Strategic Experimentation with Two-sided Private Information

Yihang Zhou[a,b,1,*]

[a] *Department of Economics, The University of Texas at Austin*
[b] *2225 Speedway, BRB 1.116, C3100, Austin, Texas 78712, United States*

Abstract

I study a symmetric two-player game of strategic experimentation where both players have private information. I find that two-sided private information improves welfare, both at the ex ante and interim stages, by mitigating the free-rider problem. Furthermore, in some states of the world, there may be over-experimentation, i.e., players may experiment more than the social planner would under complete information.

Keywords: Strategic experimentation, two-arm bandit, two-sided private information

*Corresponding author
Email address: yhzhou@utexas.edu (Yihang Zhou)
[1] Declarations of interest: none

1. Introduction

This paper studies a symmetric two-player game of strategic experimentation, as in Keller, Rady and Cripps (2005), where both players have private information. The players face a two-arm bandit problem, and the payoffs to each of these arms is common to the players. I model private information by allowing the players to be uncertain about the payoffs from both arms, but active learning to be possible only on one of the arms, the *learning arm*. That is, if a player uses this arm, it generates public lump-sum payoffs via a Poisson process, where the arrival rate is positive only in the good state. A player can also use the *no-learning arm*, which generates a constant flow payoff, which is observed by the players only in the distant future. However, each player has private information about the no-learning arm's payoff, and the players' combined information resolves the uncertainty. Because a player's experimentation strategy depends on his private information, and may reveal it, a problem of strategic experimentation with two-sided signaling arises.

I find that there exists a Markov perfect equilibrium (MPE) when the players experiment more than in the full-information benchmark (i.e., the situation where the players' private information is made public). This increased experimentation happens because of two reasons. Firstly, the use of the learning arm signals a lower flow payoff in the no-learning arm, which can encourage the other player to experiment more. Secondly, neither player knows the no-learning arm's payoff, so they may experiment more than necessary. For example, one possible situation is that players will stop experimenting if they know the true flow payoff of the no-learning arm. However, because private information exists, both players' beliefs about the flow payoff may be lower than the truth, leading to continued experimentation. Indeed, one may have over-experimentation, where the players experiment more than a utilitarian social planner would under complete information.

Despite the possibility of over-experimentation due to two-sided private information, the ex ante welfare is larger than under the full-information benchmark.[2] While over-experimentation can reduce welfare in some states, increased experimentation can mitigate the free-riding problem, resulting in increased ex ante welfare. Furthermore, the interim expected payoffs of the players, after both of them learn their private information, always exceed the payoffs in the full-information case. In contrast, in Dong (2021), the informed player of a certain type may be hurt by his private information.

The modeling innovation in my paper is the introduction of a risky no-learning arm. For example, consider passive investment in a financial asset, such as shares in a long-term investment project whose outcome will be known only in the distant future. Because the project has been undertaken and the purchase of an asset is only a transfer in ownership, such an investment does not result in any learning, but investors may have private information on the future prospects of the investment project.[3]

Related literature.

Bolton and Harris (1999) first introduced the strategic experimentation problem and characterized the unique symmetric MPE. Keller, Rady and Cripps (2005) introduced the exponential bandit model and constructed a unique symmetric MPE with under-experimentation due to free riding, as well as studying asymmetric equilibria. Keller and Rady (2010) generalized this model to the case of inconclusive good-news information, and Keller and Rady (2015) studied the bad-news case. These papers discussed the encouragement and free-riding effects in strategic experimentation and found that MPEs feature under-experimentation relative to the collaborative benchmark.

The classic analysis of strategic experimentation assumes that all information is public; however, some authors allow private information of players in strategic experimentation problems. Bonatti and Hörner (2011) discussed the effort invested in experimentation when it is unobserved by other players. Hei-

[2]Ex ante welfare is the expected payoff of a player before observing her private information.

