compensation needed to maintain participation by the targeted user type, and they are further limited by enforcement and reputational costs. Screening performance is therefore determined jointly by verification intensity, reward compensation, and deterrence incentives.

We also identify a non-monotonic comparative static in population quality. Under optimal verification, platform profit need not increase monotonically with the share of high-type users. In our numerical illustration, this non-monotonicity appears as an inverted-U pattern, with profit peaking at an intermediate level of population quality. The mechanism is that a higher share of high-type users changes both the scale of participation and the platform's optimal verification intensity. Because verification costs are convex in the verified mass, the marginal value of additional high-type participation can be outweighed by participation costs and the adjustment in verification benefits and costs.

The simulations are intended as transparent illustrations rather than empirical validation. Using a bigram language model testbed, we provide a stylized calibration for the key helpfulness parameters (η_H, η_L) and visualize the model's comparative statics in a controlled setting.

Several limitations suggest directions for future work. We assume static types, a simple verification technology, and reduced-form enforcement constraints. We also abstract from learning about user quality over time, platform competition, and richer forms of contributor heterogeneity. In practice, non-expert or diverse contributors may provide complementary information that is valuable after suitable aggregation. Extending the framework to dynamic reputation, Bayesian learning, multi-platform interaction, richer contributor diversity, and endogenous investment in verification would strengthen its applicability. Empirical work using operational data from feedback programs could further discipline parameter choices and quantify the effects of alternative verification and enforcement designs.

References

Ishani Aggarwal, Anita Williams Woolley, Christopher F Chabris, and Thomas W Malone. Cognitive diversity, collective intelligence, and learning in teams. *Proceedings of Collective Intelligence*, 1(3.1):3–3, 2015.

Mark Armstrong. Competition in two-sided markets. *The RAND journal of economics*, 37(3):668–691, 2006.

Luis Cabral and Ali Hortacsu. The dynamics of seller reputation: Evidence from ebay. *The journal of industrial economics*, 58(1):54–78, 2010.

Anirban Dasgupta and Arpita Ghosh. Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd international conference on World Wide Web*, pp. 319–330, 2013.

Paul Duetting, Vahab Mirrokni, Renato Paes Leme, Haifeng Xu, and Song Zuo. Mechanism design for large language models. In *Proceedings of the ACM Web Conference 2024*, pp. 144–155, 2024.

Douglas Gale and Martin Hellwig. Incentive-compatible debt contracts: The one-period problem. *The Review of Economic Studies*, 52(4):647–663, 1985.

Lu Hong and Scott E Page. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, 101(46):16385–16389, 2004.

Fred Kofman and Jacques Lawarrée. Collusion in hierarchical agency. *Econometrica: Journal of the Econometric Society*, pp. 629–656, 1993.

Yuqing Kong and Grant Schoenebeck. Water from two rocks: Maximizing the mutual information. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 177–194, 2018.

Yang Liu and Yiling Chen. Machine-learning aided peer prediction. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 63–80, 2017.

Richard P Mann and Dirk Helbing. Optimal incentives for collective intelligence. *Proceedings of the National Academy of Sciences*, 114(20):5077–5082, 2017.

Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.

Curtis Northcutt, Lu Jiang, and Isaac Chuang. Confident learning: Estimating uncertainty in dataset labels. *Journal of Artificial Intelligence Research*, 70:1373–1411, 2021.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

Drazen Prelec. A bayesian truth serum for subjective data. *science*, 306(5695):462–466, 2004.

Jean-Charles Rochet and Jean Tirole. Platform competition in two-sided markets. *Journal of the european economic association*, 1(4):990–1029, 2003.

Lenhart Schubert and Matthew Tong. Extracting and evaluating general world knowledge from the brown corpus. In *Proceedings of the HLT-NAACL 2003 workshop on Text meaning*, pp. 7–13, 2003.

Burr Settles. Active learning literature survey. 2009.

Nihar B Shah and Dengyong Zhou. Double or nothing: Multiplicative incentive mechanisms for crowdsourcing. *Journal of Machine Learning Research*, 17(165):1–52, 2016.

