Optimal Pricing Schemes in the Presence of Social Learning and Costly Reporting

Kaiwei Zhang[*], Xienan Cheng[†], Xi Weng[‡§]

January 21, 2026

Abstract

A monopolist sells a new product of unknown quality to sequentially arriving, short-lived buyers with unit demand. As it's costly for buyers to report their realized utility, the monopolist must offer a bonus to incentivize disclosure. We work within an exponential-bandit framework which allows the arrival of both conclusive good and bad news, and characterize the monopolist's optimal joint policy of dynamic pricing and bonus, which switches among distinct phases. In the special case of no bad conclusive news, the monopolist will set zero bonus and halt learning when public belief is sufficiently high, but doing so is generally never optimal. Moreover, a polarity dependent manipulation–paying bonus only for verifiable good news–can improve social welfare when the arrival rate of conclusive bad news is close to zero.

Keywords: Social Learning; Costly Reporting; Dynamic Pricing.
JEL. D83. C02. C61. C73.

[*]Department of Economics, Penn State University; kqz5253@psu.edu
[†]Guanghua School of Management, Peking University; xienancheng@gsm.pku.edu.cn
[‡]Guanghua School of Management, Peking University; wengxi125@gsm.pku.edu.cn.
[§]We would like to thank conference participants at the UIBE Workshop on Industrial Organization and Competition Policy (October 2023), Zhejiang University Conference on Digital Economy and Micro Theory (July 2023), Stony Brook Game Theory Conference (July 2023), Penn State Theory Workshop (Sep 2023), and Pennsylvania Economic Theory Conference (May 2025) for their helpful comments and suggestions. Zhang acknowledges support from the Ed Green Graduate Fund. Weng acknowledges support from the National Natural Science Foundation of China (Grant Nos. 72192843, and 72225001).

1 Introduction

The operation of new markets for experience goods depends on well-functioning feedback mechanisms that allow consumers to infer product quality through social learning over time. Online platforms have made it feasible and common for consumers to leave feedback after purchasing and experiencing products. However, the supply of reviews remains insufficient due to the costs associated with providing feedback. The marketing literature shows that most individuals tend to be "lurkers" who read reviews only from a small population of posters but do not contribute themselves (Moe and Schweidel, 2012). For example, only 1–2% of buyers write reviews on Amazon for the products they purchase (Kenjiroi, 2021); even in one-to-one matching markets, Li and Xiao (2014) find that "only about 50% of buyers leave feedback after transactions on eBay." An online Consumer Review Survey (BrightLocal, 2025) asks: "[W]hat's the main reason holding consumers back from writing them? Almost a quarter of US adults feel that the overall business experience was just not noteworthy enough to write about."

Consequently, review bonuses are routinely offered to incentivize consumers to provide honest feedback after purchasing experience goods. The empirical evidence (Li, 2010; Li and Xiao, 2014; Miller et al., 2005) demonstrates that (1) rebates increase the likelihood that buyers will submit reviews, and (2) monetary incentives do not compromise the honesty of reports. Consumer feedback also facilitates social learning and allows sellers to implement dynamic pricing strategies in experience goods markets, such as video games, resulting in higher revenues and profit margins in pilot categories (McKinseyCompany, 2024). Despite the dual nature of these initiatives, in the existing literature, dynamic pricing (see, e.g., Bergemann and Välimäki, 1996, 2000, 2006; Hagiu and Wright, 2020; Weng, 2015) and review incentives (see, e.g., Li, 2010; Li et al., 2016; Miller et al., 2005) are typically studied in isolation. The question of how a profit-maximizing monopolist should jointly determine dynamic pricing and review bonuses remains largely unexplored. In this article, we present a simple model in which the monopolist employs the bonus as a tool to manage the flow of public information (social learning); this use is distinct from its pricing strategy, which governs demand.

We study a social learning problem faced by a long-lived monopolist selling a risky product, the utility of which is unknown for all parties. A safe alternative with known utility is also available from a competitive market, rendering its price exogenous. We model the risky product using an extended version of the binary-state exponential-bandit model established by Keller et al. (2005). The quality of the risky product is either good or bad. In a good state, the product may yield a lump-sum payoff G with arrival rate λ_G while in a bad state, the product may yield a lump-sum damage B with arrival rate λ_B. A shared belief is formed about the quality of the risky product based on public reports from past buyers: a positive lump-sum payoff is considered conclusive good news, whereas a negative lump-sum damage is considered conclusive bad news. Without the arrival of conclusive news, buyers gain flow utility z and depending on whether $\lambda_B > \lambda_G$ or not, the public belief about quality may improve or deteriorate. Obviously, our model accommodates the classical good news model of Keller et al. (2005) as a special case with $\lambda_B = 0$, and allows more general case of $\lambda_B > 0$.

Buyers incur a private cost if they choose to report, representing the effort required to publish an online review. We allow the monopolist to offer a bonus that is not contingent on the content of the report, but can vary over time with public belief. Throughout this paper, we examine the scenario in which the upper bound on the reporting cost (denoted by $\bar{c}$) is so high that it is unprofitable for the monopolist to offer a bonus equal to $\bar{c}$ to induce full disclosure of information (see Assumption 3 in Section 3.1). The price and bonus for each period are set before the buyer makes their purchase and reporting decision. As buyers are short-lived in our model, their reporting behavior is consistent with the empirical findings: after purchasing the experience good, buyers provide a review if and only if the bonus offered exceeds the cost of reporting, and, once buyers report their experiences, they do so honestly.

Obviously, the monopolist's bonus decision is strategically intertwined with its pricing decision. For example, a high bonus can alter the purchasing decisions of agents with low reporting costs, who may be induced to buy the risky product solely to claim the reward. The monopolist must thus optimize these two instruments jointly. We show that it is without loss of generality, to analyze a pricing rule that leaves non-reporting agents indifferent between the two products (fair pricing). The monopolist's strategy then comes down to jointly setting the bonus and a tie-breaking rule for these indifferent agents. This result enables us to focus on the dynamic bonus-setting, which is non-trivial on its own: the public belief remains unchanged if the agent either purchases the safe product or purchases the risky product but does not report their experience. This latter possibility provides the monopolist with a strategic lever: it can effectively halt learning at any time, even though it is selling a risky product, simply by setting the bonus to zero. Interestingly, we show that in the general case of $\lambda_B > 0$, it is never optimal for the monopolist to use this no bonus strategy. The reason comes from the fact that fair pricing can internalize the arrival of good news but cannot internalize the arrival of bad news. This is because there is an optional value from the arrival of bad news that stops experimentation and avoids the potential loss of selling the risky product at a low price. As long as the probability of receiving conclusive bad news is strictly positive, it's never optimal to set zero bonus as this leads to a loss of the optional value.

For the general case of $\lambda_B > 0$, we characterize the monopolist's optimal policy, which follows a threshold structure based on the public belief about the product's quality, as consisting of three distinct phases along the learning path: (1) When the belief is sufficiently high, the monopolist sells the risky product while actively testing by offering a positive bonus (Full Coverage). Interestingly, in this region, the bonus is decreasing in belief. (2) Once the belief drops to a point where the risky product is myopically unattractive, the forward-looking monopolist continues to experiment. Rather than setting negative prices as in Bergemann and Välimäki (2000), our monopolist uses a novel Partial Coverage strategy. By setting a fair price that makes non-reporting agents indifferent between two products, a positive bonus induces only agents with low reporting costs to purchase the risky product, thereby subsidizing experimentation efficiently. (3) If the belief is lower than a stopping threshold, the monopolist optimally chooses to quit the market (Outside Option).

It turns out that depending on whether $\lambda_G > \lambda_B$ or not, the stopping threshold is solved differently. This is because whether $\lambda_G > \lambda_B$ or not creates a fundamental difference in the learning dynamics, since it determines the direction of smooth belief movement when no conclusive news is reported. When good news arrives more frequently, starting from a high initial belief, without realization of conclusive good news, the belief would drop so the seller faces an optimal stopping decision when considering switching from Partial Coverage to Outside Option downwards. This generates value matching and smooth pasting conditions at the stopping threshold as in the standard literature (see, e.g., Keller et al., 2005). However, when bad news arrives more frequently, starting from a low initial belief, the seller is facing an optimal starting decision about when switching from Outside Option to Partial Coverage upwards. The value function need not be smooth at the threshold. In Section 4.2, we develop different boundary conditions and solution methods depending on whether $\lambda_G > \lambda_B$ or not.

The stopping threshold partially reflects whether the learning process led by the monopolist achieves a socially efficient level. One might expect the monopolist to quit the market before reaching this level of efficiency because the difference between the offered bonus and the reporting cost imposes a profit loss, reducing its willingness to experiment. This insight is true when $\lambda_B > \lambda_G$. However, when $\lambda_B < \lambda_G$, the monopolist quits the market on reaching the efficient belief, but the learning process is inefficiently slow. This efficient stopping occurs because, when we have smooth pasting at the quitting threshold, the monopolist can always use the Partial Coverage strategy by setting a bonus close to zero; it is possible to stay in the market for a small chance of good news with a potential cost close to zero. We make the speed comparison concrete by introducing a social planner who can commit to an allocation and reporting rule based on the prevailing public belief and each agent's reported cost. The social planner is essentially solving a dynamic mechanism-design problem, where the reporting threshold chosen by the planner plays the role of the bonus in the monopolist setting for any given public belief. We show that this reporting threshold is always slightly higher than the bonus in the monopolist's setting, meaning that the social planner will encourage greater experimentation at all belief levels because the cost of doing so is lower.

Next, we consider two extensions of our benchmark scenario. In the first case, we let $\lambda_B = 0$ and consider the special case of good news model. In this case, it's indeed optimal for the monopolist to use the no bonus strategy when the belief is sufficiently close to one. This result confirms out previous intuition that the probability of receiving conclusive bad news plays an important role in shaping the monopolist's dynamic bonus-setting. When determining the bonus, the monopolist has to balance the potential gain from an additional report, which may reveal conclusive news and change future profits—against the risk of receiving no conclusive signal, which may lead to a smooth deterioration of the belief. The absence of conclusive bad news decreases the potential gain and hence makes the monopolist more willing to offer zero bonus and halt learning when the belief is sufficiently close to one.

In the second extension of the benchmark scenario, we consider a situation where the monopolist can make the reporting process polarity dependent; that is, the bonus is paid only when the review constitutes verifiable good news (Milgrom, 1981). In this scenario, the absence of a review causes the public belief in the risky product quality to diminish as in the good news model, because this implies that no good conclusive news arrives. We find that this polarity dependent manipulation has both positive and negative effects on social welfare. The positive effect mainly comes from the fact that the bonus is paid less frequently, and, as a result, the monopolist's incentive to learn is stronger, leading to more aggressive experimentation and higher social welfare. However, there is another negative effect due to pooling of bad conclusive news and no conclusive news, essentially making $\lambda_B = 0$. Therefore, the net impact on social welfare relies on how large λ_B is. In particular, if λ_B is sufficiently close to zero, the positive effect dominates the negative one, which implies that verifiable manipulation would increase social welfare. This result differs from those in the literature (e.g., Li (2010), Li and Xiao (2014)), in which manipulation is treated as a main challenge of implementing "rebate for review" systems. By showing that *verifiable* manipulation improves efficiency, we also provide a validation of the online platforms' practice to encourage buyers to include videos or photos in their reviews.

The remainder of this article is organized as follows. We review the literature in Section 2. In Section 3, we introduce the benchmark model and solve this in Section 4. We conduct welfare analysis in Section 5 and investigate two extensions of the benchmark model in Section 6. Section 7 concludes the paper.

2 Literature Review

Our model is situated within the dynamic pricing literature. Bose et al. (2006) and Bose et al. (2008) analyze dynamic pricing within a herding context (Banerjee, 1992; Bikhchandani et al., 1992), where each agent receives a private signal related to the state before deciding. In our model, each agent possesses private information (the reporting cost) that is independent of the state of the world. The aggregated information about the state is reflected in the public belief. Agents observe the utility realized following their decision and then determine whether to report. This distinguishes our model from the herding model. Regarding the information structure, our work is more aligned with (Kremer et al., 2014). However, unlike the model in Kremer et al. (2014), in which the planner can exploit information asymmetry to encourage exploration by agents, we focus on public reports. Therefore, only transfers can compensate for the reporting cost.

Agents in our model must choose between two arms, one safe and one risky, which is modeled by a bandit where agents are involved sequentially (Bolton and Harris, 1999; Keller et al., 2005; Rothschild, 1974; Weng, 2015). In multi-agent settings with public information, the positive externality of information often leads to free-rider problems, and social learning is always terminated inefficiently (Bergemann and Välimäki, 1996, 2000, 2006; Hagiu and Wright, 2020). In these studies, the termination belief serves as a comprehensive measure of efficiency as individual utility outcomes are publicly observable, ensuring that informative signals are consistently disclosed. Nevertheless, in our model, the seller must incentivize buyers to provide reports through monetary transfers; hence, informative signals may be hold as a consequence of relatively high effort (the latter one is uninformative about the state of the world). We use the seller's signal demand as a measure of learning speed and show that this demand is decreasing with respect to the distance between current belief and "the most confused belief". Zhong (2022) observes the same conclusion in a rational inattention setting.

The literature on review motivation is also closely related to our work. Although there is no doubt that online reviews help consumers decrease the uncertainty of purchasing certain goods, the current provision of reviews is clearly insufficient (Li and Xiao, 2014). Buyers may provide a review only when the incentive outweighs the costs involved, which may be endogenous because of the externality of information (Avery et al., 1999; Kremer et al., 2014), or exogenous, based on the time and effort required (Li, 2010; Miller et al., 2005). Reporting costs may lead to dishonest reporting if all agents are compelled to provide reviews (Krishna and Morgan, 2012). Our approach, by contrast, involves monetary incentives to offset agents' reporting costs, facilitating truthful reporting for effective social learning.

Several empirical studies have shown that monetary incentives effectively encourage the desired online behavior, especially that of writing reviews on online platforms (Burtch et al., 2018; Cabral and Li, 2015; Fradkin et al., 2015; Khern-am-nuai et al., 2018). These studies suggest that consumers are more likely to write reviews when they receive adequate compensation for their time and effort. Research by Li and Xiao (2014) indicates that this compensation is not necessarily understood by buyers as constituting bribery and does not affect their reporting honesty. Xu et al. (2015) point out the difficulty of detecting sellers' manipulation behaviors, such as offline ads incentivizing positive reviews only. The insights from these empirical studies strongly support our understanding of the purchasing and reporting behaviors of agents. Our analysis also suggests that allowing manipulation can enhance efficiency and, in addition to detecting sellers' manipulation, online platforms can act to enhance the credibility of positive reviews.

3 Model Setup

3.1 Environment

There are two perfect substitutes on the market. One is considered a "safe arm" as it gives a commonly known flow utility $s > 0$. The other is categorized as a "risky arm" and its utility is unknown to buyers and the seller. Production costs are normalized to zero for both. The safe arm is supplied in a perfectly competitive market; hence, its price is always 0 in equilibrium. There is a monopolist seller selling the risky arm. There is also a sequence of risk-neutral buyers (or agents) in an infinite horizon continuous-time setting, each with a lifetime of $\Delta \to 0$ periods. A single agent is present in each period. The long-lived monopolist aims to maximize its discounted profit over time, and short-lived agents, having outside utilities of 0, seek to maximize their expected utilities within period Δ by selecting at most one product between the available options. Throughout this article, we refer to the monopolist as the seller and the buyers born at time t as agent t.

The utility to the buyer of choosing the risky arm is characterized by an exponential bandit. There are two states, good ($\theta = g$) and bad ($\theta = b$). In state g, the risky arm can yield either a lump-sum payoff of g (common knowledge, good news) with commonly known arrival rate λ_G (i.e., the probability density function of the arrival time is $\lambda_G e^{-\lambda_G t}$) or a constant payoff z with the remaining probability. Similarly, in state b, the risky arm can yield either a conclusive bad news with lump-sum payoff of $-B$ or constantly flow payoff z, and the arrival rate of bad news is λ_B. Notice that our structure degenerates to standard conclusive bad or good news model by setting $\lambda_G = 0$ or $\lambda_B = 0$, but our analysis applies to all settings with $\lambda_B \geq 0, \lambda_G \geq 0$. We also assume $\min\{G, B, z\} \geq 0$. Additionally, there is a common discount factor r; that is, a payoff v received at time t generates a utility of $e^{-rt}v$ at time 0 for both the buyer and seller.

We make Assumption 1 to ensure that the learning of the unknown state of risky arm is meaningful to consider.

Assumption 1. $z_G \equiv z + \lambda_G G > s > z_B \equiv z - \lambda_B B$

After purchasing the risky arm, buyers privately learn their realized utility: either they observe a constant flow payoff of z, or a conclusive good news (at least one realization of G), or a conclusive bad news (at least one realization of B) [1]. Buyers then decide whether to disclose their information. In our main text, we simplify consumers' disclosure as a public and credible disclosure: once they report, this is publicly available to the seller and subsequent consumers.[2] We assume they would report truthfully because they are short-lived.