[3]In my formal model, I assume that the sum of the players' information resolves all uncertainty regarding the no-learning arm, but this assumption is inessential.

dhues, Rady and Strack (2015) assumed unobserved realization of payoffs on the risky arm in the two-arm bandit model. The paper most closely related to mine is Dong (2021). The author allows one of the two players in an exponential bandit model to have an extra private signal about the risky arm's state. She found that one-sided private information mitigates the under-experimentation problem without the risk of over-experimentation, which is the novel feature of her paper.

Additionally, my work is also related to other dynamic models with two-sided private information. For example, in the bargaining model with two-sided uncertainty by Cramton (1992), two players may waste time bargaining with each other only to realize that a transaction cannot be made, because the seller values the item more than the buyer does. Cho (1990) also studied a two-sided uncertainty model, but he assumed a finite horizon and one-sided offer bargaining. He also found a delay in the agreement between the two sides. Cronshaw and Alm (1995) studied the model of the government and a taxpayer, where both sides have private information. They found that the government using a concealment policy led to less compliance. Other similar results are presented in Kahn and Huberman (1988) and Banks (1993). In contrast with much of this work, in my model the overall effect of two-sided uncertainty is beneficial, because it improves the ex ante welfare by mitigating the free-riding problem.

The rest of this paper is organized as follows. Section 2 sets out the model. Section 3 presents the main results regarding Markov perfect equilibria, and Section 4, the welfare results. Section 5 discusses multiplicity of equilibria, and Section 6 concludes the paper.

2. Model

Time is continuous, and the horizon is infinite. Two players (player 1 and 2) have access to replicas of a two-arm bandit. Each player has one unit of perfectly divisible resource per unit of time to split between two arms. Furthermore, players discount future payoffs by the factor r.

2.1. Two-arm bandit

The learning arm's payoff depends on the unobserved state θ, which players do not observe. The learning arm generates a payoff of $h \in \mathbb{R}^+$ according to a Poisson process with the arrival rate λ when the state is $\theta = 1$. No payoff is generated when the state is $\theta = 0$, so a perfect good-news model is achieved. The payoff arrivals of the learning arms are public; therefore, a player can free ride the other one's experimentation.

The no-learning arm generates flow payoff $s = s_1 + s_2$, which is not observed either. Here $s_1, s_2 \in \{\underline{s}, \overline{s}\}$ are random variables drawn independently at the beginning of the game, and both have the probability of $q_0 \in (0,1)$ being $\underline{s}$, and the probability of $1 - q_0$ being $\overline{s}$ ($0 < \underline{s} < \overline{s}$). Player i observes only s_i, which is his private type. I also assume that the flow payoff s is not observed until date T, which is distant in the future, so players do not learn by using this arm. Instead, a player can infer the other one's private information from actions in equilibrium. Moreover, I assume $2\overline{s} < \lambda h$, so when $\theta = 1$, players prefer the learning arm, regardless of the realizations of s_1 and s_2.

2.2. Strategies and equilibrium

At the beginning, nature decides s_1 and s_2, and player i observes the realization of s_i. At date t, the public information includes the past actions of the two players and the past payoff outcomes on the learning arms. The space of the past actions is $[0,1]^{[0,t)}$, and the space of the past outcomes is $\{0,1\}^{[0,t)}$, where 1 represents the payoff arrival. At date t player i decides the resource $k_t \in [0,1]$ on the learning arm based on the public information and his private type. Player i's strategy at date t is represented by

$$k_t^i : \{\underline{s}, \overline{s}\} \times ([0,1]^{[0,t)})^2 \times (\{0,1\}^{[0,t)})^2 \to [0,1],$$

which is the resource on the learning arm.

Following the literature, I focus on the Markov perfect equilibrium (MPE). At date t, the sufficient state variables of player i consist of $p_t \in [0,1]$, the belief of the learning arm's state being $\theta = 1$; $q_t^1 \in [0,1]$, player 2's belief of player 1's type being $\underline{s}$; $q_t^2 \in [0,1]$, player 1's belief of player 2's type being $\underline{s}$; and $s_i \in \{\underline{s}, \overline{s}\}$, player i's private type. Variables p_t, q_t^1, and q_t^2 are common knowledge, because the past actions and past payoff outcomes on the learning arms are public information. Therefore, the sufficient state space for player i is $[0,1]^3 \times \{\underline{s}, \overline{s}\}$. The sufficient state variables for the two players are different, because the players can observe only their own type.