Haoran Sun, Yurong Chen, Siwei Wang, Wei Chen, and Xiaotie Deng. Mechanism design for llm fine-tuning with multiple reward models. *arXiv preprint arXiv:2405.16276*, 2024.

Steven Tadelis. Reputation and feedback systems in online platform markets. *Annual review of economics*, 8 (1):321–340, 2016.

Yuxuan Tang and Yifan Feng. Beyond pairwise: Empowering llm alignment with ranked choice modeling. *arXiv preprint arXiv:2510.23631*, 2025.

Robert M Townsend. Optimal contracts and competitive markets with costly state verification. *Journal of Economic theory*, 21(2):265–293, 1979.

Zhenzhe Zheng, Yanqing Peng, Fan Wu, Shaojie Tang, and Guihai Chen. An online pricing mechanism for mobile crowdsensing data markets. In *Proceedings of the 18th ACM international symposium on mobile ad hoc networking and computing*, pp. 1–10, 2017.

A Related Work

Our paper is related to work on crowdsourcing and collective intelligence, platform governance, incentive provision, and screening under costly verification. We study a platform that commits to a uniform policy (ρ, R, P) with endogenous verification and enforceable sanctions. We do not analyze a direct-revelation mechanism with type reports and menu-based allocations; instead, our focus is on screening heterogeneous contributors when expertise is difficult to certify.

Crowdsourcing, collective intelligence, and contributor heterogeneity. A large literature studies how heterogeneous contributors generate collective value. Work on collective intelligence shows that diverse groups can outperform homogeneous groups of high-ability individuals when contributors provide complementary perspectives or problem-solving heuristics (Hong & Page, 2004; Aggarwal et al., 2015). Related work also studies how incentives shape participation and collective performance (Mann & Helbing, 2017). Our analysis is complementary to this literature. Rather than focusing on aggregation benefits alone, we study a setting in which individual feedback may be harmful for model training, quality is costly to verify, and the platform uses verification, rewards, and sanctions to screen participation.

Costly verification, auditing, and enforcement. A central feature of our model is that feedback quality can be assessed through verification, but only at a cost, and the platform strategically chooses verification intensity. This connects our setting to classic contract-theoretic and principal–agent models with costly state verification and auditing (Townsend, 1979; Gale & Hellwig, 1985; Kofman & Lawarrée, 1993). Our contribution is to characterize how endogenous verification interacts with reward–penalty incentives to generate distinct screening regimes, including normal separation and reverse screening, in a feedback-quality environment.

Platforms and governance instruments. Our analysis also relates to platform economics, which emphasizes how platforms shape participation and value creation through prices, rules, and governance instruments (Rochet & Tirole, 2003; Armstrong, 2006). Dynamic feedback and learning considerations in platform environments are studied by Cabral & Hortacsu (2010). Relative to this literature, we focus on probabilistic verification and enforceable sanctions as governance tools for screening contributors when feedback quality is imperfectly observed.

Quality elicitation, peer prediction, and ML data quality. A related literature studies how to elicit informative reports when direct verification is difficult. Peer prediction and related methods use statistical relationships across reports to incentivize truthful information provision without ground truth (Miller et al., 2005; Prelec, 2004), while crowdsourcing work examines endogenous proficiency and information aggregation (Dasgupta & Ghosh, 2013; Kong & Schoenebeck, 2018). In machine learning systems, data quality is often addressed through technical tools such as data cleaning and uncertainty-based selection (Northcutt et al., 2021; Settles, 2009). Economic mechanisms for data procurement and AI-related incentive design appear in adjacent settings such as mobile crowdsensing, reputation systems, and AI training markets (Zheng et al., 2017; Tadelis, 2016; Shah & Zhou, 2016; Duetting et al., 2024; Sun et al., 2024; Liu & Chen, 2017). Our focus differs by centering costly verification and enforceable sanctions when expertise is difficult to certify and historical performance signals are limited.