However, the act of reporting incurs a cost: with each buyer having a private reporting cost of c, independently and identically distributed according to the distribution function $H(\cdot)$ with support $[0, \bar{c}]$. $\bar{c} \in \mathbb{R}_+ \cup \{+\infty\}$ and $H(\cdot)$ known to the seller. For simplicity, we assume that $H(\cdot)$ is strictly increasing and has a smooth density function $h(\cdot) > 0$ for all $x \in [0, \bar{c}]$. Moreover, we impose the following standard Assumption 2 on the density function $h(\cdot)$ and virtue-value $\beta(x) = x + \frac{H(x)}{h(x)}$. We impose Assumption 3 to simplify the analysis in our main model, and drop it in Section B.1.

Assumption 2. *$\beta(x)$ is continuously differentiable and $\beta'(x)$ is always positive.*

We want to assume out the corner solution when setting bonus equal to $\bar{c}$ is optimal. Intuitively we need a large $\bar{c}$ to make sure doing so is too expensive. We consider the case with small $\bar{c}$ in Appendix B.1. A sufficient and necessary condition is $\bar{c} \geq b(\alpha_{PC}^{FC})$ where right hand side is defined in Section 4. We also want to emphasize that Assumption 3 guarantees that setting $b = \bar{c}$ is dominated by quitting the market, hence a sufficient but not necessary condition to rule out setting $b = \bar{c}$. This sufficient condition implies the following primitive representation:

Assumption 3. $\bar{c} > z_B - s + \lambda_G G + \lambda_B B + \lambda_G \frac{z_G - s}{r}$

At time $t + \Delta$, agent t leaves the market and a new agent $t + \Delta$ arrives, so there is only one agent in the market at any time. At time t, agent t observes all history before t. If purchasing for arm $i \in \{r, s\}$ happens, the prices p_{it} are paid at the *beginning* of this period. If agent t decides to report, they pay a lump-sum reporting cost $c_t\Delta$ and obtain a lump-sum bonus $b_t\Delta$ at the *end* of period Δ.

[1] Multiple realizations of lump-sum payoffs are of higher-order terms of $\Delta \to 0$.

[2] We consider the case where the seller can partially manipulate the disclosure in another separate note. Kremer et al. (2014) considers the similiar problem.

Timeline We describe the timeline within period $[t, t+\Delta)$ as follows:

1. At time point t, agent t learns his reporting cost $c_t \sim H(\cdot)$, and then choose between the risky arm and safe arm given price p_{rt} and bonus b_t.

2. In $(t, t+\Delta)$, if risky arm is chosen, then there may be some time $t_0 \in [t, t+\Delta)$ such that the conclusive good news or bad news arrives.

3. At time point $t+\Delta$, the consumer decides whether to report his experience. If reporting happens, he pays cost $c_t\Delta$ and receives bonus $b_t\Delta$ at time point $t+\Delta$.

4. Agent t leaves the market, and agent $t+\Delta$ arrives at $t+\Delta$.

Discussion for the model set up Here we assume the agent t learns his own reporting cost c_t before making the purchase decision or knowing his own utility realization, specifically to capture the idea that some buyers are busy and unwilling to fill the feedback invitations sent to him but others are not[3]. Also, the bonus is announced at time point t and will be dispensed as long as report is made at $t+\Delta$. In this main context we do not allow bonus to be contingent on the polarity of the feedback, capturing the "Rebate for feedback" mechanisms played by real life e-commerce platforms. In reality it's also seen that sellers bribe the buyers to give positive reviews only, we consider this extension in Section 6.2. We also assume information symmetry that the seller doesn't have private information about the quality of the risky arm: we want to focus on the incentives to learn instead of the potential problem that high quality sellers signal by setting high bonus (Li and Xiao, 2014); another defense is that the seller's strategy is committed and publicly observed before the game starts and we study the pooling equilibrium. Finally, we focus on the sellers who can affect the price and set rewards for feedbacks simultaneously, which is different from the current Amazon mode where sellers only set prices and platform interacts with consumers to motivating feedbacks. We try to model the commonly seen behavior where sellers ask for feedback after the purchase to learn quality of their new product with rewards like gift cards or membership points [4].

3.2 Seller's Policy

Denote the common history of pricing, bonus, purchasing and reporting from past observations before time t as $\mathscr{F}(t)$. Both the seller and buyer t will collectively form a common state variable, which is the posterior belief $\alpha_t = \Pr(\theta = g|\mathscr{F}(t))$. We further focus on pure strategies, and limit our attention to Markovian strategies (with single state variable) and equilibrium. That is, the seller can commit to the state variable contingent (given no news

[3] Empirical literatures show that consumers may have higher incentives to write reviews when they have extreme experiences, but this dependnece is more sensitive to the context, hence we choose not to model the correlation between the realized utility and reporting cost.

[4] Examples are Weis(https://weismarkets.survey.marketforce.com/?selection=1), Adidas(https://www.adidas.com/us/help/us-adiclub/terms-and-conditions-adiclub) and SHEIN (https://roe.shein.com/Review-Guidance-a-478.html)

revealed before t) bonus scheme $b : [0, 1] \to [0, \bar{c}]$ and pricing scheme $p : [0, 1] \to \mathbb{R}$ before the game starts, where $b(\alpha)$ is the bonus for reporting news from risky arm, and $p(\alpha)$ is the price for the risky arm. We will further show that the seller's pricing strategy can be implemented equivalently with a binary tie-breaking policy $\ell(\alpha_t)$, together with the bonus scheme $b(\alpha_t)$ in this section. We use p_t, b_t for $p(\alpha_t), b(\alpha_t)$.

If agent t purchases the safe arm, he would have surplus $S = \int_0^\Delta e^{-rt} s dt = \frac{1-e^{-r\Delta}}{r} s$. If he chooses the risky arm he needs to further decide whether to report (denoted as $r_t = 1$) or not ($r_t = 0$). In equilibrium, $r_t = \mathbb{I}(b_t > c_t)$ where c_t is agent t's reporting cost and $\mathbb{I}(E)$ is an indicator function of event E. At time τ, if agent τ buys the risky arm with the above equilibrium report decision r_τ, his expected payoff is

$$\int_0^\Delta e^{-rt} \left[z + \alpha \lambda_G e^{-\lambda_G t} G - (1-\alpha) \lambda_B e^{-\lambda_B t} B \right] dt - p_\tau + r_\tau e^{-r\Delta} (b_\tau - c_\tau) \Delta.$$

Now we simplify the seller's pricing problem. We first ignore the term $r_\tau e^{-r\Delta}(b_\tau - c_\tau)\Delta$ and set a "fair price" which makes agent τ indifferent between the risky arm and the safe arm without considering the payoffs from reporting:

$$\begin{aligned} p_\tau(\alpha) \equiv p^*_\Delta(\alpha) = & \int_0^\Delta e^{-rt} \left[z + \alpha \lambda_G e^{-\lambda_G t} G - (1-\alpha) \lambda_B e^{-\lambda_B t} B \right] dt - S \\ = & \frac{\alpha \lambda_G (1 - e^{-(r+\lambda_G)\Delta})}{r + \lambda_G} G - \frac{(1-\alpha)\lambda_B (1 - e^{-(r+\lambda_B)\Delta})}{r + \lambda_B} B - \frac{1 - e^{-r\Delta}}{r} (s - z) \end{aligned} \tag{Fair Pricing}$$

Under Equation (Fair Pricing), agent τ with reporting cost $c_\tau < b_\tau$ would get positive utility from reporting and purchase the risky arm for sure. If reporting cost $c_\tau > b_\tau$, agent τ would choose not to report and hence be indifferent between the risky arm and the safe arm. We introduce a tie-breaking policy $\ell : [0, 1] \to \{s, r\}$ in this situation where $\ell(\alpha) = r$ would guide all indifferent buyers to choose the risky arm, and s for the safe arm.[5] When $b > 0$, we call the case of $\ell(\alpha) = r$ *Full Coverage* as all of the market demand is served by the seller; and the case of $\ell(\alpha) = s$ *Partial Coverage* as the seller has a positive probability to sell a risky arm when the bonus is positive.[6]

Next we show that any of the seller's pricing strategies can be implemented equivalently with ℓ_t and p_t. For ($p_t = p^* + \delta, b_t$) with $\delta > 0$, we can equivalently set ($\tilde{p}_t = p^*, \tilde{b}_t = b_t - \frac{\delta}{e^{-r\Delta}\Delta}$) and $\ell_t = s$, as the price and bonus are perfectly substitutable and all indifferent agents will purchase the safe arm under $p_t = p^* + \delta$. Similarly, for ($p_t = p^* - \delta, b_t$) with $\delta > 0$, we can equivalently set ($\tilde{p}_t = p^*, \tilde{b}_t = b_t + \frac{\delta}{e^{-r\Delta}\Delta}$) and $\ell_t = r$.

With the above argument, we can fix the seller's price to be the fair price $p^*(\alpha; \Delta)$, and model the seller's policy as a combination of bonus scheme $b(\alpha)$ and tie-breaking policy

[5] We can follow the standard argument to implement the tie-breaking policy $\ell(\alpha)$. For example, by setting $p = p^*(\alpha; \Delta) - \epsilon$ with $\epsilon \to 0^+$, the seller can leave a positive surplus $S + \epsilon$ for buyers, and hence guarantee that all buyers would choose the risky arm. On the contrary, when the price is set at $p = p^*(\alpha; \Delta) + \epsilon$ with $\epsilon \to 0^+$, all buyers who will not report would buy the safe arm.

[6] Theoretically speaking, the tie-breaking policy can also be random: the seller may sell to the indifferent agents with probability x. But as shown by Keller et al. (2005), the seller's maximization problem is linear in the probability x, and hence the optimal solution is bang-bang.

$\ell(\alpha)$, both contingent on common belief α. The bonus scheme $b : [0,1] \to \mathbb{R}$ is a measurable function that gives the bonus the seller would pay to the buyer under common belief α, if the buyer reports his utility of purchasing the risky arm. The tie-breaking policy $\ell : [0,1] \to \{s,r\}$ is a measurable function that guides an indifferent buyer's choice between the risky arm and safe arm. Given the fair pricing formula, agent τ's purchasing decision $a : [0,\bar{c}] \times \{s,r\} \times [0,\bar{c}] \to \{r,s\}$ depends on the seller's action profile (ℓ, b) and the his own reporting cost:

$$a(b,\ell,c_\tau) = \begin{cases} r & \text{if } b > c_\tau \\ r & \text{if } \ell = r, b \leq c_\tau \\ s & \text{if } \ell = s, b \leq c_\tau \end{cases} \tag{1}$$

After purchase, buyers' reporting strategy $r : [0,\bar{c}]^2 \to \{0,1\}$ is a simple indicator

$$r(b, c_\tau) = \mathbb{I}(b > c_\tau) \tag{2}$$

As a result, by setting bonus b, the ex-ante probability of successfully motivating the reporting is $H(b)$. Correspondingly, the seller's stage payoff within $[\tau, \tau + \Delta)$ is given by:

$$\tilde{\pi}_\Delta(b,\ell,c_\tau) = \begin{cases} p^*_\Delta(\alpha) - e^{-r\Delta} b\Delta & \text{if } b > c_\tau \\ p^*_\Delta(\alpha) & \text{if } \ell = r, b \leq c_\tau \\ 0 & \text{if } \ell = s, b \leq c_\tau \end{cases} \tag{3}$$

3.3 Seller's Value Function

We first pin down some corner cases, given the price is set by (Fair Pricing). If the state is known to be $\theta = b$, it's optimal for the seller to quit the market and get a value function of 0. If it's known to be $\theta = g$, the value function is $\tilde{V} = \frac{z+\lambda_G G - s}{r}$. We denote $z_G = z + \lambda_G G$ and $z_B = z - \lambda_B B$ for the expected stage payoff under stage g and b. We define

$$\gamma_B(\alpha, \Delta) = (1-\alpha)(1 - e^{-\lambda_B \Delta})$$

$$\gamma_G(\alpha, \Delta) = \alpha(1 - e^{-\lambda_G \Delta})$$

as the probability of realizing a bad/good news within period Δ when common prior of good state is α, and denote $\lambda(\alpha) = \alpha\lambda_G + (1-\alpha)\lambda_B$ as the expected arrival rate for some conclusive news. We then have

$$\frac{\gamma_B}{\Delta} \to (1-\alpha)\lambda_B, \frac{\gamma_G}{\Delta} \to \alpha\lambda_G$$

Consider what happens to agent t within period $[t, t+\Delta)$, when the common belief is $\alpha_t = \alpha$. There are 4 contingencies with respect to reported information:

- With probability $H(b)\gamma_B(\alpha, \Delta)$ a conclusive bad news is reported hence belief drops to 0 immediately.
- With probability $H(b)\gamma_G(\alpha, \Delta)$ a conclusive good news is reported hence belief jumps to 1 immediately.

- With probability $H(b)[1-\gamma_B(\alpha,\Delta)-\gamma_G(\alpha,\Delta)]$ a constant flow payoff z is reported hence belief moves smoothly to α' which is defined later. Notice here no lump-sum payoff is informative: if $\lambda_G > \lambda_B$, then the absence of conclusive news is bad news hence $\alpha' < \alpha$, and vice versa. We can calculate the updated belief α' when no conclusive news is reported within period Δ via Bayes' rule:

$$\alpha' = \frac{\alpha e^{-\lambda_G \Delta}}{\alpha e^{-\lambda_G \Delta} + (1-\alpha)e^{-\lambda_B \Delta}}$$

- With probability $1-H(b)$ no report is made, hence belief is still α.

We now establish the HJB equation based on above four contingencies. Denote the value function of the seller as $\Pi(\alpha)$ when the common belief is α, we have the following HJB equation:

$$\begin{aligned}\Pi_\Delta(\alpha) = &\max_{\ell,b} \pi_\Delta(b,\ell) + e^{-r\Delta}H(b)\gamma_B(\alpha,\Delta)\cdot 0\\ &+ e^{-r\Delta}H(b)\gamma_G(\alpha,\Delta)\cdot\frac{z_G - s}{r}\\ &+ e^{-r\Delta}H(b)[1-\gamma_B(\alpha,\Delta)-\gamma_G(\alpha,\Delta)]\Pi_\Delta(\alpha')\\ &+ e^{-r\Delta}[1-H(b)]\Pi_\Delta(\alpha)\end{aligned} \tag{4}$$

where $\pi_\Delta(b,\ell) = \mathbb{E}_{c\sim H(\cdot)}\tilde{\pi}_\Delta(b,\ell,c)$. We further define $\pi(b,\ell) = \lim_{\Delta\to 0}\frac{\pi(b,\ell)}{\Delta}$ and $\Pi(\alpha) = \lim_{\Delta\to 0}\Pi_\Delta(\alpha)$. We show in the Appendix A.1 that above equation can be simplified to:

$$r\Pi(\alpha) = \max_b[\max_\ell \pi(b,\ell)] + H(b)\left[\lambda_G\alpha\tilde{V} + \alpha(1-\alpha)(\lambda_B-\lambda_G)\Pi'(\alpha) - \lambda(\alpha)\Pi(\alpha)\right] \tag{5}$$

We define the equilibrium to conclude the model setup:

Definition. *(Markov Perfect Equilibrium) A Markov Perfect Equilibrium constitutes: (i) Buyers' purchasing strategy and reporting strategy given by Equations (1) and (2), respectively; and (ii) seller's strategy (ℓ, b) which solves seller's problem defined in (5).*

4 Model Analysis

4.1 Preliminary Analysis

In the main context, we assume a large $\bar{c}$, so in a Markov perfect equilibrium, the bonus chosen is always less than $\bar{c}$. However, there is another corner case in which the bonus is set to zero. There are thus four possible combinations of the selling policy and bonus scheme, as shown in Table 1.

within $[t, t+\Delta)$:	$\ell(\alpha_t) = r$	$\ell(\alpha_t) = s$
$\bar{c} > b(\alpha_t) > 0$	Full Coverage	Partial Coverage
$b(\alpha_t) = 0$	Non Bonus	Outside Option

Table 1: Seller's Four Possible Strategies in the Main Context

Notice that $\mathbb{I}(\ell = r)$ doesn't appear in the $H(b)$ term in equation (5). That is, the difference between setting $\ell = r$ (Full Coverage) or $\ell = s$ (Partial Coverage) only arises in the stage payoff term $\max_\ell \pi(b, \ell)$, while whether bonus is positive or zero only affects the continuation term. If the seller sets $b = 0$, then no report will be made hence the belief won't be updated. Together with $\ell = s$, the seller actully quits the market and the market demand is served by the safe arm, we call this strategy *Outside Option*. If the seller sets $b = 0$ and $\ell = r$, then all indifferent buyers will purchase the risky arm but no report will be made, we call this strategy *Non Bonus*. We call strategy $L \in \{OO, NB, FC, PC\}$ according to the combination of (ℓ, b) in Table 1.