I define the pure Markov strategies for players following Dong (2021). Consider two players: player i and player j. Let s_i be player i's private type, q_t^i be player j's belief about player i's type, and q_t^j be player i's belief about player j's type. Player i's strategy is $k^i(p, q^j; p_t, q_t^i, q_t^j, s_i)$, which says that when the state is (p_t, q_t^i, q_t^j) at some point, type s_i plays according to $k^i(p, q^j; p_t, q_t^i, q_t^j, s_i)$ for any $p \leq p_t$, and q^j is type s_i's belief about the other player's type.[4] The strategy is Markov if any history causing the state (p_t, q_t^i, q_t^j) has the same action. For example, $\forall p_t, p_t'$ s.t. $p_t > p_t'$, starting from state (p_t, q_t^i, q_t^j), if the two players play according to $k^i(p, q^j; p_t, q_t^i, q_t^j, s_i)$ and $k^j(p, q^i; p_t, q_t^j, q_t^i, s_j)$ and the state reaches $(p_t', q_t^{i\prime}, q_t^{j\prime})$, then the two players' strategies satisfy $k^i(p, q^j; p_t, q_t^i, q_t^j, s_i) = k^i(p, q^j; p_t', q_t^{i\prime}, q_t^{j\prime}, s_i)$ and $k^j(p, q^i; p_t, q_t^j, q_t^i, s_j) = k^j(p, q^i; p_t', q_t^{j\prime}, q_t^{i\prime}, s_j)$ for $p \leq p_t'$. For simplicity, I use k_t as the action at date t when there is no confusion. Finally, following the literature, I require the strategy to be left continuous and piecewise Lipschitz continuous in p_t, to avoid further problems (Keller and Rady, 2010).

Besides the pure strategies, I also allow for a specific kind of mixed strategies: players can randomly stop experimentation and move to the no-learning arm (i.e., mixing stopping experimentation and continuing experimentation at a date). More precisely, I assume players can choose the arrival rates of the Poisson process for the time they stop experimentation. This mixed strategy is

[4] A simpler version of Markov strategies is $k^i(p_t, q_t^i, q_t^j; s_i)$, but the player's action is affected by q_t^i at the last moment, and vice versa, which can cause problems.

also equivalent to having a random variable drawn from $[0,1]$ at the beginning of the game and having the player choose the stopping time according to the realization of the random variable.

An MPE consists of a strategy function $k^i(p, q^j; p_t, q_t^i, q_t^j, s_i)$ and belief updating $\mu(k_j, p_t, q_t^i, q_t^j)$, which is the new q_t^i when the state is (p_t, q_t^i, q_t^j) and player i observes an action of k_j from the other player. At each date, the strategy $k^i(p, q^j; p_t, q_t^i, q_t^j, s_i)$ maximizes the expected payoff of player i,

$$E\Big[\int_t^{\infty} re^{-rs}[(1-k)(s_1+s_2)+k\lambda h p_t]dt\Big],$$

given the other player's strategy and belief updating. Additionally, the belief updating should satisfy Bayes' rule when possible.

Another restriction I put on the equilibria is that after both players reveal their private information,[5] they will play the unique symmetric equilibrium presented in Keller, Rady and Cripps (2005), referred to as the KRC strategy henceforth.

2.3. State variables

Variable p_t is the belief about the state of the learning arm, and its updating is the same as in Keller, Rady and Cripps (2005). If there is payoff arrival on the learning arm, then the two players know $\theta = 1$. If there is no payoff arrival, the update of p_t follows the law of motion:

$$\frac{dp_t}{dt} = -K_t \lambda p_t (1-p_t),$$

where K_t is the two players' total resource at date t on the learning arms. The belief of the learning arm p_t decreases with time.