B Proofs of Main Results

B.1 Constant Risk Adjustment Approximation

Lemma B.1 (Constant risk adjustment approximation)**.** *Let $q_t \equiv \rho(1-\eta_t) \in [0,1]$. For users who choose $a = F$, mean-variance preferences imply*

$$u_t^{MV}(P, q_t) = \bar{U} + R - c - Pq_t - \psi_t P^2 q_t(1-q_t),$$

where $\bar{U} \equiv U(s_{i,k})$ is deterministic and $\psi_t \geq 0$ is the type-specific risk aversion parameter.

Fix a reference mechanism $(\bar{P}, \bar{q}_t)$ with $\bar{P} \geq 0$ and $\bar{q}_t \in [0,1]$, and define

$$\phi_t \equiv 1 + \psi_t \bar{P}(1-\bar{q}_t).$$

Then for any (P, q_t),

$$u_t^{MV}(P, q_t) = \bar{U} + R - c - \phi_t Pq_t + \Delta_t(P, q_t),$$

where the approximation error is given by

$$\Delta_t(P, q_t) = -\psi_t P q_t \left(P(1-q_t) - \bar{P}(1-\bar{q}_t)\right).$$

Moreover, the error satisfies the bound

$$|\Delta_t(P, q_t)| \leq \psi_t Pq_t \left(|P - \bar{P}| + |q_t - \bar{q}_t| \max\{P, \bar{P}\}\right).$$

In particular, when (P, q_t) lies in a neighborhood of $(\bar{P}, \bar{q}_t)$, mean-variance utility is well approximated by the reduced form

$$u_t^{RF}(P, q_t) \equiv \bar{U} + R - c - \phi_t Pq_t,$$

which preserves the endogenous incentive term Pq_t while absorbing the second-order risk premium into a constant type-specific coefficient.

Proof. Starting from mean-variance utility,

$$u_t^{MV}(P, q_t) = \bar{U} + R - c - Pq_t - \psi_t P^2 q_t(1-q_t) = \bar{U} + R - c - Pq_t[1 + \psi_t P(1-q_t)].$$

Add and subtract $\psi_t \bar{P}(1-\bar{q}_t)$ inside the bracket to obtain

$$u_t^{MV}(P, q_t) = \bar{U} + R - c - Pq_t[1 + \psi_t \bar{P}(1-\bar{q}_t)] - \psi_t Pq_t[P(1-q_t) - \bar{P}(1-\bar{q}_t)].$$

Defining $\phi_t = 1 + \psi_t \bar{P}(1 - \bar{q}_t)$ yields the stated decomposition, with

$$\Delta_t(P, q_t) = -\psi_t P q_t [P(1 - q_t) - \bar{P}(1 - \bar{q}_t)].$$

The bound follows from the triangle inequality and the facts that $P(1-q_t) \in [0, P]$ and $\bar{P}(1-\bar{q}_t) \in [0, \bar{P}]$. □

Fixing a reference mechanism $(\bar{P}, \bar{q}_t)$, Lemma B.1 shows that mean–variance utility can be locally written as

$$\bar{U} + R - c - \phi_t P q_t, \qquad \phi_t = 1 + \psi_t \bar{P}(1 - \bar{q}_t),$$

up to an approximation error that is small near the reference point. Under the maintained assumptions in this illustrative mean–variance specification, $\phi_t \geq 1$. In the main text, however, $\phi_t > 0$ is treated as a reduced-form effective sanction-exposure coefficient. Thus, the screening analysis does not impose the mean–variance restriction $\phi_t \geq 1$; values below one capture cases in which formal penalties are attenuated by institutional or behavioral factors such as limited enforceability, imperfect salience, appeal opportunities, or lower perceived reputational costs.

B.2 Proof of Theorem 3.1

We prove both necessity and sufficiency of the separation condition.

Necessity: Suppose perfect separation equilibrium exists with $p_H = 1$ and $p_L = 0$.