By setting $b_\tau > 0$ and $\ell = r$, agent τ will always buy the risky arm and report their utility if the bonus covers their cost, with ex ante probability $H(b_\tau)$. In this case, the seller uses the risky arm to fully cover the market demand, hence we denote this as the "Full Coverage" strategy (FC).

By setting $b_\tau > 0$ and $\ell = s$, the positive bonus now may affect the market demand. Notice that given (Fair Pricing), purchasing the risky arm and report would generate an additional expected utility of $(b_\tau - c_\tau)\Delta e^{-r\Delta}$ for agent τ. Thus, for agent whose reporting cost c_τ is less than the bonus b_τ, these agents would purchase the risky arm and report for sure. For other agents whose reporting cost c_τ is higher than the bonus b_τ, since they won't report, they are indifferent in purchasing the risky arm or the safe arm, hence they choose the safe arm according to the tie-breaking rule $\ell = s$. We call this strategy "Partial Coverage" (PC), in the sense that the seller is still learning the quality of the risky arm, but only partially covering the market demand.

Note that a report will be generated by agent τ if and only if $b_\tau > c_\tau$, regardless of whether $L = PC$ or FC: given an identical bonus scheme, agents who would report under the Full Coverage strategy will do so if, and only if, they would also report under the Partial Coverage strategy.

4.2 Seller's Optimal Strategy

Now we solve seller's problem defined in (5) and characterize the optimal strategy in this section. We denote $M(\alpha) = z + \alpha\lambda_G G - (1-\alpha)\lambda_B B$ as the expected stage payoff of the risky arm when belief is α.

Lemma 4.1 states one of the above four strategies defined in Table 1 is weakly dominated unless $\lambda_B = 0$. We show that when the value function takes the linear form given by NB, $b = 0$ doesn't maximize the value function. The suboptimality of NB is not obvious ex ante, and we explain the intuition after we investigate the $\lambda_B = 0$ case in Section 6.1.

Lemma 4.1. *NB is suboptimal unless $\lambda_B = 0$, that is, $\Pi(\alpha) \geq \frac{M(\alpha)}{r}, \forall \alpha \in [0,1]$. Moreover, $\Pi(1) = \frac{M(1)}{r}$*

Lemma 4.2 is intuitive, since the stage payoff when using OO is 0, and the stage payoff when using PC should also be lower than that of FC, since the seller chooses to use risky arm to cover only part of the market under PC. We state the solution of special case $\lambda_B = 0$ in section 6.1. In section 4 we focus on the $\lambda_B > 0$ case.

Lemma 4.2. *When $\lambda_B > 0$, the optimal strategy order is always $OO \rightarrow PC \rightarrow FC$. Seller's value function $\Pi(\alpha)$ is continuous, strictly increasing in PC and FC region, and convex.*

We then specify the optimal switching conditions given the strategy order stated in Lemma 4.2. For notational convenience, we use $\alpha_L^{L'}$ for the switching threshold where strategy $L(L')$ is used on the left(right) small neighborhood, and say a function is smooth if it has continuous derivatives.

Whether $\lambda_G > \lambda_B$ or not creates a fundamental difference in the learning dynamics, since it determines the direction of smooth belief movement when no conclusive news is reported. Intuitively, when good news arrives more frequently, no conclusive good news is bad news, and after such a situation, the common belief drop downwards smoothly. However, when bad news arrives more frequently, no conclusive bad news is good news, and the belief moves upwards smoothly. In the first case, starting from a high initial belief, without realization of conclusive good news, the belief would drop so the seller faces a quitting decision when considering switching from PC to OO downwards. However, in the second case, starting from a low initial belief, the low initial belief may rise so the seller faces an entering decision when switching from OO to PC upwards.

This difference determines the condition of optimal swicthing between OO and PC. In the first case, the smooth downward movement of belief generates a smooth pasting condition between PC and OO: starting from a high initial belief, along the learning path, either conclusive good news brings the belief to 1 or 0 immediately, or the belief goes down smoothly. The seller's quitting decision can be smoothed out via the bonus scheme at the switching cutoff α_{PC}^{OO}. This feature guarantees the value function Π is smooth around α_{PC}^{OO} when $\lambda_G > \lambda_B$. However, in the second case, the order is reversed and the seller faces an optimal starting problem (i.e., switching from OO to PC when belief α is higher than α_{PC}^{OO}). Hence, the value function Π need not be smooth at α_{OO}^{PC} as shown by Weng (2015).

Value matching and smooth pasting at α_{PC}^{FC} are always required by optimality, and such conditions imply $\Pi^+(\alpha_{PC}^{FC}) = \Pi^-(\alpha_{PC}^{FC})$ and $\Pi^{+'}(\alpha_{PC}^{FC}) = \Pi^{-'}(\alpha_{PC}^{FC})$. As expected, it coincides with the solution of $M(\alpha) = s$ since the latter one is the switching cutoff between $\ell = s$ and $\ell = r$, the myopic cutoff between risky arm and safe arm. We get

$$\alpha_{PC}^{FC} = \frac{s - z + \lambda_B B}{\lambda_G G + \lambda_B B} \tag{6}$$

Assumption 1 guarantees $\alpha_{PC}^{FC} \in (0, 1)$.

When $\lambda_G > \lambda_B$, previous analysis shows that α_{OO}^{PC} is pinned down by value matching and smooth pasting conditions, which require $\Pi^+(\alpha_{OO}^{PC}) = \Pi^-(\alpha_{OO}^{PC})$ and $\Pi^{+'}(\alpha_{OO}^{PC}) = \Pi^{-'}(\alpha_{OO}^{PC})$. Since $Pi^-(\alpha)$ is simply 0, we get

$$\alpha_{OO}^{PC} = \frac{s - z + \lambda_B B}{\lambda_G G + \lambda_B B + \lambda_G \frac{z_G - s}{r}} \tag{7}$$

So starting from a low initial belief, this cutoff serves as the boundary condition of $\Pi(\cdot)$ and pins down the value function together with (5). It's straightforward to observe that in this case, $\alpha_{PC}^{FC} \geq \alpha_{OO}^{PC}$ and they are equal iff $\lambda_G = 0$. That is, as long as good state gives some chance to learn, PC is optimal for some middle region of α.

When $\lambda_G < \lambda_B$, α_{OO}^{PC} only needs to satisfy value matching, i.e.

$$\Pi(\alpha_{OO}^{PC}) = 0 \tag{8}$$

In this case, we need another boundary condition $\Pi(1) = \tilde{V}$ to first pin down the value function in FC region. Once we have $\Pi(\cdot)$ in FC region, we can apply Value matching and smooth pasting at α_{PC}^{FC} to pin down $\Pi(\cdot)$ in PC region, and then use Equation (8) to solve α_{OO}^{PC}. In this case, the above solution procedure guarantees that $\alpha_{PC}^{FC} > \alpha_{OO}^{PC}$ holds strictly for sure, since $\Pi(\alpha_{PC}^{FC}) > 0$. This implies that the learning process would optimally start at a belief lower than the myopic one, a feature which also appears in Weng (2015).

We use Proposition 1 to describe the seller's optimal strategy.

Proposition 1. *There exsists a unique Markov Perfect Equilibrium, where seller's value function $\Pi(\alpha)$ and bonus scheme $b(\alpha)$ is continuous. There exist cutoffs $0 < \alpha_{OO}^{PC} < \alpha_{PC}^{FC} < 1$, where α_{PC}^{FC} is defined in (6), such that:*

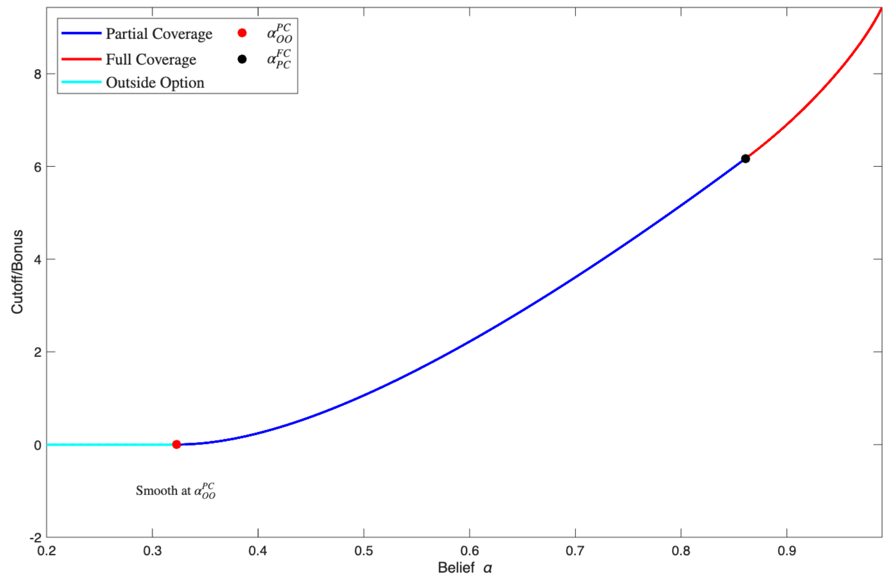
- *if $0 \leq \alpha < \alpha_{OO}^{PC}$, the monopolist would use the Outside Option.*
- *if $\alpha_{OO}^{PC} < \alpha < \alpha_{PC}^{FC}$, the monopolist would use the Partial Coverage strategy. $b(\alpha)$ is increasing in α in this region.*
- *if $\alpha_{PC}^{FC} < \alpha \leq 1$, the monopolist would use the Full Coverage strategy. $b(\alpha)$ is decreasing in α in this region, and $b(1) = 0$.*

Moreover:

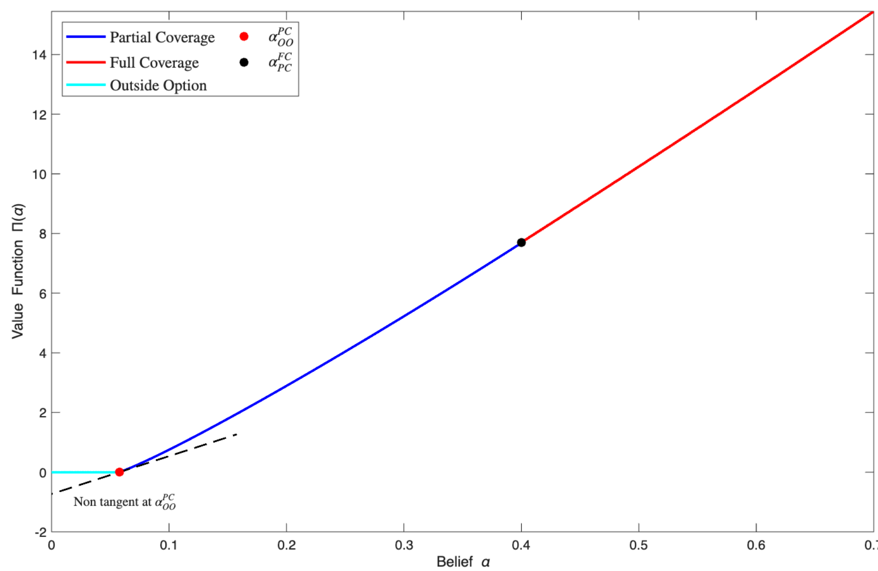
- *If $\lambda_G > \lambda_B$, α_{OO}^{PC} is defined in (7), and $\Pi(\alpha)$ is smooth on $[0, 1]$ interval.*
- *If $\lambda_B > \lambda_G$, α_{OO}^{PC} is defined in (8), and $\Pi(\alpha)$ is non-smooth only at α_{OO}^{PC}.*

Figure 1 illustrates Propositon 1[7]. Besides the switching order illustrated above, we want to emphasize that in Figure 1a value function is smooth pasting at α_{OO}^{PC}, while in Figure 1b it's not. We illustrate the optimal bonus scheme in Figure 2, where we want to emphasize the inverted U-shape of the bonus scheme predicted in Proposition 1.

[7]Paramter used: We assume $b \sim U[0, 1]$, hence $H(b) = b, \forall b \in [0, 1], \bar{c} = 1$. In left panel we use $(r = 0.05; \lambda_G = 0.6; \lambda_B = 0.3; z = 10; G = 5; B = 2; s = 12.5)$, and in right panel $(r = 0.05; \lambda_G = 0.2; \lambda_B = 0.8; z = 8; G = 6; B = 1; s = 8)$. They satisfy Assumption 1. We use the same parameter setting in Figure 2 where we can see $\bar{c} = 1$ is also large enough for this setting, so Assumption 3 is also satisfied.



(a) When $\lambda_G > \lambda_B$



(b) When $\lambda_G < \lambda_B$

Figure 1: Monopolist's Value Function

Optimal strategy at the bottom being OO should be intuitive. If the belief is extremely pessimistic, the risky arm is not only costly to sell (the seller needs to set a negative price), but also worthless for learning; the seller should stay outside of the market and never try to start the learning process (i.e., set $b(\alpha) = 0$ for all $\alpha < \alpha_{OO}^{PC}$).

If the belief is extremely optimistic, the optimal bonus is determined by the complex interaction between two forces. On the one hand, the bonus is set for the possibility that good news is revealed and hence the seller can further charge a higher price (exploration) at a cost of direct term $H(b(\alpha))b(\alpha)$. But on the other hand, exploration also leads to the possibility of bad news. When α is higher, though the seller himself believes good news comes more likely, the net improvement is marginal, captured by $b(\alpha) \to 0$ when $\alpha \to 1$. Nonetheless, as long as bad news has positive arrival rate (i.e., $\lambda_B > 0$), the seller would still find it optimal to set a positive bonus to encourage reporting even when the belief is very high, as shown in Lemma 4.1.

We conclude discussion of monopolist seller's optimal strategy with the tradeoff between FC and PC. The seller is able to use the Partial Coverage strategy and efficiently screens out those consumers whose reporting cost is sufficiently low, and motivates these low-cost buyers to purchase the risky arm and report. So PC is optimal in regiorn $[\alpha_{OO}^{PC}, \alpha_{PC}^{FC}]$ where the price is negative ($M(\alpha) < s$ so myopically risky arm is worse than the safe arm), but the risky arm is still worth exploring. In this region, the motivation of learning decreases if the common belief decreases further, hence the bonus under Partial Coverage is monotonically increasing with respect to belief α. In FC region the seller is exploiting the optimistic belief to charge a positive price, and the motivation of learning shrinks if the common belief increases further.

5 Welfare Analysis

5.1 Social Planner's Problem

We consider a setting where a social planner commits to an allocation rule $p(\alpha, \tilde{c})$, a transfer rule $t(\alpha, \tilde{c})$, and a reporting rule $q(\alpha, \tilde{c})$, where $\tilde{c}$ is the current agent i's reported value of c, which may or may not be equal to c. Agent can report any $\tilde{c} \in [0, \bar{c}]$. The social planner's objective is to maximize the total discounted sum of buyer surplus.

If the current belief is α, and the current agent with true cost c arrives and reports $\tilde{c}$, the social planner will allocate to that agent a risky arm (with probability $p(\alpha, \tilde{c})$) or a safe arm (with probability $1 - p(\alpha, \tilde{c})$), together with a transfer $t(\alpha, \tilde{c})$. Then, conditional on gettin allocated with the risky arm, the agent is obliged to report his news with probability $q(\alpha, \tilde{c})$, or not to report with probability $1 - q(\alpha, \tilde{c})$. News reporting and allocation will be determined at the beginning of period $[t, t + \Delta)$ based on common belief α and agent's cost $\tilde{c}$, and the news report and transfer $\Delta \times t(\alpha, \tilde{c})$ (if any) will happen at the end.

The following technique lemma shows it's without any loss to consider such a mechanism for social planner's problem.