The updates of q_t^1 and q_t^2 depend on the strategies in equilibrium. If player i of type $\underline{s}$ and type $\overline{s}$ choose the same action (i.e., pooling), then q_t^i does not change. Otherwise, it updates according to the Bayes rule. For example, suppose in the whole history, type $\underline{s}$ chooses $k_t = 1$ and type $\overline{s}$ jumps to $k_t = 0$

[5] In this case, the model becomes the same as in Keller, Rady and Cripps (2005).

from 1 at an arrival rate e_t; then the belief of type $\underline{s}$ at time t, seeing only $k = 1$ in the past, is

$$q_t^i = \frac{q_0}{q_0 + (1 - q_0)e^{-\int_{m=0}^{t} e_m dm}}.$$

Off-path beliefs will follow the belief updating $\mu(k_j, p_t, q_t^i, q_t^j)$ in equilibrium.

3. Equilibria Results

3.1. The cooperative solution and benchmark

If I assume the two players share the private information when working cooperatively, the model degenerates into the model in Keller, Rady and Cripps (2005) (the KRC model). As a result, the cooperative solution of my model is the same as that of the KRC model. The players will invest all resources into the learning arm when $p_t > p_c(s)$, and into the no-learning arm when $p_t \leq p_c(s)$, where the threshold $p_c(s)$ is a function of the realization of the flow payoff in the no-learning arm:

$$p_c(s) = \frac{rs}{(\lambda h - s)(r + 2\lambda) + rs}.$$

My model nests on the KRC model so that it will be the benchmark. Specifically, I use the unique symmetric MPE in the KRC model as the benchmark, because I also look at a symmetric MPE. Free riding leads to under-experimentation in this solution. Figure 1 illustrates the KRC strategy[6] and the cooperative threshold $p_c(s)$.

3.2. Equilibrium with over-experimentation

With $\overline{s}$ large enough, $\underline{s}$ small enough, and q_0 large enough, I construct an equilibrium with possible over-experimentation. Here $\overline{s}$ and $\underline{s}$ should be large and small enough, respectively, so that at $p_t = p_c(\underline{s} + \overline{s})$, the KRC strategy of $s = 2\underline{s}$ is $k_t = 1$; at $p_t = p_c(2\overline{s})$, the KRC strategy of $s = \underline{s} + \overline{s}$ is $k_t = 1$ (Figure 2).

[6]The strategy is presented as a function of p_t.

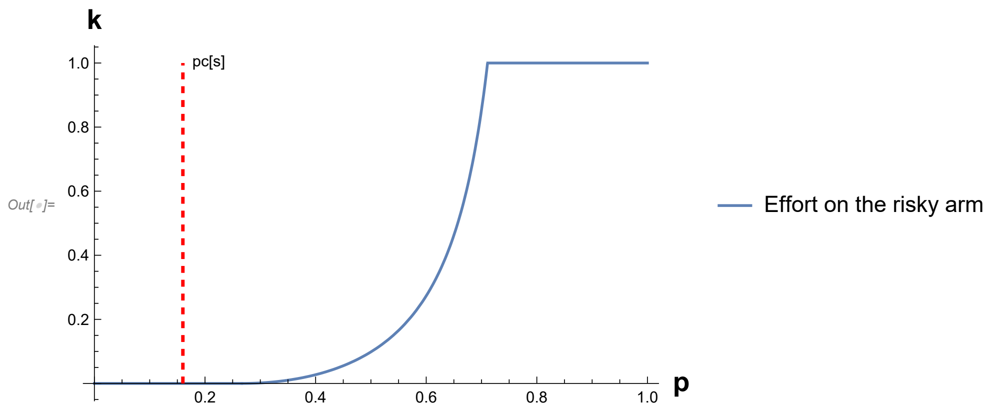


Figure 1: The KRC strategy (the blue line) and the cooperative threshold (the red dotted line)