Since high-types participate with probability 1, their participation constraint must be satisfied:

$$R - c - \phi_H P \rho (1 - \eta_H) \geq 0 \tag{PC-H}$$

Since low-types abstain with probability 1, they must strictly prefer not participating:

$$R - c - \phi_L P \rho (1 - \eta_L) < 0 \tag{PC-L}$$

For the platform to minimize costs while maintaining separation, the high-type constraint binds:

$$R = c + \phi_H P \rho (1 - \eta_H) \tag{44}$$

Substituting equation 44 into (PC-L):

$$c + \phi_H P \rho (1 - \eta_H) - c - \phi_L P \rho (1 - \eta_L) < 0 \tag{45}$$

$$P \rho [\phi_H (1 - \eta_H) - \phi_L (1 - \eta_L)] < 0 \tag{46}$$

If $P = 0$ or $\rho = 0$, then both participation constraints become $R - c \geq 0$ and $R - c < 0$, which is impossible. Therefore, we must have $P > 0$ and $\rho > 0$ in any meaningful separation equilibrium.

With $P\rho > 0$, we obtain:

$$\phi_H (1 - \eta_H) < \phi_L (1 - \eta_L) \tag{47}$$

Sufficiency: Suppose $\phi_H(1 - \eta_H) < \phi_L(1 - \eta_L)$ holds.

Set $R = c + \phi_H P \rho (1 - \eta_H)$ for some $P > 0$ and $\rho > 0$.

High-types' expected utility from participation is:

$$U_H = R - c - \phi_H P \rho (1 - \eta_H) = 0 \geq 0 \tag{48}$$

So high-types are willing to participate ($p_H = 1$ is optimal).

Low-types' expected utility from participation would be:

$$U_L = R - c - \phi_L P\rho(1-\eta_L) \tag{49}$$
$$= \phi_H P\rho(1-\eta_H) - \phi_L P\rho(1-\eta_L) \tag{50}$$
$$= P\rho[\phi_H(1-\eta_H) - \phi_L(1-\eta_L)] < 0 \tag{51}$$

So low-types strictly prefer to abstain ($p_L = 0$ is optimal).

Therefore, perfect separation equilibrium exists.

The equivalent ratio form follows by dividing both sides of the inequality by $\phi_L(1-\eta_H) > 0$:

$$\frac{\phi_H}{\phi_L} < \frac{1-\eta_L}{1-\eta_H} \tag{52}$$

B.3 Proof of Theorem 3.2

Under perfect separation, we have $\theta^{F,H} = \lambda$, $\theta^{F,L} = 0$, $\theta^{v,H} = \rho\lambda$, $\theta^{v,L} = 0$, and $\theta^u = (1-\rho)\lambda$. The platform's profit function is

$$\Pi(\rho, R, P) = \gamma_v \eta_H \rho\lambda + \gamma_u(\eta_u^+ - \eta_u^-)(1-\rho)\lambda - \frac{k}{2}(\rho\lambda)^2 - R\lambda + P\rho(1-\eta_H)\lambda - \alpha P^2\lambda. \tag{53}$$

In any perfect separation equilibrium, the high-type participation constraint binds, so

$$R = c + \phi_H P\rho(1-\eta_H). \tag{54}$$

Substituting into Π yields the reduced objective

$$\Pi(\rho, P) = \lambda\Big[\gamma_v\eta_H\rho + \gamma_u(\eta_u^+ - \eta_u^-)(1-\rho) - \frac{k\lambda}{2}\rho^2 - c\Big] + \lambda P\rho(1-\eta_H)(1-\phi_H) - \lambda\alpha P^2. \tag{55}$$

The choice variables satisfy $\rho \in [0,1]$ and $P \geq 0$.

Case 2(i) in Theorem 3.2: $\phi_H \geq 1$. When $\phi_H \geq 1$, we have $(1-\phi_H) \leq 0$. Hence, for any fixed $\rho \geq 0$, the term

$$\lambda P\rho(1-\eta_H)(1-\phi_H) - \lambda\alpha P^2 \tag{56}$$

is weakly decreasing in P on $[0,\infty)$, so the optimal penalty is

$$P^* = 0. \tag{57}$$

With $P^* = 0$, maximizing equation 55 over $\rho \in [0,1]$ gives

$$\left.\frac{\partial\Pi}{\partial\rho}\right|_{P=0} = \lambda\Big[\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-) - k\lambda\rho\Big], \tag{58}$$

so the optimal verification rate is

$$\rho^* = \min\left\{1, \max\left\{0, \frac{\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-)}{k\lambda}\right\}\right\}. \tag{59}$$

Finally, $R^* = c$ follows from the binding constraint $R = c + \phi_H P\rho(1-\eta_H)$ with $P^* = 0$. This establishes Case 2(i) of the theorem.