Lemma 5.1. *There exists a mechanism (p, q, t) solves the social planner's problem, where $p : [0, 1] \times [0, \bar{c}] \to [0, 1]$ stands for the probability of allocating the risky arm, $q : [0, 1] \times [0, \bar{c}] \to [0, 1]$ stands for the probability of enforcing the agent to report the news conditional on getting the risky arm, and $t : [0, 1] \times [0, \bar{c}] \to [0, 1]$ stands for the transfer from the agent (agent pays t).*

In period Δ, with the current agent truthfully reporting cost c, given $p(\cdot, \cdot), q(\cdot, \cdot), t(\cdot, \cdot)$, the immediate gain of buyer surplus net with transfer, which is also the stage payoff of the social planner, is

$$\begin{aligned}\tilde{\pi}_{\Delta}(\alpha, c) =& p(\alpha, c) \int_0^{\Delta} [z + \alpha \lambda_G e^{-\lambda_G t} G - (1-\alpha)\lambda_B e^{-\lambda_B t} B] e^{-rt} dt \\ & - p(\alpha, c) q(\alpha, c) \int_0^{\Delta} e^{-rt} c dt + [1 - p(\alpha, c)] \int_0^{\Delta} s e^{-rt} dt\end{aligned} \tag{9}$$

We denote $U(c, \tilde{c}, \alpha) = \lim_{\Delta \to 0} \frac{\tilde{\pi}_{\Delta}(\alpha, c)}{\Delta}$. The incentive compatible constraint for agent with cost c thus requires

$$c \in \arg\max_{\tilde{c}} U(c, \tilde{c}, \alpha) - t(\alpha, \tilde{c})$$

We establish the HJB for contingent value function $W(\alpha, c)$ based on (9) and similar discussion of four contingencies in section 3.3. We then define the value function $W(\alpha)$ by taking expectation over $W(\alpha, c)$. We get a new HJB function:

$$\begin{aligned}rW(\alpha) =& \max_{p,q} \mathbb{E}(p)[M(\alpha) - s] - \mathbb{E}(pqc) + s \\ & + \mathbb{E}(pq) \left[\lambda_G \alpha \frac{z_G}{r} + \lambda_B (1-\alpha) \frac{s}{r} + \alpha(1-\alpha)(\lambda_B - \lambda_G) W'(\alpha) - \lambda(\alpha) W(\alpha) \right]\end{aligned} \tag{10}$$

5.2 Social Planner's Optimal Strategy

In the appendix, we establish an analogue of Lemma 4.1 to show that NB is also suboptimal for the social planner when $\lambda_B > 0$.

Proposition 2 describes the optimal strategy for the social planner. It shows that the optimal $p(\cdot, \cdot)$ and $q(\cdot, \cdot)$ take value only in $\{0, 1\}$, i.e., the social planner won't allocate randomly. Also remember that when $p = 0$, q doesn't matter since only the safe arm is allocated. We first claim the optimal strategy and then match it to the monopolist's strategy.

Proposition 2. *There exist optimal mechanisms (p, q, t) to guarantee that solves social planner's problem (10) and satisfies agents' IC and IR. When $\lambda_B > 0$, there exist cutoffs of belief $\tilde{\alpha}_{OO}^{PC}$, $\tilde{\alpha}_{PC}^{FC}$ and cutoffs of cost $C_1(\alpha)$ and $C_2(\alpha)$, such that*

- *if $\lambda_G > \lambda_B > 0$, then $\tilde{\alpha}_{OO}^{PC} = \alpha_{OO}^{PC}$, and the social planner's value function $W(\alpha)$ is smooth.*
- *if $\lambda_B > \lambda_G \geq 0$, then $\tilde{\alpha}_{OO}^{PC}$ is defined by $W(\tilde{\alpha}_{OO}^{PC}) = 0$, which is smaller than α_{OO}^{PC}, and the social planner's value function $W(\alpha)$ is non-smooth only at $\tilde{\alpha}_{OO}^{PC}$.*
- *$\tilde{\alpha}_{PC}^{FC} = \alpha_{PC}^{FC}$.*

In terms of the optimal allocation and reporting strategy,

- *if* $0 < \alpha < \tilde{\alpha}_{OO}^{PC}$*, the social planner would set* $p(\alpha, c) = q(\alpha, c) = 0, \forall c$ *(i.e., using the Outside Option).*

- *if* $\tilde{\alpha}_{OO}^{PC} \leq \alpha \leq \tilde{\alpha}_{PC}^{FC}$*, the social planner would set* $q(\alpha, c) = 1, \forall c$ *and* $p(\alpha, c) = 1$ *if* $c < C_1(\alpha)$ *and* $p(\alpha, c) = 1$ *otherwise (i.e., using Partial Coverage).*

- *if* $\tilde{\alpha}_{PC}^{FC} < \alpha \leq 1$*, the social planner would set* $p(\alpha, c) = 1, \forall c$ *and* $q(\alpha, c) = 1$ *if* $c < C_2(\alpha)$*, and set* $q(\alpha, c) = 0$ *otherwise (i.e., using Full Coverage).*

The transfer rule is stated in the Proof of Proposition 2 in the appendix.

To see how the social planner's strategy matches that of the monopolist accordingly, consider the following three cases:

- If $\forall c$ we set $p(\alpha, c) = 0$, this means the current agent is allocated to the safe arm. This is an analogue of the Outside Option (OO) strategy.

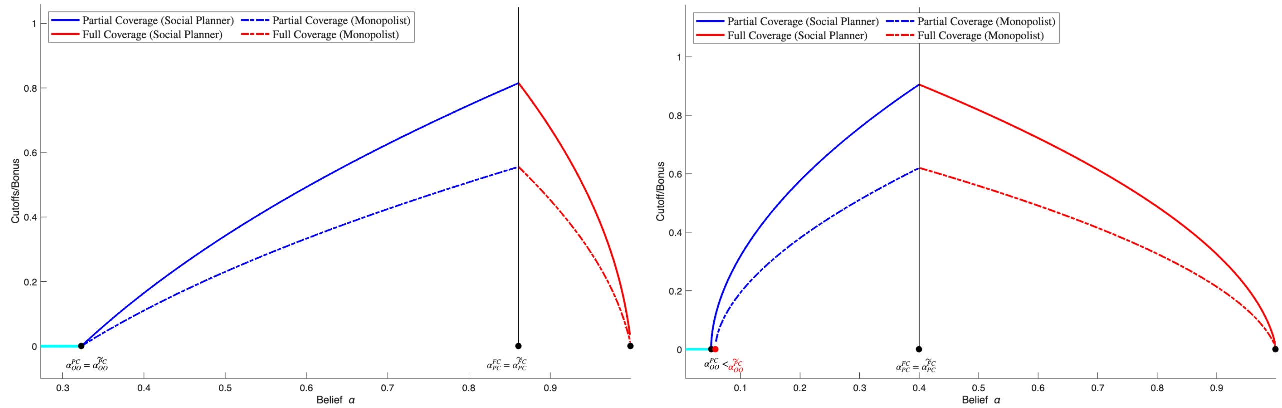
- If $\alpha > \alpha_{PC}^{FC}$, we have $p(\alpha, c) = 1, \forall c$. At this time, if $C_2(\alpha) > 0$, we have $q(\alpha, c) = 1$ whenever $0 \leq c \leq C_2(\alpha)$. This combination means the social planner would always allocate the risky arm to the current agent and have them report if their reporting cost is sufficiently low. This is an analogue of FC.

- If $\alpha \leq \alpha_{PC}^{FC}$ and $c \leq C_1(\alpha)$, we have $p(\alpha, c) = 1$, $c \leq C_1(\alpha)$ further implies $c \leq C_2(\alpha)$ so $q(\alpha, c) = 1$. Otherwise, if $\alpha \leq \alpha_{PC}^{FC}$ and $c > C_1(\alpha)$ we set $p(\alpha, c) = 0$. This combination means only low-cost agents would be allocated the risky arm, and they will definitely report. Otherwise, they are allocated the safe arm. This is an analogue of the PC strategy.

The fact that both the social planner and the monopolist seller use the same switching cutoff between Full Coverage and Partial Coverage is intuitive, since exploration generates negative stage payoff whenever $\alpha < \alpha_{PC}^{FC}$, hence it's efficient to screen out the lost cost agents and do exploration via allocating the risky arm to agents who report for sure. In monopolist setting, it's done by setting tie breaking rule $\ell = s$ and set a positive bonus $b(\alpha)$, while in social planner setting, it's done by setting $p(\alpha, c) = 1$ only for low cost agents whose cost is less than $C_1(\alpha)$, and set $q(\alpha, c) = 1$ whenever $p(\alpha, c) = 1$.

5.3 Welfare Comparison

We are now ready for welfare results. The efficiency of the learning process are usually measured by checking whether the insufficient or excessive learning appears (i.e., the learning stops too early or too late) compared to the social optimum (Bergemann and Välimäki, 1996; Keller et al., 2005; Weng, 2015, etc.). In our model, PC and FC strategies both involve learning, while OO strategy stops learning. Since FC ends at 1 unless $\lambda_B = 0$, it's enough to compare α_{OO}^{PC} and $\tilde{\alpha}_{OO}^{PC}$. Proposition 2 has answered this question. When $\lambda_G > \lambda_B$, both the social planner and the monopolist seller stop learning at the same cutoff, while when $\lambda_B > \lambda_G$, monopolist stops learning insufficiently early. Figure 2 illustrates this comparison.



(a) When $\lambda_G > \lambda_B$ (b) When $\lambda_G < \lambda_B$

Figure 2: Cutoff vs Bonus schemes

Our setting allows us to go further and compare the speed of learning under both settings. The bonus $b(\cdot)$ in the monpolist setting and the allocation and reporting rules $\mathbb{E}(pq)$ in the social planner setting both reflects the speed of revealing new signals and reducing the social uncertainty (Zhong, 2022). Our Proposition 3 confirms that the speed is higher under social planner setting, and further states that the low speed under monopolist setting leads to a social welfare loss. As an illustration, the cutoff of social planner is always higher in Figure 2, and the (expected) social surplus under monopolist setting is always lower than the value function of the social planner in Figure 3.

Proposition 3. *Speed under monopolist is always lower than the social planner. As a consequence, the social welfare under monopolist is always lower than the value function of social planner.*

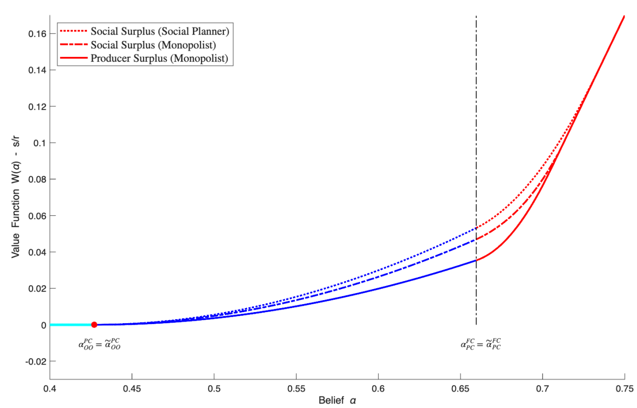


Figure 3: Social Welfare Comparison

The key difference between the social planner and the monopolist seller lies in the cost structure. In monopolist setting, when $b(\alpha) > c$, the cost of the report for the seller is $b(\alpha)$, so succeeding in motivating a report comes with a dead weight cost $b(\alpha) - c$ to the seller which comes from the lack of knowledge of c. However, for the social planner, if $C_i(\alpha) > c, i \in \{1, 2\}$ the cost is c instead of C_i, hence the social planner only bears the true cost of report and dead weight cost is 0. Hence, the social planner is more willing to do exploration via motivating reports, and this explains why the switching cutoff between OO and PC for the social planner $\tilde{\alpha}_{OO}^{PC}$ is weakly lower than that of the monopolist seller α_{OO}^{PC}. In section 6.1, when we introduce another non-learning strategy Non-Bonus, we can also show that the social planner switches from FC to NB at a higher cutoff than the monopolist seller, so the learning region is larger than that of the monopolist seller in both directions.

When $\lambda_G > \lambda_B$, the quitting decision is smooth, both the social planner and the monopolist seller set positive bonus for learning in the same region $\alpha \in (\alpha_{OO}^{PC}, 1)$. α_{OO}^{PC} stands for the belief where the strategy switches between Partial Coverage and the Outside Option, at which point, in the long run, the decision maker feels indifferent between experimenting further or not. One more experiment is of no value when the common belief is sufficiently low, and it is irrelevant to the cost of the experiment. Intuitively, given the quitting decision is smooth (only when $\lambda_G > \lambda_B$), if one more peace of news is valuable for the social planner, it is also valuable for the monopolist; the monopolist can always set a low enough bonus, and keep waiting until agents with sufficiently low costs arrive, try the risky arm, and report. The deadweight loss $\max\{b(\alpha_t) - c_t\} \to 0$ when $\alpha \to \alpha_{OO}^{PC+}$ so the monopolist stops at the same cutoff as the social planner.

When $\lambda_B > \lambda_G$, the quitting decision is non-smooth, and quitting decision happens via value matching (net with constant shift s/r, both monopolist's value function or social planner's value function are going to intersect 0 transversally). Because social planner has higher value function, it's strictly positive at α_{OO}^{PC}, leading to the intersection point $\tilde{\alpha}_{OO}^{PC} < \alpha_{OO}^{PC}$. The difference of two quitting cutoffs simply comes from the cost structure, that at α_{OO}^{PC} it's already not worthy for the monopolist to further explore, but since the cost is lower for the social planner, it's still socially efficient to explore at a non trivial bonus.

6 Extensions

In this section, we first consider the special case when bad news never arrives, and show that Non-Bonus (NB) strategy appears in a non-trivial way. Based on this characterization, we further investigate a polarity-contingent feedback system in which the bonus is only paid if the agent reports verifiable and conclusive good news.

6.1 Special Case of Good News Model

In this special case of $\lambda_B = 0$, we can apply the $\lambda_G > \lambda_B$ case in Proposition 1 to characterize the optimal strategy. However, as shown in Lemma 4.1, the seller may also find it optimal to set zero bonus and halt learning (NB). Our next proposition shows that if NB ever happens, it must be the case that α is sufficiently high. Intuitively, only when the belief is very high, it's optimal to always sell the risky arm and set zero bonus without motivating any risk to drop the belief.

Then the seller actually faces two different decisions about halting learning. One decision is the same as before: when the belief is low, it's optimal to quit learning and stay outside of the market (OO). The other decision happens when the belief is high: it's optimal to give up the NB strategy and start learning when the belief falls below a certain threshold.

As shown in Section 4.2, the smooth pasting is downward when $\lambda_G > \lambda_B$. Therefore, we should also have value matching and smooth pasting between PC and OO. But regarding the optimal starting problem between NB and FC, we only need value matching and don't require smooth pasting between NB and FC as in the case of $\lambda_B > \lambda_G$.

To illustrate the difference, in the proof of Proposition 4, we show that if FC and NB switch at some cutoff $\alpha_{FC}^{NB} < 1$ given by value matching condition, then we must have $\Pi'^{-}(\alpha_{FC}^{NB}) < \frac{M'(\alpha)}{r}$, so the value function is non smooth at α_{FC}^{NB}.

Proposition 4. *When $\lambda_B = 0$, there exist cutoffs $0 < \alpha_{OO}^{PC} < \alpha_{PC}^{FC} < \alpha_{FC}^{NB} \leq 1$, where α_{PC}^{FC} is defined in (6), α_{OO}^{PC} is defined in (7), α_{FC}^{NB} is the solution of $\Pi(x) = \frac{M(x)-s}{r}$, and $\Pi(\alpha)$ is non smooth only at α_{FC}^{NB}, such that:*

- *if $0 \leq \alpha < \alpha_{OO}^{PC}$, the monopolist would use the Outside Option.*
- *if $\alpha_{OO}^{PC} < \alpha < \alpha_{PC}^{FC}$, the monopolist would use Partial Coverage.*
- *if $\alpha_{PC}^{FC} < \alpha \leq \alpha_{FC}^{NB}$, the monopolist would use Full Coverage.*
- *if $\alpha_{FC}^{NB} < \alpha \leq 1$, the monopolist would use Non Bonus.*

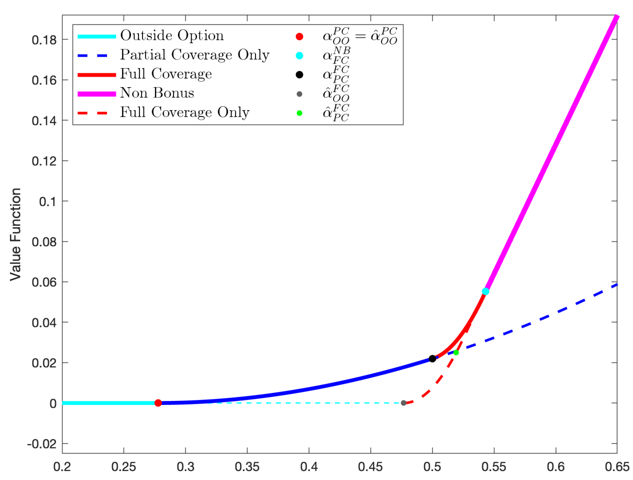


Figure 4: Seller's Value function when $\lambda_B = 0$

To understand why NB can be optimal only when $\lambda_B = 0$, we observe from Lemma 4.1 that fair pricing can internalize the arrival of good news but cannot internalize the arrival of bad news. This is because under the arrival of good news, it's optimal to keep selling the risky arm while under the arrival of bad news, it's optimal to stop selling the risky arm and avoid the potential loss of selling the risky arm at a low price. As a result, there is an optional value from the arrival of bad news which cannot be captured by the NB strategy.

The key difference between $\lambda_B = 0$ and $\lambda_B > 0$ lies in the fact that the probability of receiving bad news is zero in the former case but strictly positive in the latter case. As the probability of receiving conclusive bad news is strictly positive, it's never optimal to set zero bonus as this leads to a loss of the optional value from the arrival of bad news. Nonetheless, when the probability of receiving conclusive bad news is zero, it can be optimal to set zero bonus when the belief α is sufficiently close to 1.