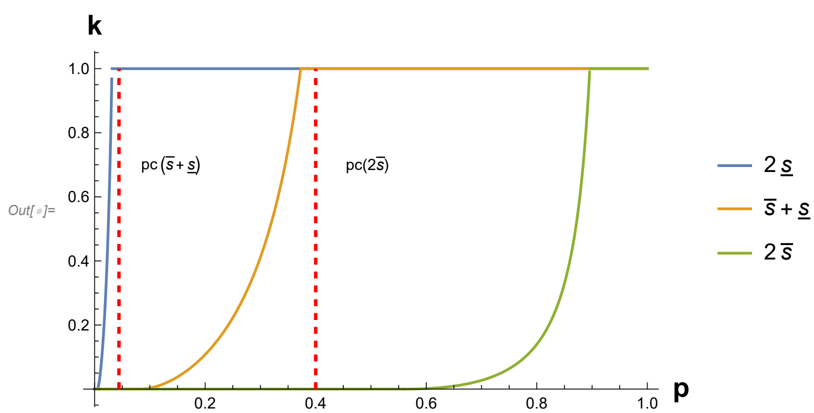


Figure 2: The KRC strategies of three possible realizations of s

Clearly, if the learning arm's payoff arrives, the players know $\theta = 1$ and will put all their resources into the learning arm. Next, I will discuss the equilibrium strategies without payoff arrival.

This equilibrium is characterized by two cutoff points, $p_c(\underline{s}+\overline{s})$ and p_s, where $p_c(\underline{s}+\overline{s}) < p_s < p_c(2\overline{s})$. The following strategies are conditional on no payment arrival on the learning arm and one player's beliefs about the other player's type (q_t^1, q_t^2) staying on path (discussed later).

Suppose both beliefs (q_t^1, q_t^2) stay on path:

When $p_t > p_s$, both types experiment with full resources, so q_t^i, q_t^j will stay at q_0. This is *the pooling stage.*

When $p_c(\underline{s}+\overline{s}) < p_t \leq p_s$, type $\underline{s}$ experiments with resource $k = 1$ while type $\overline{s}$ randomly stops; the two players' strategies are partially separating. More specifically, during the time interval $(t, t+dt)$, type $\overline{s}$ stops experimenting with probability, $e_t dt$[7] i.e., the stopping of experimentation happens with an arrival rate of e_t, which may change with time (as detailed in the next subsection). If one player stops experimentation, the other player knows that the first player's type is $\overline{s}$; if experimentation does not stop, q_t^j increases with time (it also decreases in p_t, because p_t decreases with time). This is *the partial separating stage.*

When $p_t \leq p_c(\underline{s}+\overline{s})$, the two players' strategies will be fully separating. Type $\underline{s}$ plays the KRC strategy of $2\underline{s}$ if $(q_t^i, q_t^j) = (1, 1)$, which is the on-path belief. I will show later that this is the only possible on-path case. This stage is fully separating, because type $\overline{s}$ will reveal himself before p_t falls below $p_c(\underline{s}+\overline{s})$. For the off-path cases, if $(q_t^i, q_t^j) = (1, 0)$, player i of type $\underline{s}$ plays $k = 0$, which is the KRC strategy of $s = \underline{s}+\overline{s}$; if (q_t^i, q_t^j), where $q_t^i < 1$, player i of type $\underline{s}$ plays $k = 1$ to reveal himself; if (q_t^i, q_t^j), where $q_t^i = 1$, $0 < q_t^j < 1$, player i of type $\underline{s}$ plays $k = 1$ to keep $q_t^i = 1$ and wait for the other player to disclose his type. Type $\overline{s}$ does not experiment in this stage. The fully separating of the players' strategies can also be explained as follows: type $\overline{s}$ always plays $k = 0$ and type

[7] Here $(dt)^n$ with a power no less than 2 is omitted.

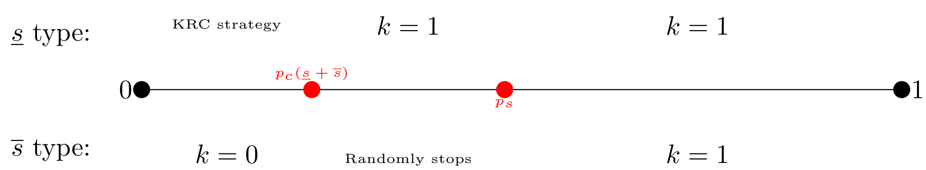


Figure 3: On-path strategies

$\underline{s}$ plays $k = 1$ when his type is not known by the other player in this stage.[8] This is *the full separating stage*. Figure 3 summarizes these strategies.