Case 1 and Case 2(ii) in Theorem 3.2: $\phi_H < 1$. When $\phi_H < 1$, we have $(1 - \phi_H) > 0$. For any fixed $\rho \in [0, 1]$, the reduced profit equation 55 is a strictly concave quadratic in P (since the coefficient on P^2 is $-\lambda\alpha < 0$). Therefore, for each ρ the unique maximizer over $P \geq 0$ satisfies the first-order condition

$$\frac{\partial \Pi}{\partial P} = \lambda\Big[\rho(1 - \eta_H)(1 - \phi_H) - 2\alpha P\Big] = 0, \tag{60}$$

which yields

$$P^*(\rho) = \frac{\rho(1 - \eta_H)(1 - \phi_H)}{2\alpha}. \tag{61}$$

Substituting equation 61 back into equation 55 gives a one-dimensional problem in ρ:

$$\Pi(\rho, P^*(\rho)) = \lambda\Big[\gamma_u(\eta_u^+ - \eta_u^-) - c + \rho\Big(\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-)\Big) - \frac{k\lambda}{2}\rho^2\Big] + \lambda \cdot \frac{\rho^2(1 - \eta_H)^2(1 - \phi_H)^2}{4\alpha}. \tag{62}$$

Equivalently, up to a ρ-independent constant,

$$\Pi(\rho, P^*(\rho)) = \text{const} + \lambda\Big(\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-)\Big)\rho - \frac{\lambda}{2}\Big(k\lambda - \frac{(1 - \eta_H)^2(1 - \phi_H)^2}{2\alpha}\Big)\rho^2. \tag{63}$$

Case 1 of the theorem (interior/concave region): Suppose $\phi_H < 1$ and condition equation 6 holds, i.e.,

$$k\lambda - \frac{(1 - \eta_H)^2(1 - \phi_H)^2}{2\alpha} > 0. \tag{64}$$

Then equation 63 is strictly concave in ρ, and the unconstrained maximizer solves

$$\frac{\partial}{\partial\rho}\Pi(\rho, P^*(\rho)) = \lambda\Big(\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-)\Big) - \lambda\Big(k\lambda - \frac{(1 - \eta_H)^2(1 - \phi_H)^2}{2\alpha}\Big)\rho = 0, \tag{65}$$

so

$$\tilde{\rho} = \frac{\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-)}{k\lambda - \frac{(1-\eta_H)^2(1-\phi_H)^2}{2\alpha}}. \tag{66}$$

Imposing $\rho \in [0, 1]$ yields

$$\rho^* = \min\left\{1, \max\left\{0, \frac{\gamma_v\eta_H - \gamma_u(\eta_u^+ - \eta_u^-)}{k\lambda - \frac{(1-\eta_H)^2(1-\phi_H)^2}{2\alpha}}\right\}\right\}. \tag{67}$$

Given ρ^*, the optimal penalty follows from equation 61:

$$P^* = \frac{\rho^*(1 - \eta_H)(1 - \phi_H)}{2\alpha}, \tag{68}$$

and the reward is pinned down by the binding participation constraint:

$$R^* = c + \phi_H P^* \rho^*(1 - \eta_H). \tag{69}$$

This proves Case 1 of the theorem.

Case 2(ii) of the theorem (non-concave region): Now suppose $\phi_H < 1$ and condition equation 6 is violated, i.e.,

$$k\lambda - \frac{(1 - \eta_H)^2(1 - \phi_H)^2}{2\alpha} \leq 0. \tag{70}$$

Then equation 63 is weakly convex (linear if equality holds and strictly convex if the inequality is strict) in ρ. Therefore, the maximizer over the compact set $[0, 1]$ must lie at a boundary point, so

$$\rho^* \in \{0, 1\}. \tag{71}$$

To select between $\rho^* = 0$ and $\rho^* = 1$, compare $\Pi(1, P^*(1))$ and $\Pi(0, P^*(0))$ using equation 62:

$$\Pi(1, P^*(1)) - \Pi(0, P^*(0)) = \lambda\Big[\gamma_v \eta_H - \gamma_u(\eta_u^+ - \eta_u^-) - \frac{k\lambda}{2} + \frac{(1-\eta_H)^2(1-\phi_H)^2}{4\alpha}\Big]. \tag{72}$$