6.2 Exponential Bandit with Manipulation

In the main context, we assume the seller cannot condition the bonus scheme on the signal reported by the agent, i.e., the seller pays b_t to publish the report even when report is negative. In this section, we allow the seller to set a polarity-contingent feedback system, and show that allowing such manipulation may lead to higher learning speed and hence higher social welfare.

We consider a specific form of manipulation[8], where the bonus is only paid if the agent reports conclusive good news. If the agent has no conclusive good news or his cost is above bonus, the bonus is not paid hence the agent won't report. Three other different situations are hence pooled together as a signal "no conclusive good news": agent has cost above b_t, agent observes conclusive bad news, and agent observes only flow utility (no conclusive news)[9].

Consequently, in this case, starting from t with belief α, at time $t + \Delta$ there are only two possible outcomes: either good news is realized within $[t, t + \Delta)$ and reported at the end of this period (at the end, the agent pays cost $c_t\Delta$ for bonus $b_t\Delta$), or there is no report at all (so no cost or bonus).

In detail, as the agents are indifferent between the two arms and they don't know the utility they get when they choose between the safe arm and risky arm, if the seller uses the Partial Coverage strategy, agents with $c_t > b_t$ will purchase the safe arm, and agents with $c_t < b_t$ will purchase the risky arm. Among the second group, only those agents who observe the good signal will report and obtain the bonus. Below, we show that setting $H(b) \in (0, 1)$ is always suboptimal, mainly because, in this case, the cost of bonus goes to 0 since conclusive good news is rare when $\Delta \to 0$, hence the seller will always motivate all agents to experiment when it's worthwhile.

Clearly if good news is reported, the belief jumps to 1 immediately. If not, the belief is updated according to Bayes' rule given no conclusive good news is reported. By martingale property of beliefs, $\alpha' < \alpha$. We show in Appendix that when $\Delta \to 0$,

$$(\alpha' - \alpha)/\Delta \to -\alpha(1 - \alpha)H(b)\lambda_G$$

[8]Discussion of optimal disclosure strategy is beyond the focus of this paper. This manipulation is mostly observed in real life, and we point out this manipulation is not always optimal in Proposition X

[9]Here a report of conclusive good news is seen as a standard verifiable disclosure from Milgrom (1981).

There are only two contingencies, either no report, or a report of conclusive good news. We show seller's HJB equation is now:

$$r\Pi(\alpha) = \max_{b,\ell} \pi(b,\ell) + H(b)\alpha\lambda_G\tilde{V} - H(b)\alpha\lambda_G\Pi(\alpha) + H(b)(\alpha^2-\alpha)\lambda_G\Pi'(\alpha) \tag{11}$$

where

$$\pi(b,\ell) = \begin{cases} M(\alpha)-s & \text{if FC or NB or IR} \\ H(b_t)[M(\alpha)-s] & \text{if PC} \\ 0 & \text{if OO} \end{cases} \tag{12}$$

Equation (12) shows the stage payoff is now independent of the bonus. Bonus is given at probability close to 0 since the realization of conclusive good news is rare. The right-hand side of (11) is hence linear in $H(b)$, since higher b only affects future belief and brings no stage cost. Noticing the term $M(\alpha)-s$ is increasing in α, there must be a cutoff $\bar{\alpha}$ such that $H(b)=1$ iff $\alpha \geq \bar{\alpha}$. When $H(b)$ is set to be 0, clearly $\Pi(\alpha)=0$. Otherwise, $\Pi(\alpha)$ satisfies the following ODE:

$$r\Pi(\alpha) = M(\alpha) - s + \alpha\lambda_G\tilde{V} - \alpha\lambda_G\Pi(\alpha) + (\alpha^2-\alpha)\lambda_G\Pi'(\alpha) \tag{13}$$

This ODE has a boundary condition coming from smooth pasting at $\bar{\alpha}$. Again it can be written as $\Pi(\bar{\alpha})=0$ where $\bar{\alpha}$ is defined by $\bar{\alpha} = \frac{s-z_B}{\lambda_G G+\lambda_B B+\lambda_G\frac{z_G-s}{r}}$, which is the same as α^{PC}_{OO}.

Except for the stage payoff difference, (13) is also different from (5) since the coefficient of Π' and Π on right hand side is now proportional to λ_G instead of $\lambda(\alpha)$. This is because the downward movement of belief only happens when no conclusive good news is reported with a probability related only to λ_G. We later show this change may allow λ_B to affect the comparison of two value functions.

We summarize the optimal strategy given the manipulation takes the specific form studies in this section with the following proposition:

Proposition 5. *Assume $\bar{c} < \infty$. Given the manipulation takes the form of disclosing conclusive good news and pooling the other three outcomes, then seller's optimal strategy is a cutoff strategy such that when $\alpha < \bar{\alpha}$ the seller sets bonus 0 and $\ell = s$, and when $\alpha > \bar{\alpha}$ the seller sets bonus $b = \bar{c}$ and $\ell = r$. The value function is smooth everywhere.*

We now show that such manipulation may increase the social welfare. The intuition comes from an observation discussed above: when $\lambda_B = 0$, the two ODEs are the same except for π, where the social welfare under no manipulation has an additional cost $\mathbb{E}(c \mid c < b)$. Since under manipulation there is not cost term in stage payoff, the seller maximizes the learning speed by setting $H(b) = 1$ at no cost, but doing so is too costly without manipulation. The continuity over λ_B hence implies that for small enough λ_B, such manipulation increases social welfare.However, as we show in Figure 5, this manipulation doesn't always dominate no manipulation when λ_B is large.

Proposition 6. *There exists a cutoff $\bar{\lambda_B}$ such that when $\lambda_B \leq \bar{\lambda_B}$, allowing such manipulation increases social welfare.*

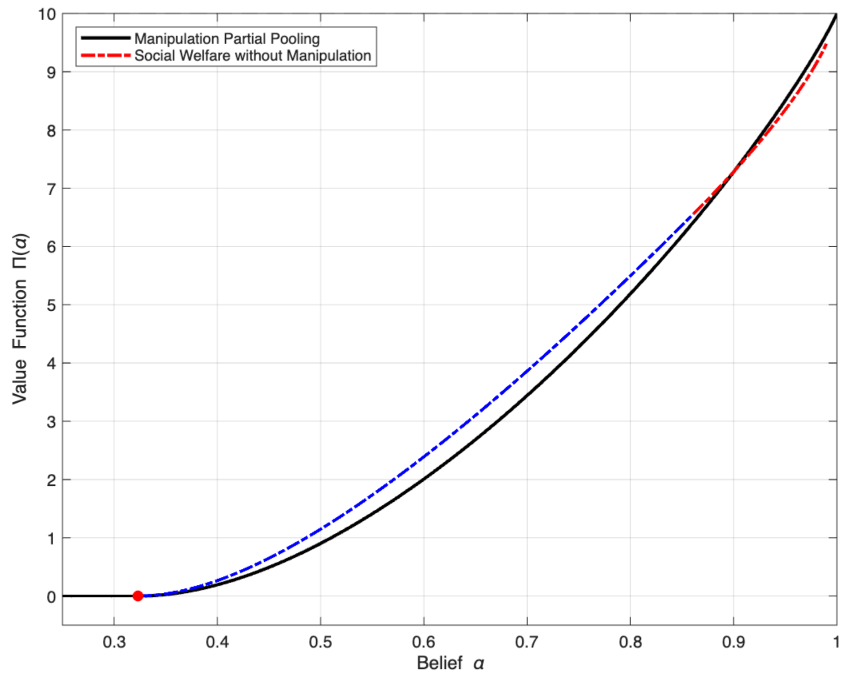


Figure 5: Such manipulation doesn not dominates no manipulation when λ_B is large

We briefly discuss the trade off of hiding bad news. Intuitively, given above manipulation form, conclusive bad news is pooled together with either a smooth drift α' or no update α. The benifit from hiding bad news is the belief won't drop to zero, which is shown in $\lambda_B(1-\alpha)\Pi(\alpha)$ term (if no manipulation, the right hand side of ODE has $\lambda(\alpha)\Pi$ but with manipulation it's $\alpha\lambda_G\Pi$). The cost is, now smooth drift and no update situations are interpreted as bad news since they are pooled with bad news. This loss is shown in $\lambda_B\alpha(1-\alpha)\Pi'$ term. Notice Π is convex increasing and start from $\Pi(\bar{\alpha}) = 0$, we always have $\Pi'(\alpha)\alpha > \Pi(\alpha)$. The net benefit of hiding bad news is always a negative term scaled up by $\lambda_B(1-\alpha)$. However, manipulation in this way has another additional benefit from saving the bonus cost, hence when λ_B is small enough or α close to 1, the net benefit of manipulation is positive.

7 Conclusion

We provide a simple model in which a seller can manage the interplay between product demand and the supply of learning signals through the joint design of pricing and reporting bonus schemes. A signal can be generated only by selling the risky arm, but this signal is not necessarily given if the bonus offered is too low. This feature makes it impossible to fully separate the acts of selling and learning. However, as we highlight, it can be optimal for the seller to decouple these acts as much as possible when two incentives are negatively correlated by treating price as a selling incentive and a bonus as a learning incentive. This occurs in the Partial Coverage region, where only the group of agents from whom the seller is actively learning is induced to purchase the risky arm. This strategy is optimal when the risky arm is perceived to be inferior to the safe alternative.

The difference between the paid bonus and the agents' true reporting costs imposes a profit loss on the seller, which is the primary source of inefficiency. The inefficiency in our model arises from both an early termination of learning and inefficiently slow learning speed. Nevertheless, because the optimal bonus approaches zero as the public belief nears the terminal threshold, this inefficiency does not distort the stopping decision when bonus can be used as a smoothing tool (i.e. $\lambda_G > \lambda_B$ in our model).

Our results have several practical implications. We show seller has higher willingness to halt the learning process when the social belief is extremely positive and the arrival rate of bad news is low. Furthermore, we find that if the seller can selectively motivate good news, the social surplus may also increase. This is because the bonus is paid less frequently, which strengthens the monopolist's incentive to experiment and accelerates learning. This leads to a seemingly counterintuitive policy insight: allowing platforms some latitude to manage their review sections might, by boosting their incentive to learn, ultimately enhance overall efficiency, if positive reviews are verifiable.

As we focus on the monopolist's incentive design, this article assumes that an agent's reporting cost is independent of their realized utility. The empirical literature suggests, however, that consumer incentives can be more complex; for example, they may be more motivated to report particularly negative or extreme experiences. It is also observed that some consumers are motivated to provide in-depth, high-quality reviews to establish a reputation as knowledgeable experts. This opens the door to exploring greater review heterogeneity in buyers' incentives and content. One can expect richer bonus contracts when the bonus is contingent on a review's observable characteristics, such as its deviation from others or helpfulness. Another simplification is that we assume the seller has no knowledge of buyers' reporting costs. It is possible to consider long-lived agents interacting with the seller, and, in the long run, the seller identifying a group of content creators with low reporting costs, as currently seen on YouTube.

References

Avery, C., P. Resnick, and R. Zeckhauser (1999). The market for evaluations. *American Economic Review 89*(3), 564–584.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics 107*(3), 797–817.

Bergemann, D. and J. Välimäki (1996). Learning and strategic pricing. *Econometrica: Journal of the Econometric Society*, 1125–1149.

Bergemann, D. and J. Välimäki (2000). Experimentation in markets. *The Review of Economic Studies 67*(2), 213–234.

Bergemann, D. and J. Välimäki (2006). Dynamic pricing of new experience goods. *Journal of Political Economy 114*(4), 713–743.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy 100*(5), 992–1026.

Bolton, P. and C. Harris (1999). Strategic experimentation. *Econometrica 67*(2), 349–374.

Bose, S., G. Orosel, M. Ottaviani, and L. Vesterlund (2006). Dynamic monopoly pricing and herding. *The RAND Journal of Economics 37*(4), 910–928.

Bose, S., G. Orosel, M. Ottaviani, and L. Vesterlund (2008). Monopoly pricing in the binary herding model. *Economic Theory 37*(2), 203–241.

BrightLocal (2025). Local consumer review survey 2025. https://www.brightlocal.com/research/local-consumer-review-survey/. Accessed: 2025-06-17.

Burtch, G., Y. Hong, R. Bapna, and V. Griskevicius (2018). Stimulating online reviews by combining financial incentives and social norms. *Management Science 64*(5), 2065–2082.

Cabral, L. and L. Li (2015). A dollar for your thoughts: Feedback-conditional rebates on ebay. *Management Science 61*(9), 2052–2063.

Cohen, A. and E. Solan (2013). Bandit problems with lévy processes. *Mathematics of Operations Research 38*(1), 92–107.

Fradkin, A., E. Grewal, D. Holtz, and M. Pearson (2015). Bias and reciprocity in online reviews: Evidence from field experiments on airbnb. *EC 15*, 15–19.

Hagiu, A. and J. Wright (2020). Data-enabled learning, network effects and competitive advantage. In *working paper*.

Keller, G., S. Rady, and M. Cripps (2005). Strategic experimentation with exponential bandits. *Econometrica 83*(1), 39–68.

Kenjiroi (2021). How to get reviews on amazon in 2021. https://kenjiroi.com/how-to-get-reviews-on-amazon-in-2021/.

Khern-am-nuai, W., K. Kannan, and H. Ghasemkhani (2018). Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research 29*(4), 871–892.

Kremer, I., Y. Mansour, and M. Perry (2014). Implementing the "wisdom of the crowd". *Journal of Political Economy 122*(5), 988–1012.

Krishna, V. and J. Morgan (2012). Voluntary voting: Costs and benefits. *Journal of Economic Theory 147*(6), 2083–2123.

Li, L. (2010). Reputation, trust, and rebates: How online auction markets can improve their feedback mechanisms. *Journal of Economics & Management Strategy 19*(2), 303–331.

Li, L., S. Tadelis, and X. Zhou (2016). Buying reputation as a signal of quality: Evidence from an online marketplace. Technical report, National Bureau of Economic Research.

Li, L. and E. Xiao (2014). Money talks: Rebate mechanisms in reputation system design. *Management Science 60*(8), 2054–2072.

McKinseyCompany (2024). Dynamic pricing. https://www.mckinsey.com/capabilities/growth-marketing-and-sales/how-we-help-clients/dynamic-pricing. Accessed: 2025-06-17.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 380–391.

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

Moe, W. W. and D. A. Schweidel (2012). Online product opinions: Incidence, evaluation, and evolution. *Marketing Science 31*(3), 372–386.

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory 9*(2), 185–202.

Stokey, N. L. and R. E. Lucas Jr (1989). *Recursive methods in economic dynamics*. Harvard University Press.

Weng, X. (2015). Dynamic pricing in the presence of individual learning. *Journal of Economic Theory 155*, 262–299.

Xu, H., D. Liu, H. Wang, and A. Stavrou (2015). E-commerce reputation manipulation: The emergence of reputation-escalation-as-a-service. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 1296–1306.

Zhong, W. (2022). Optimal dynamic information acquisition. *Econometrica 90*(4), 1537–1582.

A Ommitted Proofs

A.1 Proofs of Law of Motion and HJB in section 3.3.

We start from the following HJB equation with discrete time step Δ:

$$\begin{aligned}\Pi_\Delta(\alpha) = &\max_{\ell,b} \pi_\Delta(b,\ell) + e^{-r\Delta}H(b)\gamma_B(\alpha,\Delta)\cdot 0\\ &+ e^{-r\Delta}H(b)\gamma_G(\alpha,\Delta)\cdot\frac{z_G - s}{r}\\ &+ e^{-r\Delta}H(b)[1-\gamma_B(\alpha,\Delta)-\gamma_G(\alpha,\Delta)]\Pi_\Delta(\alpha')\\ &+ e^{-r\Delta}[1-H(b)]\Pi_\Delta(\alpha)\end{aligned}$$

and show that when $\Delta \to 0$, we have

$$r\Pi(\alpha) = \max_b[\max_\ell \pi(b,\ell)] + H(b)\left[\lambda_G\alpha\tilde{V} + \alpha(1-\alpha)(\lambda_B-\lambda_G)\Pi'(\alpha) - \lambda(\alpha)\Pi(\alpha)\right]$$

Notice that when $\Delta \to 0$,

$$\frac{\gamma_B}{\Delta} \to (1-\alpha)\lambda_B, \frac{\gamma_G}{\Delta} \to \alpha\lambda_G$$

and

$$\frac{\alpha' - \alpha}{\Delta} \to \alpha(1-\alpha)(\lambda_B - \lambda_G)$$

We use the Taylor expansion with Lagrangian remainder for $\Pi(\alpha)$, so there exists $\alpha \geq \xi \geq \alpha'$ such that

$$\Pi(\alpha') = \Pi(\alpha) + \Pi'(\alpha)(\alpha'-\alpha) + \frac{\Pi''(\xi)}{2}(\alpha'-\alpha)^2$$

Plug it into the discrete time HJB equation and drop the higher order terms, we get

$$\begin{aligned}\Pi_\Delta(\alpha)(1-e^{-r\Delta}) = &\max_{\ell,b} \pi_\Delta(b,\ell)\\ &+ e^{-r\Delta}H(b)\gamma_G(\alpha,\Delta)\cdot\frac{z_G - s}{r}\\ &+ \left[e^{-r\Delta}H(b)[1-\gamma_B(\alpha,\Delta)-\gamma_G(\alpha,\Delta)] - H(b)\right]\Pi_\Delta(\alpha)\\ &+ e^{-r\Delta}H(b)[1-\gamma_B(\alpha,\Delta)-\gamma_G(\alpha,\Delta)]\Pi'_\Delta(\alpha)(\alpha'-\alpha)\end{aligned}$$

Notice that when $\Delta \to 0$ the coefficient of the third line goes to $-rH(b)$ since γ_B, γ_G goes to 0. We then divide both sides by Δ and let $\Delta \to 0$, we get the desired HJB equation.