It is possible to have the on-path beliefs q_t^i when there is no stopping of the experimentation or no payoff arrival on the learning arm as a function of p_t by the Bayes rule:

$$q_i(p_t) = \begin{cases} q_0, & p_t > p_s \\ \dfrac{q_0}{q_0 + (1 - q_0)e^{-\int_{m=0}^{t} e_m dm}}, & p_c(\underline{s} + \overline{s}) < p_t \leq p_s \\ 1, & p_t \leq p_c(\underline{s} + \overline{s}) \end{cases}$$

which is also presented in Figure 4. When player i stops in the partial revealing stage, q_t^i jumps to 0. I will show in the next subsection that $q_i(p_t)$ reaches 1 when p_t arrives at $p_c(\underline{s} + \overline{s})$.

As for off-path beliefs, I will assign $q_t^i = 0$ for any k_t smaller than the strategy of type $\underline{s}$.

After one player reveals his private signal, the problem becomes a one-sided private information problem similar to that in Dong (2021). A detailed construction of the equilibrium of this subgame is presented in section 3.4 and Appendix A.2.

Finally, after both players reveal their private signal, the problem becomes the same as in Keller, Rady and Cripps (2005), and I will assume the players play the unique symmetric MPE (the KRC strategies).

In this equilibrium, over-experimentation may happen. When the true state

[8] When p_t is very small, a player does not experiment regardless of the other player's type, so there is no need to distinguish this situation from the full-separation case.

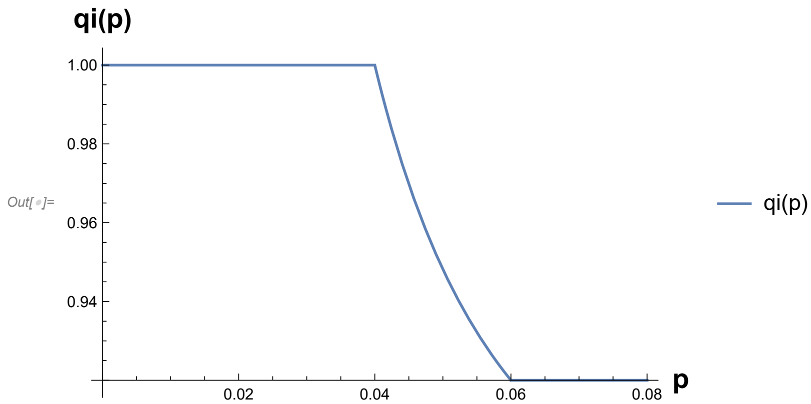


Figure 4: On-path belief

is $s = 2\overline{s}$, the outcome of this equilibrium is as follows:[9]

Suppose the initial prior about the learning arm is $p_0 > p_s$. At the beginning, both players play $k_t = 1$. As p_t goes down and reaches p_s, both players (type $\overline{s}$) randomly stop experimentation at an arrival rate e_t. With probability 1, the stopping of experimentation happens before p_t reaches $p_c(\underline{s} + \overline{s})$. Then, because $p_t \leq p_s < p_c(2\overline{s})$, both players use the no-learning arm according to the equilibrium in the one-sided private information problem (see section 3.4).