Hence,

$$\rho^* = \begin{cases} 1, & \text{if } \gamma_v \eta_H - \gamma_u(\eta_u^+ - \eta_u^-) - \dfrac{k\lambda}{2} + \dfrac{(1-\eta_H)^2(1-\phi_H)^2}{4\alpha} \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{73}$$

Given ρ^*, the penalty is again pinned down by equation 61:

$$P^* = \frac{\rho^*(1-\eta_H)(1-\phi_H)}{2\alpha}, \tag{74}$$

(which implies $P^* = 0$ when $\rho^* = 0$ and $P^* > 0$ when $\rho^* = 1$), and the reward follows from binding participation:

$$R^* = c + \phi_H P^* \rho^* (1 - \eta_H). \tag{75}$$

This proves Case 2(ii) of the theorem and completes the proof.

B.4 Reverse Screening: Proof of Theorem 3.4

Proof. Fix any mechanism (ρ, R, P). A type $i \in \{H, L\}$ participates if and only if its expected payoff from participation is nonnegative, i.e.,

$$R - c \;\geq\; \phi_i \, P \, \rho \, (1 - \eta_i). \tag{76}$$

Therefore, a *strict* reverse-screening outcome $(p_H, p_L) = (0, 1)$ requires

$$\text{(PC-L)} \quad R - c \;\geq\; \phi_L P \rho (1 - \eta_L), \tag{77}$$

$$\text{(NP-H)} \quad R - c \;<\; \phi_H P \rho (1 - \eta_H). \tag{78}$$

(We implicitly require $P\rho > 0$; otherwise the two right-hand sides coincide and it is impossible to have $p_L = 1$ while $p_H = 0$.)

($\Rightarrow$) Necessity. If a strict reverse-screening equilibrium exists, then there is an R satisfying equation 77 and equation 78. Combining them yields

$$\phi_L P \rho (1 - \eta_L) \;\leq\; R - c \;<\; \phi_H P \rho (1 - \eta_H). \tag{79}$$

Since $P\rho > 0$, dividing through by $P\rho$ gives

$$\phi_L (1 - \eta_L) \;<\; \phi_H (1 - \eta_H), \tag{80}$$

which is equation 22. The ratio form equation 23 follows by dividing both sides of equation 22 by $\phi_L(1-\eta_H) > 0$ (when $\phi_L > 0$; if $\phi_L = 0$, the strict inequality reduces to $\phi_H(1 - \eta_H) > 0$).

(⇐) Sufficiency. Assume equation 22 and $P\rho > 0$. Then the interval

$$\Big[c + \phi_L P\rho(1-\eta_L),\ \ c + \phi_H P\rho(1-\eta_H)\Big) \tag{81}$$

is nonempty. Pick any R in this interval. By construction, (PC-L) holds and (NP-H) holds, so $(p_H, p_L) = (0, 1)$ is implemented. Hence a strict reverse-screening equilibrium exists.

Fix any (ρ, P) that admits strict reverse screening. If R satisfies (PC-L) with slack, i.e., $R > c+\phi_L P\rho(1-\eta_L)$, then lowering R to $\tilde{R} := c + \phi_L P\rho(1-\eta_L)$ preserves low-type participation and weakly decreases high-type incentives to participate, so $(p_H, p_L) = (0, 1)$ is preserved. Since R enters the platform objective as a pure transfer when low types participate, an optimizing platform prefers the smallest feasible reward. Therefore, in any optimal reverse-screening mechanism,

$$R \ = \ c + \phi_L \, P \, \rho \, (1 - \eta_L), \tag{82}$$

which is equation 24. □

B.5 Proof of Theorem 3.5

Under reverse screening, low types participate and high types abstain, so

$$\theta^{F,L} = 1 - \lambda, \qquad \theta^{F,H} = 0,$$

and therefore

$$\theta^{v,L} = \rho(1-\lambda), \qquad \theta^{v,H} = 0, \qquad \theta^{u} = (1-\rho)(1-\lambda).$$