A.2 Proof of Lemma 4.1

Proof. We show by contradiction. If NB is optimal in some interval of α, we first use $\ell = r$ to rewrite equation (5) to

$$0 = \max_b H(b)[-b + (1-\alpha)\lambda_B\frac{s - z_B}{r}]$$

Notice that the term $s - z_B$ is positive by our assumption 1, right hand side is maximized by some $b^* > 0$ instead of $b = 0$, unless when $\lambda_B = 0$ or at α_1. Hence, we have a contradiction. Above argument also shows $\Pi(1) = \frac{M(1)}{r}$. □

A.3 Proof of Lemma 4.2

Proof. We know $\lim_{\Delta\to 0} \frac{p^*_\Delta(\alpha)}{\Delta} = z - s + \alpha\lambda_G G - (1-\alpha)\lambda_B B = M(\alpha) - s$ from (Fair Pricing), and

$$\pi(b,\ell) = H(b)[\lim_{\Delta\to 0} \frac{p^*_\Delta(\alpha)}{\Delta} - b] + [1 - H(b)]\mathbb{I}(\ell = r) \lim_{\Delta\to 0} \frac{p^*_\Delta(\alpha)}{\Delta}$$

So optimal $\ell = r$ iff $\lim_{\Delta\to 0} \frac{p^*_\Delta(\alpha)}{\Delta} > 0$ or $M(\alpha) > s$. Notice $M(\alpha)$ is increasing in α, hence the optimal $\mathbb{I}(\ell = r)(\alpha)$ is also increasing in α. That is, there exists a cutoff of α, $\bar{\alpha}$, such that below $\bar{\alpha}$, the optimal strategy is to set $\ell = s$ and above $\bar{\alpha}$, the optimal strategy is to set $\ell = r$. Given NB is suboptimal by Lemma 4.1, FC is optimal when $\alpha \in [\bar{\alpha}, 1]$.

When $\alpha \in [0, \bar{\alpha}]$, $\ell = s$, we show there exists another cutoff $\hat{\alpha}$ such that OO is better than PC when $\alpha \in [0, \hat{\alpha}]$ and PC is better than OO when $\alpha \in (\hat{\alpha}, \bar{\alpha}]$. Since we know OO generates zero value function, and PC generates a non negative one (since the seller can always set $b(\alpha) = 0$ to get 0), we only need to show the value function is increasing with respect to α in this region. Notice when $\ell = s$, $\pi(b,\ell) = H(b)[M(\alpha) - s - b]$ is increasing in α, and Π is the solution of the following Bellman equation:

$$\Pi(\alpha) = T(\Pi) \equiv \max_b \frac{\pi(b,\ell = s) + H(b)\left[\lambda_G\alpha\tilde{V} + \alpha(1-\alpha)(\lambda_B - \lambda_G)\Pi'(\alpha) - \lambda(\alpha)\Pi(\alpha)\right]}{r}.$$

We know that the solution $\Pi(\alpha)$ is increasing, by standard argument in Lemma 9.5 of Stokey and Lucas Jr (1989). Since $\pi(\alpha)$ is strictly increasing in α, Lemma 9.5 of Stokey and Lucas Jr (1989) also guarantees $\Pi(\alpha)$ is strictly increasing in PC and FC region.

The continuous and convex part follows Proposition 2 in Cohen and Solan (2013), since our value function is linear combination of value functions given state is good and bad, and the risky arm structure is a Levy process. The value function is then the supremum of a set (over bonus schemes space) of linear functions, hence convex and continuous. □

A.4 Proof of Proposition 1

Proof. We prove proposition 1 by showing $\Pi(\alpha)$ and the bonus scheme $b(\alpha)$ satisfy the following properties:

1. if $0 \le \alpha < \alpha^{PC}_{OO}$, $\Pi(\alpha) = 0$, $b(\alpha) = 0$.

2. if $\alpha_{OO}^{PC} < \alpha < \alpha_{PC}^{FC}$, we have

$$\begin{aligned} \beta(b) =& M(\alpha) - s + \lambda_G \alpha \frac{z_G - s}{r} + \alpha(1-\alpha)(\lambda_B - \lambda_G)\Pi'(\alpha) - \lambda(\alpha)\Pi(\alpha) \\ r\Pi(\alpha) =& \frac{H^2(b)}{h(b)} \end{aligned} \tag{14}$$

and $b(\alpha)$ is monotonically increasing in α.

3. if $\alpha_{PC}^{FC} < \alpha \leq 1$, we have

$$\begin{aligned} \beta(b) =& \lambda_G \alpha \frac{z_G - s}{r} + \alpha(1-\alpha)(\lambda_B - \lambda_G)\Pi'(\alpha) - \lambda(\alpha)\Pi(\alpha) \\ r\Pi(\alpha) =& M(\alpha) - s + \frac{H^2(b)}{h(b)} \end{aligned} \tag{15}$$

and $b(\alpha)$ is monotonically decreasing in α.

4. If $\lambda_G > \lambda_B$, value function is solved forward: $\Pi(\alpha)$ in PC region satisfies the ODE in point 2 and the boundary conditions $\Pi(\alpha_{OO}^{PC}) = 0$, then $\Pi(\alpha)$ in FC region satisfies the ODE in point 3 and the boundary condition $\Pi(\alpha_{PC}^{FC}) = \Pi^-(\alpha_{PC}^{FC})$.

5. If $\lambda_G < \lambda_B$, value function is solved backward: $\Pi(\alpha)$ in FC region satisfies the ODE in point 3 and the boundary conditions $\Pi(1) = \tilde{V}$, then $\Pi(\alpha)$ in PC region satisfies the ODE in point 2 and the boundary condition $\Pi(\alpha_{PC}^{FC}) = \Pi^+(\alpha_{PC}^{FC})$.

FOC is easy to get, simply differentiate (5) with respect to b and we get the first line in (14) and (15). Plug the FOC back to HJB, we get the second lines in (14) and (15). Given the second lines, clearly $b(\alpha)$ is continuous on whole $[0, 1]$ interval, since Π has the same property by Lemma 4.2.

When $\lambda_G > \lambda_B$ $\alpha' < \alpha$, so no conclusive news drops the belief smoothly downward, hence both the decision of downward quitting (switching from PC to OO) and the decision of switching from FC to PC are smooth, which implies smooth pasting at both α_{OO}^{PC} and α_{PC}^{FC}, we know $\Pi(\alpha)$ is smooth on whole $[0, 1]$ interval.

When $\lambda_B > \lambda_G$, $\alpha' > \alpha$, so no conclusive news raises the belief smoothly upward, hence the downward quitting decision need not to be smooth, value matching is sufficient at α_{OO}^{PC}. The upward quitting decision (switching from FC to NB), and switching from PC to FC is still smooth. By Lemma 4.1 the first condition is slack, but we have $\Pi'(1) = M'(1)/r$.

Then we want to figure out the monotonicity of the bonus scheme $b(\alpha)$ in PC and FC regions.

We define the drift coefficient $\mu(\alpha) \equiv \alpha(1-\alpha)(\lambda_B - \lambda_G)$.

In the PC region, the Bellman equation derived from the HJB is:

$$r\Pi(\alpha) = \frac{H^2(b)}{h(b)} \tag{16}$$

So $b(\alpha)$ shares the same monotonicity with $\Pi(\alpha)$ hence both increasing.

Now we investigate the monotonicity of $b(\alpha)$ in the FC region. From the Bellman equation $r\Pi(\alpha) = M(\alpha) - s + \frac{H^2(b)}{h(b)}$, differentiating with respect to α yields:

$$r\Pi'(\alpha) - M'(\alpha) = \left[\frac{d}{db}\left(\frac{H^2(b)}{h(b)}\right)\right] b'(\alpha)$$

With Assumption 2, we know $H^2(b)/h(b)$ is increasing in b. We define $k(\alpha) \equiv r\Pi'(\alpha) - M'(\alpha)$. The sign of $b'(\alpha)$ is the same as the sign of $k(\alpha)$.

First notice that $M(\alpha)$ is linear hence $M'(\alpha) \equiv M'(1)$. Since $\Pi(1) = M(1)/r$ and $\Pi(\alpha) > M(\alpha)/r, \forall \alpha < 1$, we must have $\Pi'(1) \leq M'(1)/r$. Since $\Pi(\alpha)$ is convex by Lemma 4.2, we know $\Pi'(\alpha) \leq \Pi'(1) \leq M'(1)/r$ for all $\alpha < 1$, hence $\Pi'(\alpha) \leq M'(\alpha)/r$, which shows $k(\alpha)$ is negative. So in FC part $b'(\alpha) < 0$.

□

A.5 Proof of Proposition 2 and Lemma 5.1

Proof. We use p, q, t for $p(\alpha, c), q(\alpha, c), t(\alpha, c)$ for simplicity. Notice now if the state is knwon to be good, $\tilde{V}^* = \frac{z_G}{r}$. t is set to satisfy IC once given p, q. So we first solve for p, q.

We add tilde to distinguish social planner's stage payoff from seller's one, i.e. $\tilde{\pi}$ and $\tilde{\alpha}$.

We first show NB is still suboptimal when $\lambda_B > 0$. Notice NB requires to allocate the risky arm always, and never enforce the agent to report the news, i.e. for some α interval, we set $p(\alpha, c) = 1, q(\alpha, c) = 0, \forall c$. To show it's suboptimal, we simplify (10) by plugging $W(\alpha) = \frac{M(\alpha)}{r}$ and $p \equiv 1$. We get

$$0 = \max_q s - \mathbb{E}(qc) + \mathbb{E}(q)\lambda_B \frac{(1-\alpha)(s - z_B)}{r}$$

NB requires $q \equiv 0$, which clearly fails to maximize above equation.

Then we solve optimal p, q. We have

$$\frac{\tilde{\pi}_\Delta(\alpha, c)}{\Delta} = pM(\alpha) + (1-p)s - pqc$$

$$\begin{aligned} W_\Delta(\alpha, c) = &\max_{p,q} \tilde{\pi}_\Delta(\alpha, c) + e^{-r\Delta} pq\gamma_B(\alpha, \Delta) \cdot \frac{s}{r} \\ &+ e^{-r\Delta} pq\gamma_G(\alpha, \Delta) \cdot \frac{z_G}{r} \\ &+ e^{-r\Delta} pq[1 - \gamma_B(\alpha, \Delta) - \gamma_G(\alpha, \Delta)]W_\Delta(\alpha', c) \\ &+ e^{-r\Delta}[1 - pq]W_\Delta(\alpha, c) \end{aligned} \tag{17}$$

Hence

$$
\begin{aligned}
W_\Delta(\alpha) = &\max_{p,q} \mathbb{E}[\tilde{\pi}_\Delta(\alpha,c)] + e^{-r\Delta}\mathbb{E}(pq)\gamma_B(\alpha,\Delta)\cdot\frac{s}{r} \\
&+ e^{-r\Delta}\mathbb{E}(pq)\gamma_G(\alpha,\Delta)\cdot\frac{z_G}{r} \\
&+ e^{-r\Delta}\mathbb{E}(pq)[1-\gamma_B(\alpha,\Delta)-\gamma_G(\alpha,\Delta)]W_\Delta(\alpha') \\
&+ e^{-r\Delta}[1-\mathbb{E}(pq)]W_\Delta(\alpha)
\end{aligned}
\tag{18}
$$

With the same calculation as in section 3.3, let

$$
\pi(p,q) \equiv \mathbb{E}\lim_{\Delta\to 0}\frac{\tilde{\pi}_D(\alpha,c)}{\Delta} = \mathbb{E}(p)[M(\alpha)-s] - \mathbb{E}(pqc) + s
$$

we can simplify above equation to:

$$
\begin{aligned}
rW(\alpha) = &\max_{p,q} \mathbb{E}(p)[M(\alpha)-s] - \mathbb{E}(pqc) + s \\
&+ \mathbb{E}(pq)\left[\lambda_G\alpha\frac{z_G}{r} + \lambda_B(1-\alpha)\frac{s}{r} + \alpha(1-\alpha)(\lambda_B-\lambda_G)W'(\alpha) - \lambda(\alpha)W(\alpha)\right]
\end{aligned}
\tag{19}
$$

Given above linear structure, we know the optimal p, q are corner solutions. We denote

$$
\lambda_G\alpha\frac{z_G}{r} + \lambda_B(1-\alpha)\frac{s}{r} + \alpha(1-\alpha)(\lambda_B-\lambda_G)W'(\alpha) - \lambda(\alpha)W(\alpha) \equiv B(\alpha)
$$

which is irrelevant to c. So optimal $q(\alpha,c)$ maximizes $\mathbb{E}(pq)B(\alpha) - \mathbb{E}(pqc)$. If $p(\alpha,c) = 0$ then any q is optimal. Otherwise, optimal $q(\alpha,c) = 1$ if $B(\alpha) - c > 0$ and $q(\alpha,c) = 0$ if $B(\alpha) - c < 0$. We denote cutoff $C_2(\alpha) = B(\alpha)$, we have

$$
q(\alpha,c) = \begin{cases} 1 & \text{if } c \le C_2(\alpha) \text{ and } p(\alpha,c) > 0 \\ 0 & \text{if } c > C_2(\alpha) \text{ or } p(\alpha,c) = 0, \end{cases}
\tag{20}
$$

Plug this result back to (10), we have

$$
\begin{aligned}
rW(\alpha) - s &= \max_p \mathbb{E}(p)[M(\alpha)-s] + \int_c p(\alpha,c)\max\{B(\alpha)-c,0\}dH(c) \\
&= \max_p \int_c p(\alpha,c)\left[M(\alpha)-s+\max\{B(\alpha)-c,0\}\right]dH(c).
\end{aligned}
\tag{21}
$$

Remember $M(\alpha) > s$ iff $\alpha > \alpha_{PC}^{FC}$. That is $\alpha_{PC}^{\tilde{FC}} = \alpha_{PC}^{FC}$. Let $C_1(\alpha) = B(\alpha) + M(\alpha) - s$, we have

$$
p(\alpha,c) = \begin{cases} 1 & \text{if } c \le C_1(\alpha) \text{ or } \alpha > \alpha_{PC}^{FC} \\ 0 & \text{otherwise,} \end{cases}
\tag{22}
$$

Since $\alpha > \alpha_{PC}^{FC}$ iff $C_1(\alpha) > C_2(\alpha)$, we finally obtain the law of motion for the social planner:

$$
rW(\alpha) = \begin{cases} M(\alpha) + \int_{0<c<C_2(\alpha)} B(\alpha) - c\,dH(c), & \text{if } \alpha > \alpha_{PC}^{FC} \\ s + \int_{0<c<C_1(\alpha)} M(\alpha) - s + B(\alpha) - c\,dH(c), & \text{if } \alpha_{PC}^{FC} > \alpha \end{cases}
\tag{23}
$$

We now specify the optimal switching rule. We have shown that the social planner's value func tion is smooth pasting between PC to FC at $\tilde{\alpha}_{PC}^{FC}$ exactly the same cutoff as the monopolist α_{PC}^{FC}.

At $\tilde{\alpha}_{OO}^{PC}$, if $\lambda_G > \lambda_B$, we have smooth pasting condition, i.e. $W^+(\tilde{\alpha}_{OO}^{PC}) = s/r$ and $W^{+'}(\tilde{\alpha}_{OO}^{PC}) = 0$. We solve (23).

$$rW(\alpha) = s + \int_{0<c<C_1(\alpha)} M(\alpha) - s + B(\alpha) - cdH(c)$$

and get

$$\tilde{\alpha}_{OO}^{PC} = \frac{r(s - z_B)}{(r + \lambda_G)(z_G - s) + r(s - z_B)} \tag{24}$$

which is exactly the same as α_{OO}^{PC} defined in (7).

If $\lambda_G < \lambda_B$, we use value matching condition $W(\tilde{\alpha}_{OO}^{PC}) = s/r$ to pin down $\tilde{\alpha}_{OO}^{PC}$. Since the social planner's net value function $W(\alpha) - \frac{s}{r}$ is point wise larger than that of the monopolist, we actually get now the $\tilde{\alpha}_{OO}^{PC}$ of the social planner is strictly lower than that of the monopolist, α_{OO}^{PC}.

Finally we show there exists a transfer rule $t(\alpha, c)$ to implement the above allocation and reporting rules (that is, makes agents' IC and IR (with lower bound of utility 0), but social planner may not have budget balance).

Consider the following trivial transfer rule:

$$t(\alpha, c) = \begin{cases} s & \text{if } \alpha < \alpha_{PC}^{FC}, c > C_1(\alpha) \\ M(\alpha) - C_1(\alpha) & \text{if } \alpha < \alpha_{PC}^{FC}, 0 \le c \le C_1(\alpha) \\ M(\alpha) & \text{if } \alpha \ge \alpha_{PC}^{FC}, c > C_2(\alpha) \\ M(\alpha) - C_2(\alpha) & \text{if } \alpha \ge \alpha_{PC}^{FC}, 0 \le c \le C_2(\alpha) \end{cases} \tag{25}$$

We can see at any α, the social planner only distinguish whether the agent has cost below or above the cutoff, and set a transfer insensitive to the cost. We can then check the IC by simply check whether the agent wants to deviate to report any cost in the other region. The following table shows the net utility of agent reporting truthfully is non-negative and larger than the deviation utility of agent misreporting to the other region. So both IC and IR are satisfied.

condition1	condition2	$p(\alpha, c)$	$q(\alpha, c)$	$t(\alpha, c)$	net utility	deviation utility
$\alpha \ge \alpha_{PC}^{FC}$	$0 \le c \le C_2(\alpha)$	1	1	$M(\alpha) - C_2(\alpha)$	$C_2(\alpha) - c > 0$	0
$\alpha \ge \alpha_{PC}^{FC}$	$C_2(\alpha) < c < \bar{c}$	1	0	$M(\alpha)$	0	$C_2(\alpha) - c < 0$
$\alpha < \alpha_{PC}^{FC}$	$0 \le c \le C_1(\alpha)$	1	1	$M(\alpha) - C_1(\alpha)$	$C_1(\alpha) - c > 0$	0
$\alpha < \alpha_{PC}^{FC}$	$C_1(\alpha) < c < \bar{c}$	0	-	s	0	$C_1(\alpha) - c < 0$

Table 2: Optimal mechanism of social planner

□

A.6 Proof of Proposition 3

Proof. Once we map p, q, C_1, C_2 back to b, ℓ, we have $H(b) = 1$ when $p = q = 1$, and $[1 - H(b)]\mathbb{I}(\ell = r) = 1$ when $p = 1, q = 0$. So $W(\alpha)$ also fits the same structure as in (5), with only difference in the stage payoff:

$$\pi(\ell, b) = H(b)[M(\alpha) - s - \mathbb{E}(c \mid b > c)] + [1 - H(b)]\mathbb{I}(\ell = r)[M(\alpha) - s]$$

Notice that $\mathbb{E}(c \mid b > c) \leq b$, so only stage payoff increases, and cost is less with the same bonus. Hence optimal b is weakly higher in social planner setting, leading to higher speed. $\square$

A.7 Proof of Proposition 4

Proof. $\lambda_B = 0$ falls into the case $\lambda_G > \lambda_B$ in Proposition 1, and the switching order of OO, PC and FC follows the same logic as in Lemma 4.2. We just need to show that if NB ever happens, it must be the case that α is sufficiently high.

Suppose in contrast that FC happens for $\alpha \in (1 - \varepsilon, 1)$ for some $\varepsilon > 0$. Then we must have $\Pi(1) = \frac{M(1)}{r}$ and $\Pi(\alpha) > \frac{M(\alpha)}{r}$ for $\alpha \in (1 - \varepsilon, 1)$. Due to the convexity of $\Pi(\cdot)$, it must be the case that $\Pi(\alpha) > \frac{M(\alpha)}{r}$ also holds that $\alpha = 1 - \varepsilon$ and hence NB can never be optimal as $\frac{M(\alpha)}{r}$ is the value of NB.

From Proposition 1, together with $b(\alpha_{FC}^{NB}) = 0$, we have:

$$\begin{aligned} 0 = \beta(b) =& \lambda_G \alpha \frac{z_G - s}{r} - \alpha(1 - \alpha)(\lambda_G)\Pi'(\alpha) - \lambda_G \alpha \Pi(\alpha) \\ r\Pi(\alpha) =& M(\alpha) - s \end{aligned}$$

which shows

$$\lambda_G \frac{s}{r} = (1 - \alpha)\lambda_G(\frac{\lambda_G G}{r} - \Pi'(\alpha))$$

Since LHS is positive, the solution $\alpha < 1$, and $\Pi'(\alpha) < \frac{\lambda_G G}{r}$. That is to say, at α_{FC}^{NB} (defined by $\Pi(\alpha) = \frac{M(\alpha) - s}{r}$, we have $\Pi'(\alpha_{FC}^{NB}) < M'(1)/r$, so Π is non-smooth at α_{FC}^{NB}. $\square$

A.8 Proof of Proposition 5

Clearly if good news is reported, the belief jumps to 1 immediately. If not, the belief is updated according to Bayes' rule given no conclusive good news is reported. By martingale property of beliefs, $\alpha' < \alpha$.

If state is good, such probability is:

$$P(\text{no conclusive good news} \mid \theta = g) = 1 - H(b)(1 - e^{-\lambda_G \Delta})$$

If state is bad, then probability of no conclusive good news is 1. So

$$\alpha' = \frac{\alpha[1 - H(b)(1 - e^{-\lambda_G \Delta})]}{\alpha[1 - H(b)(1 - e^{-\lambda_G \Delta})] + 1 - \alpha}$$

hence

$$(\alpha' - \alpha)/\Delta \to -\alpha(1-\alpha)H(b)\lambda_G$$

The seller's discrete time HJB is hence:

$$\Pi_\Delta(\alpha) = \max_{b,\ell} \tilde{\pi}_\Delta(b,\ell) + H(b)\alpha(1-e^{-\lambda_G \Delta})e^{-r\Delta}\tilde{V} + [1 - H(b)\alpha(1-e^{-\lambda_G\Delta})]e^{-r\Delta}\Pi_\Delta(\alpha') \quad (26)$$

Using the same tricks (divide by $\Delta \to 0$) as before generates (11).

Setting $H(b) = 1$ we get the ODE and the boundary condition described in the main context. This boundary problem has a unique solution with closed form expression:

$$\Pi(\alpha) = \frac{M(\alpha) - s}{r} + C\alpha^{-r/\lambda_G}(1-\alpha)^{(r+\lambda_G)/\lambda_G}$$

Where

$$C = -\frac{M(\bar{\alpha}) - s}{r}\bar{\alpha}^{r/\lambda_G}(1-\bar{\alpha})^{-(r+\lambda_G)/\lambda_G}$$

A.9 Proofs of Proposition 6

Proof. First notice that with manipulation, the social welfare equals the seller's payoff, since buyers get zero surplus (no report) with probability 1. We use Π_1 to denote the value function standing for the social welfare when manipulation is allowed, and Π_2 when not allowed. We start from PC region.

If manipulate, Π_1 satisfies:

$$r\Pi(\alpha) = M(\alpha) - s + \alpha\lambda_G\tilde{V} - \alpha\lambda_G\Pi(\alpha) + (\alpha^2 - \alpha)\lambda_G\Pi'(\alpha) \quad (27)$$

If not, in PC region, Π_2' satisfies:

$$r\Pi(\alpha) = H(b)\left[M(\alpha) - s - \mathbb{E}(c \mid c < b) + \alpha\lambda_G\tilde{V} - \alpha\lambda_G\Pi(\alpha) + (\alpha^2-\alpha)\lambda_G\Pi'(\alpha)\right] \quad (28)$$

they have the same boundary condition that at α_{OO}^{PC} the value function is 0. We show the claim by expanding Π_2' to Π_2 and show $\Pi_1 > \Pi_2$ instead. Π_2 is defined as:

$$r\Pi(\alpha) = H(b)\left[M(\alpha) - s + \alpha\lambda_G\tilde{V} - \alpha\lambda_G\Pi(\alpha) + (\alpha^2-\alpha)\lambda_G\Pi'(\alpha)\right] \quad (29)$$

Clearly $\Pi_2 > \Pi_2'$ since the stage cost is decreased to 0.

Lemma A.1. *For all $\alpha \in (\bar{\alpha}, 1)$, we have $\Pi_1(\alpha) > \Pi_2(\alpha)$.*

Proof. We define the linear differential operator $\mathcal{L}$ acting on a smooth function $V(\alpha)$ as:

$$\mathcal{L}[V](\alpha) \equiv (\alpha^2-\alpha)\lambda_G V'(\alpha) - (r + \alpha\lambda_G)V(\alpha) + \left[M(\alpha) - s + \alpha\lambda_G\tilde{V}\right] \quad (30)$$

So

$$\mathcal{L}[\Pi_1] = 0$$

Next, we evaluate the operator $\mathcal{L}$ on Π_2. We rearrange second ODE by dividing both sides by $H(b)$ (noting $H(b) > 0$):

$$\frac{r\Pi_2}{H(b)} = M(\alpha) - s + \alpha\lambda_G\tilde{V} - \alpha\lambda_G\Pi_2 + (\alpha^2 - \alpha)\lambda_G\Pi_2'$$

Moving terms related to Π_2 and derivatives to the right-hand side to match the structure of $\mathcal{L}$:

$$\begin{aligned}\frac{r\Pi_2}{H(b)} &= \left[(\alpha^2 - \alpha)\lambda_G\Pi_2' - \alpha\lambda_G\Pi_2 + M(\alpha) - s + \alpha\lambda_G\tilde{V}\right] \\ &= \mathcal{L}[\Pi_2] + r\Pi_2\end{aligned}$$

Solving for $\mathcal{L}[\Pi_2]$:

$$\mathcal{L}[\Pi_2] = r\Pi_2(\alpha)\left(\frac{1}{H(b)} - 1\right)$$

Since in PC region (except for boundary) we have $0 < H(b) < 1$, we have $\frac{1}{H(b)} - 1 > 0$. So it follows that:

$$\mathcal{L}[\Pi_2] > 0, \forall\alpha \in (\bar{\alpha}, \alpha_{PC}^{FC}].$$

Now, let $w(\alpha) = \Pi_1(\alpha) - \Pi_2(\alpha)$. By the linearity of the operator $\mathcal{L}$:

$$\mathcal{L}[w] = \mathcal{L}[\Pi_1] - \mathcal{L}[\Pi_2] = 0 - \mathcal{L}[\Pi_2] \leq 0$$

Substituting the definition of $\mathcal{L}[w]$:

$$(\alpha^2 - \alpha)\lambda_G w'(\alpha) - (r + \alpha\lambda_G)w(\alpha) \leq 0$$

Rearranging the inequality:

$$(\alpha^2 - \alpha)\lambda_G w'(\alpha) \leq (r + \alpha\lambda_G)w(\alpha)$$

For $\alpha \in (\bar{\alpha}, 1)$, the drift term coefficient $(\alpha^2 - \alpha)\lambda_G$ is strictly negative. Dividing by this negative term reverses the inequality:

$$w'(\alpha) \geq \underbrace{\frac{r + \alpha\lambda_G}{(\alpha^2 - \alpha)\lambda_G}}_{K(\alpha)} w(\alpha) \implies w'(\alpha) - K(\alpha)w(\alpha) \geq 0$$

Using the integrating factor $I(\alpha) = \exp\left(-\int_{\bar{\alpha}}^{\alpha} K(x)dx\right) > 0$:

$$\frac{d}{d\alpha}\left(w(\alpha)I(\alpha)\right) \geq 0$$

Integrating from the boundary $\bar{\alpha}$ to α:

$$w(\alpha)I(\alpha) - w(\bar{\alpha})I(\bar{\alpha}) \geq 0$$

Given the boundary condition $w(\bar{\alpha}) = \Pi_1(\bar{\alpha}) - \Pi_2(\bar{\alpha}) = 0$, we conclude:

$$w(\alpha) \geq 0 \implies \Pi_1(\alpha) \geq \Pi_2(\alpha)$$

And above inequality is tight only at the boundary $\alpha = \bar{\alpha}$. So Π_1 strictly dominates Π_2 in $(\bar{\alpha}, \alpha_{PC}^{FC})$. This completes the proof. □

The left part is to show $\Pi_1(\alpha) > \Pi_2(\alpha)$ for all $\alpha \in (\bar{\alpha}, \alpha_{PC}^{FC})$. It's another similar application of above comparison principle, only differentce is now at the boundary α_{PC}^{FC}, we have $\Pi_1 > \Pi_2$ so w is increasing and starting from a positive value, hence positive on the whole interval.

□

B Online Appendix

B.1 Immediate Revelation

In our main analysis, Assumption 3 allowed us to ignore the potential for an Immediate Revelation strategy. In this section, we drop that assumption and consider the case of a small $\bar{c}$. When $\bar{c}$ is sufficiently low, the seller may find it optimal to set a bonus $b_t = \bar{c}$, in terms of which an agent will report their utility with certainty after purchasing the risky arm. We refer to this as the Immediate Revelation (IR) strategy. In this extension, we firt establish the optimal strategy and then argue that although the social planner and the monopolist respond differently to a small $\bar{c}$, their limiting behaviors are similar: as $\bar{c} \to 0$, Full and Partial Coverage degenerates to IR and hence shrinks, hence the learning process is implemented purely through IR.

IR can be treated as a corner case of PC and FC by setting the bonus $b_t = \bar{c}$. In this case, almost all buyers receive a strictly positive surplus from purchasing the risky arm and reporting, rendering the tie-breaking rule redundant. As stated in Proposition 1, the equilibrium bonus structure is single-peaked such that the highest bonus is paid at α_{PC}^{FC}. So the lower bound of $\bar{c}$ can be captured by calculating $b(\alpha_{PC}^{FC})$ as if IR is not allowed. When $\bar{c}$ is larger than this cutoff, naturally IR is suboptimal. But when $\bar{c}$ is smaller than this cutoff, IR appears around α_{PC}^{FC}. By the continuity and piece wise monotonicity of $b(\cdot)$, if $\bar{c} > b(\alpha_{PC}^{FC})$, then $\exists \underline{\alpha} < \alpha_{PC}^{FC} < \bar{\alpha}$ such that $b(\underline{\alpha}) = b(\bar{\alpha}) = \bar{c}$. Given IR appears to be optimal, the seller's strategy would further adjust, hence there would be new cutoffs $\alpha_{PC}^{IR} \in (\underline{\alpha}, \alpha_{PC}^{FC})$ and $\alpha_{IR}^{FC} \in (\alpha_{PC}^{FC}, \bar{\alpha})$ to smoothly connect IR with PC and FC.

We now characterize the seller's value function. The stage payoff for this strategy is $\lim_{\Delta \to 0} \frac{\pi_L(\Delta, b_t = \bar{c})}{\Delta} = M(\alpha) - s - \bar{c}$ when the seller uses IR. The law of motion in IR region is analogous to (5):[10] (we denote Y as the value function restricted in IR region to better state the optimal switching condition):

$$rY(\alpha) = M(\alpha) - s - \bar{c} + \lambda_G \alpha \tilde{V} + \alpha(1-\alpha)(\lambda_B - \lambda_G)Y'(\alpha) - \lambda(\alpha)Y(\alpha) \tag{31}$$

This solution must satisfy the following smooth-pasting conditions: [11]

$$\Pi(\alpha_{PC}^{IR}) = Y(\alpha_{PC}^{IR}), \quad \Pi'(\alpha_{PC}^{IR}) = Y'(\alpha_{PC}^{IR}) \tag{32}$$

and

$$Y(\alpha_{IR}^{FC}) = \Pi(\alpha_{IR}^{FC}), \quad Y'(\alpha_{IR}^{FC}) = \Pi'(\alpha_{IR}^{FC}) \tag{33}$$

[10]Notice here that if we allow this $Y(\cdot)$ to optimally switch to Π^{OO}, i.e., $\exists \tilde{\alpha}$ such that $Y(\tilde{\alpha}) = \frac{s}{r}$, and $Y'(\tilde{\alpha}) = 0$, then the cutoff belief should satisfy $\tilde{\alpha} = \frac{r(s+\bar{c})}{rz_G + \lambda(z_G - s)}$. A sufficient condition to ignore immediate revelation is $\tilde{\alpha} \geq 1$, which provides a concrete lower bound of $\bar{c}$ in our Assumption 3.

[11]As before, we have smooth pasting upwards or downwards, depending $\lambda_G > \lambda_B$ or not. However since IR, PC, FC are all optimally deciding b, switching at α_{PC}^{IR} and α_{IR}^{FC} are both smooth. If $\lambda_G > \lambda_B$, we first solve $\Pi(\cdot)$ restricted in the PC region using the same law of motion and boundary conditions stated in proposition 1, then solve for α_{PC}^{IR} and $Y(\cdot)$ using the first smooth-pasting condition (32) and law of motion 31. Then we solve α_{IR}^{FC} and $\Pi(\cdot)$ for the left region by the second smooth-pasting condition (33). The solving order is reversed when $\lambda_B > \lambda_G$ and we start from $\Pi(1) = M(1)/r$.