The platform's profit function becomes

$$\begin{aligned}\Pi(\rho, R, P) = \gamma_v \eta_L \rho(1-\lambda) + \gamma_u(\eta_u^+ - \eta_u^-)(1-\rho)(1-\lambda) - \frac{k}{2}\big(\rho(1-\lambda)\big)^2 \\ - R(1-\lambda) + P\rho(1-\eta_L)(1-\lambda) - \alpha P^2(1-\lambda).\end{aligned} \tag{83}$$

In any reverse-screening outcome with $(p_H, p_L) = (0, 1)$, the low-type participation constraint binds:

$$R = c + \phi_L P\rho(1-\eta_L). \tag{84}$$

Moreover, deterring high types requires

$$R \le c + \phi_H P\rho(1-\eta_H), \tag{85}$$

which, using equation 84, is equivalent to the reverse-screening condition $\phi_L(1-\eta_L) \le \phi_H(1-\eta_H)$ in equation 26 (strict inequality yields strict deterrence).

Substituting equation 84 into equation 83 yields the reduced objective

$$\begin{aligned}\Pi(\rho, P) = (1-\lambda)\Big[\gamma_v \eta_L \rho + \gamma_u(\eta_u^+ - \eta_u^-)(1-\rho) - \frac{k(1-\lambda)}{2}\rho^2 - c\Big] \\ + (1-\lambda)P\rho(1-\eta_L)(1-\phi_L) - (1-\lambda)\alpha P^2,\end{aligned} \tag{86}$$

with $\rho \in [0, 1]$ and $P \ge 0$.

Case 2(i) in Theorem 3.5: $\phi_L \ge 1$. When $\phi_L \ge 1$, we have $(1-\phi_L) \le 0$. Hence, for any fixed $\rho \ge 0$, the term $(1-\lambda)\big[P\rho(1-\eta_L)(1-\phi_L) - \alpha P^2\big]$ is weakly decreasing in P on $[0, \infty)$, so $P^* = 0$. With $P^* = 0$, maximizing equation 86 over $\rho \in [0, 1]$ yields

$$\rho^* = \min\left\{1, \max\left\{0, \frac{\gamma_v \eta_L - \gamma_u(\eta_u^+ - \eta_u^-)}{k(1-\lambda)}\right\}\right\}, \tag{87}$$

and $R^* = c$ follows from equation 84. This establishes Case 2(i).

Case 1 and Case 2(ii) in Theorem 3.5: $\phi_L < 1$. When $\phi_L < 1$, we have $(1-\phi_L) > 0$. For any fixed $\rho \in [0,1]$, the reduced profit equation 86 is strictly concave in P, so the unique maximizer satisfies

$$\frac{\partial \Pi}{\partial P} = (1-\lambda)\Big[\rho(1-\eta_L)(1-\phi_L) - 2\alpha P\Big] = 0, \tag{88}$$

giving

$$P^*(\rho) = \frac{\rho(1-\eta_L)(1-\phi_L)}{2\alpha}. \tag{89}$$

Substituting equation 89 into equation 86 yields, up to a ρ-independent constant,

$$\Pi(\rho, P^*(\rho)) = \text{const} + (1-\lambda)\Big(\gamma_v\eta_L - \gamma_u(\eta_u^+ - \eta_u^-)\Big)\rho - \frac{(1-\lambda)}{2}\Big(k(1-\lambda) - \frac{(1-\eta_L)^2(1-\phi_L)^2}{2\alpha}\Big)\rho^2. \tag{90}$$

Case 1 (interior/concave region): If

$$k(1-\lambda) > \frac{(1-\eta_L)^2(1-\phi_L)^2}{2\alpha}, \tag{91}$$

then equation 90 is strictly concave in ρ, and the unconstrained maximizer is

$$\tilde{\rho} = \frac{\gamma_v\eta_L - \gamma_u(\eta_u^+ - \eta_u^-)}{k(1-\lambda) - \frac{(1-\eta_L)^2(1-\phi_L)^2}{2\alpha}}. \tag{92}$$

Imposing $\rho \in [0,1]$ yields the stated ρ^*. Then the expressions for P^* and R^* follow from equation 89 and equation 84. This proves Case 1.