Proposition 7 is an analogue of Proposition 1 when Assumption 3 is dropped.

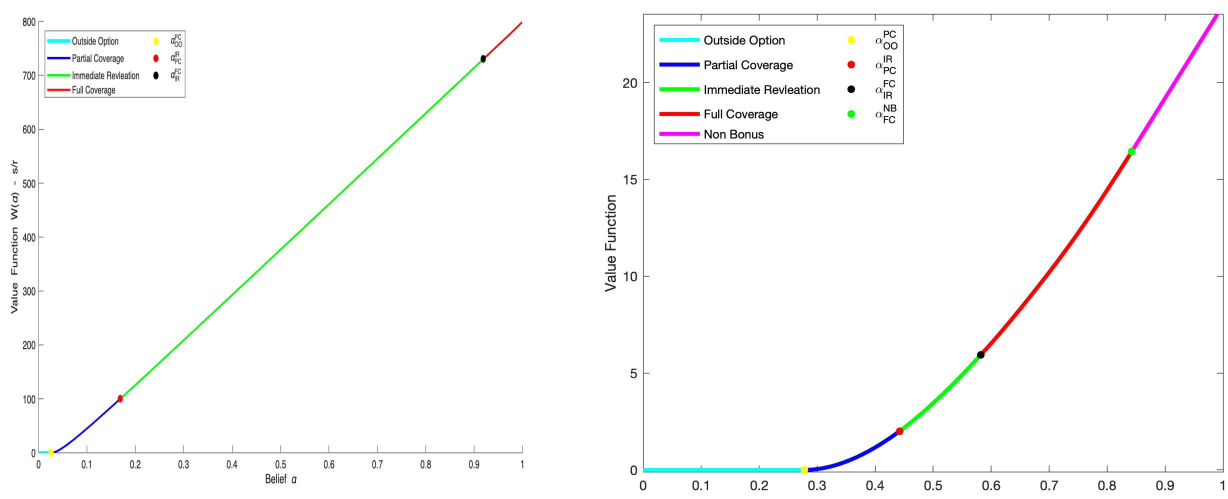
Proposition 7. *Assume $\lambda_B > 0$. There exists a cutoff $\bar{c}^*$ such that, for all $\bar{c} \in (0, \bar{c}^*)$, there is a continuous value function $\Pi(\alpha)$ and bonus scheme $b(\alpha)$ solving seller's problem. There exist cutoffs $0 < \alpha_{OO}^{PC} < \alpha_{PC}^{IR} < \alpha_{IR}^{FC} < 1$, such that*

- *if $0 < \alpha < \alpha_{OO}^{PC}$, the monopolist would use the Outside Option;*
- *if $\alpha_{OO}^{PC} < \alpha < \alpha_{PC}^{IR}$, the monopolist would use Partial Coverage;*
- *if $\alpha_{PC}^{IR} < \alpha < \alpha_{IR}^{FC}$, the monopolist would use Immediate Revelation;*
- *if $\alpha_{IR}^{FC} < \alpha < 1$, the monopolist would use Full Coverage.*

Moreover,

- *If $\lambda_G > \lambda_B$, then Π is smooth, $\alpha_{OO}PC$ is defined in (7), and α_{PC}^{IR} is defined by $r\Pi^-(\alpha) = \frac{1}{h(\bar{c})}$, α_{IR}^{FC} is defined by $rY(\alpha) = M(\alpha) - s + \frac{1}{h(\bar{c})}$.*
- *If $\lambda_B > \lambda_G$, then Π is non-smooth only at α_{OO}^{PC}. α_{IR}^{FC} is defined by $r\Pi^+(\alpha) = M(\alpha) - s + \frac{1}{h(\bar{c})}$, α_{PC}^{IR} is defined by $Y(\alpha) = \frac{1}{h(\bar{c})}$, and α_{OO}^{PC} is defined by $\Pi(\alpha_{OO}^{PC}) = 0$.*

Figure 6 illustrates the monopolist's optimal switching results. We can also apply the analysis in Section 6.1 to get the results when $\lambda_B = 0$. The only difference is that FC switches to NB at some cutoff $\alpha_{FC}^{NB} \in (\alpha_{IR}^{FC}, 1)$.



(a) When $\lambda_G > \lambda_B > 0$ (b) When $\lambda_G > \lambda_B = 0$

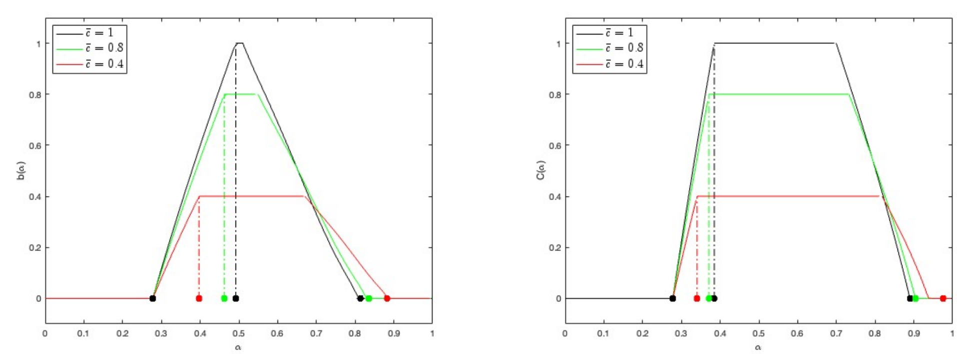
Figure 6: Optimal Switching Monopolist with IR

We can apply the same methods in Section 5 to establish the social planner's optimal strategy when $\bar{c}$ is small. The intuition is as before: the social planner has a lower cost of motivating reports, hence is more willing to experiment. Figure 7 illustrates this point by showing for the same $\bar{c}$, right panel is point wise weakly higher than the left panel, and right panel has larger learning region than the left panel.

We can also establish comparative statics with respect to $\bar{c}$. Proposition 7 clearly states that smaller $\bar{c}$ expands the IR region, and Figure 7 also illustrates this point, by showing for each panel, smaller $\bar{c}$ generates a wider IR region. Intuitively, consider the case when $\lambda_G > \lambda_B > 0$, we know FC region has right end 1 and PC region has fixed left end α_{OO}^{PC}, so expansion of IR leads to shrinkage of both PC and FC regions. Our Proposition 8 establishes an extreme limit result, that under Assumption 4, as $\bar{c} \to 0$, PC and FC disappears. We set $\lambda_B = 0$ in Figure 7 to show that NB (no learning region) also shrinks as $\bar{c}$ decreases.

Assumption 4. $\lim_{\bar{c}\to 0} h(\bar{c}) = \infty$.

Remarks Notice we assume $h(x) > 0, \forall x \in [0, \bar{c}]$ in the main context, so $H^2(\bar{c})/h(\bar{c})$ is bounded above. Notice also that $\beta(\cdot)$ increasing is not enough to guarantee above condition. A counter example is $h(x) = \frac{3}{2} - x$ and $H(x) = \frac{3}{2}x - \frac{1}{2}x^2, \bar{c} = 1$, β is still increasing, but $h(\bar{c}) \equiv 0.5 > 0$. We actually know that the average mass of $h(\cdot)$ over $[0, \bar{c}]$ is $\frac{1}{\bar{c}}$ which goes to ∞, so a sufficient condition for Assumption 4 is that $h(\cdot)$ is weakly increasing, and the numerical example we use is $h(b) = \frac{1}{\bar{c}}$ which satisfies this assumption. If $0 < \lim_{\bar{c}\to 0} h(\bar{c}) < \infty$, then the optimal strategy is fully captured by Proposition 7, and PC and FC regions are still present.



(a) Monopolist (b) Social Planner

Figure 7: Reporting cutoffs with small $\bar{c}$

Proposition 8 says, under Assumption 4, when $\bar{c} \to 0$, the Partial Coverage, Full Coverage, and Non Bonus regions vanish entirely.

Proposition 8. *When $\bar{c} = 0$, there exists a smooth value function $\Pi(\alpha)$. There exist cutoff $0 < \alpha_{OO}^{IR} < 1$, where $\alpha_{OO}^{IR} = \frac{sr}{\lambda(z_G - s) + rz_G}$, such that:*

- *if $0 < \alpha < \alpha_{OO}^{IR}$, the monopolist would use the Outside Option, and*

- *if $\alpha_{OO}^{IR} < \alpha < 1$, the monopolist would use immediate revelation.*

An immediate corollary is that the social planner would follow the same switching rule. As established above, the social planner always has a slightly greater willingness to experiment than the monopolist (and is bounded above by $\bar{c} = 0$) and terminates experimentation at the same cutoff; therefore, the above switching rule and cutoff also work for the social planner. Because the social planner and monopolist switch from the Outside Option to immediate revelation at the same cutoff and, within the immediate revelation region, the learning speed is maximized under both settings, we conclude that when $\bar{c} = 0$, the monopolist also maximizes social surplus: maximal learning speed brings efficiency.

This limit result is consistent with the literature. Bergemann and Välimäki (2000) consider a model with long-lived agents, a monopolist seller of the risky arm competing with one seller of the safe arm, and automatic disclosure of every signal after purchasing[12]. They show that there is a socially efficient stopping level $\hat{\alpha}$ at which the risky arm is myopically dominated by the safe arm, and sellers may engage in excess learning and may sell the risky arm with negative profits. As reporting is costless, the first-best allocation is to have all agents buy the risky arm when $\alpha > \hat{\alpha}$ and all agents buy the safe arm when $\alpha < \hat{\alpha}$. This result can be treated as a special case of our results by setting $\bar{c} = 0$.

When $\bar{c} > 0$, our result can be viewed as an interior solution: if there is reporting cost with support $[0, \bar{c}]$, the socially efficient allocation is, under moderate levels of belief $[\alpha_{OO}^{PC}, \alpha_{FC}^{NB}]$, to screen out the agents with lower reporting costs for experimentation on the risky arm (and report), while allocating the other agents with the myopically dominant arm.

B.2 Proof of Proposition 7

Proof. In FC and PC region, FOC and simplified HJB is given by equation (15) and (14) respectively, however in IR region $b = \bar{c}$ is not from FOC.

When $\lambda_G > \lambda_B$, as before, we consider smooth pasting upwards, since the smooth belief movement $\alpha' - \alpha < 0$. So we first pin down the PC region by solving (14) together with boundary condition $\exists \alpha_{OO}^{PC}, s.t. \Pi(\alpha_{OO}^{PC}) = \Pi'(\alpha_{OO}^{PC}) = 0$, we get (7). Then we solve IR region by solving (31) together with smooth pasting condition at α_{PC}^{IR} (where bonus is $\bar{c}$). However we can simply get α_{IR}^{PC} by solving $r\Pi(\alpha_{IR}^{PC}) = H^2(\bar{c})/h(\bar{c}) = 1/h(\bar{c})$. Since Π on the left region of α_{IR}^{PC} is known, above equation pins down α_{IR}^{PC}, and we can use the boundary condition for Y as $Y(\alpha_{IR}^{PC}) = \Pi(\alpha_{IR}^{PC})$ to pin down Y. Similarly smooth pasting condition at α_{IR}^{FC} gives us $rY(\alpha_{IR}^{FC}) = M(\alpha_{IR}^{FC}) - s + 1/h(\bar{c})$, and we can use boundary condition $Y(\alpha_{IR}^{FC}) = \Pi(\alpha_{IR}^{FC})$ to solve FC region.

When $\lambda_B > \lambda_G$, as before we start from above. So we first pin down FC region and then pin down α_{IR}^{FC} by $r\Pi(\alpha_{IR}^{FC}) = M(\alpha_{IR}^{FC}) - s + 1/h(\bar{c})$. Then we get Y, and pin down PC region by $rY(\alpha_{PC}^{IR} = \frac{1}{h(\bar{c}}$. We finally pin down Π on PC region and α_{OO}^{PC} by $\Pi(\alpha_{OO}^{PC}) = 0$.

[12] Unlike Bergemann and Välimäki (2000) we do not consider long-lived agents. However, because in our model there is no information asymmetry and information is aggregated via social belief, the long-lived-agents setting does not make a difference.

$\square$

B.3 Proof of Proposition 8

Proof. When Assumption 4 holds, we know when $\bar{c} \to 0$, $1/h(\bar{c}) \to 0$.

The last two points of Proposition 7 show that smooth pasting conditions at α_{PC}^{IR} requires $r\Pi(\alpha) = rY(\alpha) = 1/h(\bar{c}) \to 0$, which coincides with the boundary condition at α_{OO}^{IR}, so PC region shrinks. Similarly boundary condition at α_{IR}^{FC} requires $rY(\alpha_{IR}^{FC}) = M(\alpha_{IR}^{FC}) - s + 1/h(\bar{c}) \to M(\alpha_{IR}^{FC}) - s$, which coincides with the boundary condition at α_{FC}^{NB} (we know $\alpha_{FC}^{NB} = 1$ when $\lambda_B > 0$ from Lemma 4.1). So in conclusion, we know when $\lambda_B > 0$, FC and PC disappears when $\bar{c} \to 0$, and NB disappears naturally.

We then show that NB also disappears when $\lambda_B = 0$. We denote $\lambda_G = \lambda$. We know $Y(\alpha)$ can be uniquely determined by (31) and the boundary condition

$$Y(\alpha_{OO}^{IR}) = 0$$

Notice that Equation (31) has a closed-form general solution:

$$Y(\alpha) = \frac{\alpha\lambda z - s}{r} + C \cdot \alpha^{-\frac{r}{\lambda}}(1-\alpha)^{\frac{r+\lambda}{\lambda}}, \tag{34}$$

where C can be uniquely identified by boundary condition $Y(\alpha_{OO}^{IR}) = 0$:

$$C = -\frac{\frac{\alpha_{OO}^{IR}\lambda G - s}{r}}{(\alpha_{OO}^{IR})^{-\frac{r}{\lambda}}(1-\alpha_{OO}^{IR})^{\frac{r+\lambda}{\lambda}}}$$

with

$$\alpha_{OO}^{IR} = \frac{rs}{\lambda(z_G - s) + rz_G}$$

The fact that $\alpha_{OO}^{IR}\lambda G < s$ guarantees $C > 0$, hence $Y(\alpha) \geq \frac{\alpha\lambda G - s}{r}$ and the inequality is tight only if $\alpha = 1$. That is to say, the value function of adopting the IR strategy, $Y(\alpha)$, is strictly higher than NB as long as $\alpha < 1$, so in the limit, the latter does not appear.

In conclusion, when $\bar{c} \to 0$, as long as Assumption 4 holds, the optimal switching rule for the monopolist is

$$OO \to IR$$

, and the switch happens at

$$\alpha_{OO}^{IR} = \frac{rs}{\lambda(z_G - s) + rz_G}$$

$\square$

B.4 A Numerical Example: Social Planner with Large $\bar{c}$

Notice that in the social planner's value function, we do not need to take s into consideration. However, to compare their value function with that of the monopolist, we shift the social planner's value function downward by $\frac{s}{r}$ in the end.

We use U, V, Y for the social planner's value function W, restricted on the PC, FC, and IR regions.

If $\alpha < \alpha_{PC}^{FC}$, we have

$$rU(\alpha) - s = \int_{0<c<C_1(\alpha)} B(\alpha) + M(\alpha) - s - cdH(c) = \frac{[B(\alpha) + M(\alpha) - s]^2}{2\bar{c}}$$

Hence, we can obtain the following motion law:

$$U'(\alpha) = \frac{s - M(\alpha) - \frac{\alpha\lambda_G z_G + (1-\alpha)\lambda_B s}{r} + \lambda(\alpha)U(\alpha) + \sqrt{2\bar{c}(rU(\alpha) - s)}}{(\alpha^2 - \alpha)(\lambda_G - \lambda_B)}, \tag{35}$$

with boundary condition

$$U(\alpha_{OO}^{PC}) = \frac{s}{r}. \tag{36}$$

Similarly, when $\alpha > \alpha_{PC}^{FC}$, we have

$$V'(\alpha) = \frac{\lambda(\alpha)V(\alpha) - \frac{\alpha\lambda_G z_G + (1-\alpha)\lambda_B s}{r} + \sqrt{2\bar{c}(rV(\alpha) - M(\alpha))}}{(\alpha^2 - \alpha)(\lambda_G - \lambda_B)}, \tag{37}$$

with a boundary condition requiring

$$V(\alpha_{PC}^{FC}) = U(\alpha_{PC}^{FC}) \tag{38}$$

The social planner's value function is depicted in Figure **??**. For purposes of comparison with the monopolist's producer surplus, we shift the social planner's function downward by a constant $\frac{s}{r}$. The social planner's thresholds are depicted in Figure **??**. Both use the same parameters as in the monopolist setting.