Case 2(ii) (non-concave region): If equation 91 is violated, then equation 90 is weakly convex in ρ, so the maximizer over $[0,1]$ satisfies $\rho^* \in \{0,1\}$. Comparing $\Pi(1, P^*(1))$ and $\Pi(0, P^*(0))$ yields

$$\Pi(1, P^*(1)) - \Pi(0, P^*(0)) = (1-\lambda)\Big[\gamma_v\eta_L - \gamma_u(\eta_u^+ - \eta_u^-) - \frac{k(1-\lambda)}{2} + \frac{(1-\eta_L)^2(1-\phi_L)^2}{4\alpha}\Big], \tag{93}$$

which implies the boundary rule for ρ^*. The expressions for P^* and R^* follow from equation 89 and equation 84. This completes the proof.

B.6 Proof of Proposition 1

Proof. Work under perfect separation, so $\theta^{F,H} = \lambda$ and $\theta^{F,L} = 0$. Under the interior case with $\phi_H < 1$, the reduced objective after substituting the binding high-type participation constraint is

$$\Pi(\rho, P; \lambda) = \lambda\Big[\gamma_v\eta_H\rho + \gamma_u(\eta_u^+ - \eta_u^-)(1-\rho) - c\Big] - \frac{k}{2}(\rho\lambda)^2 + \lambda P\rho(1-\eta_H)(1-\phi_H) - \alpha P^2\lambda. \tag{94}$$

For fixed ρ, the optimal penalty is

$$P^*(\rho) = \frac{\rho(1-\eta_H)(1-\phi_H)}{2\alpha}. \tag{95}$$

Substituting this into the reduced objective gives

$$\Pi(\rho, P^*(\rho); \lambda) = \lambda\big[\gamma_u(\eta_u^+ - \eta_u^-) - c\big] + \lambda N\rho - \frac{\lambda}{2}(k\lambda - D)\rho^2. \tag{96}$$

When the interior verification solution is feasible, the first-order condition gives

$$\rho^*(\lambda) = \frac{N}{k\lambda - D}. \tag{97}$$

The assumption $0 < N < k - D$ implies that $\rho^*(1) \in (0, 1)$. Hence, by continuity, the interior expression is valid for all λ in some neighborhood of 1.

Substituting $\rho^*(\lambda)$ into the objective gives

$$\Pi^*(\lambda) = \lambda\big[\gamma_u(\eta_u^+ - \eta_u^-) - c\big] + \frac{\lambda N^2}{2(k\lambda - D)}. \tag{98}$$

Differentiating with respect to λ yields

$$\frac{d\Pi^*}{d\lambda} = \big[\gamma_u(\eta_u^+ - \eta_u^-) - c\big] - \frac{DN^2}{2(k\lambda - D)^2}. \tag{99}$$

At $\lambda = 1$,

$$\left.\frac{d\Pi^*}{d\lambda}\right|_{\lambda=1} = \big[\gamma_u(\eta_u^+ - \eta_u^-) - c\big] - \frac{DN^2}{2(k - D)^2}. \tag{100}$$

By the stated condition,

$$c > \gamma_u(\eta_u^+ - \eta_u^-) - \frac{DN^2}{2(k - D)^2}, \tag{101}$$

we have

$$\left.\frac{d\Pi^*}{d\lambda}\right|_{\lambda=1} < 0. \tag{102}$$

Since the derivative is continuous in a neighborhood of $\lambda = 1$, there exists some $\bar{\lambda} \in (0, 1)$ sufficiently close to 1 such that

$$\left.\frac{d\Pi^*}{d\lambda}\right|_{\lambda=\bar{\lambda}} < 0. \tag{103}$$

Therefore, $\Pi^*(\lambda)$ is not monotonically increasing on $(0, 1)$. □