Both players will experiment at $p_t < p_c(2\overline{s})$ because $p_s < p_c(2\overline{s})$. However, working cooperatively, they will stop experimentation at $p_c(2\overline{s})$. There is over-experimentation because of two reasons. The first is that experimenting signals a lower opportunity cost of experimentation, leading to more experimentation from the other player. This encouragement effect mitigates the free-riding problem and causes more experimentation but does not lead to over-experimentation. The second reason is the ignorance towards the true s. In this example, no player knows the true state $s = 2\overline{s}$. Instead, each player believes that the other player might be type $\underline{s}$. Since both players think the opportunity cost of experimenting might be low, they will experiment more than needed. They know they have

[9] I will discuss the outcome when no payoff arrives on the learning arm, because the result is trivial with payoff arrival.

over-experimentation if the state is $2\overline{s}$, but the expected gain from other possible realizations of s can cover the loss from over-experimentation. Therefore, both players continue to experiment, even if they might pass the efficient threshold. In this equilibrium, selfishness will cause more experimentation, but only with ignorance toward the true s will the experimentation level exceed the efficient level.

The equilibrium above is summarized in the following proposition.

Proposition 1. *If there exist $c_1(r,\lambda,h)$, $c_2(r,\lambda,h)$ s.t. for any $\overline{s} \in (c_1(r,\lambda,h), \frac{\lambda h}{2})$ and $\underline{s} \in (0, c_2(r,\lambda,h))$, $\exists\ \underline{q}(\overline{s},\underline{s},r,l,h) < 1$ s.t. with any $q_0 > \underline{q}(\overline{s},\underline{s},r,\lambda,h)$ and $p_0 > p_s$,[10] it is possible to construct an equilibrium with over-experimentation featured by cutoff points p_s and $p_c(\underline{s}+\overline{s})$:*
(1) Type $\underline{s}$ plays $k_t = 1$ if $p_t > p_c(\underline{s}+\overline{s})$; it plays the KRC strategy of the true s if $p_t \leq p_c(\underline{s}+\overline{s})$.
(2) Type $\overline{s}$ plays $k_t = 1$ if $p_t > p_s$; it plays $k_t = 0$ if $p_t \leq p_c(\underline{s}+\overline{s})$; and it randomly stops to the no-learning arm if $p_s < p_t \leq p_c(\underline{s}+\overline{s})$.
(3) The on-path belief when the stopping of experimentation has not happened is $q_i(p_t)$.
(4) The off-path belief with a deviation k_t smaller than the strategy of type $\underline{s}$ is 0.

3.3. Construction of the partial separating stage

In this section, I discuss the stage of $p_t \in (p_c(\underline{s}+\overline{s}), p_s]$. In this stage, type $\underline{s}$ uses the learning arm with his full resource, and type $\overline{s}$ randomly stops at an arrival rate of e_t.

If a player is type $\underline{s}$ and deviates to any $k < 1$, then the other player will think that he is type $\overline{s}$ and thus experiment less because of the higher opportunity cost of experimentation. As a result, this deviation hurts type $\underline{s}$.

Because the type $\overline{s}$ player randomly stops, he needs to be indifferent between continuing on $k_t = 1$ and revealing his type by $k_t = 0$. Suppose that type $\overline{s}$'s

[10] As discussed in section 3.3, p_s depends on q_0.

expected payoff with no one revealing his type in this stage is $u(p_t, q_i(p_t))$; then it should satisfy the following Hamilton–Jacobi–Bellman equation:

$$\begin{aligned} u(p_t, q_i(p_t)) =& p\lambda h + \frac{1}{r}\Big\{(1 - q_i(p_t))e_t[2\overline{s} - u(p_t, q_i(p_t))] \\ &+ [2p\lambda(\lambda h - u(p_t, q_i(p_t))) + \frac{du(p_t, q_i(p_t))}{dp_t}\frac{dp_t}{dt}]\Big\}. \end{aligned} \tag{1}$$

Here $(1 - q_i(p_t))e_t dt$ is the probability of the other player being type $\overline{s}$ and revealing his type by $k_t = 0$ during the interval dt.[11]

Furthermore, the expected payoff of type $\overline{s}$ should equal the expected payoff of revealing his type by $k = 0$, because a player with type $\overline{s}$ is indifferent between mimicking type $\underline{s}$ and revealing his type in this stage. I will make p_s at least small enough so that the KRC strategy of $s = \underline{s} + \overline{s}$ is $k = 0$ at p_s, which can be achieved by assuming q_0 is large enough (discussed later). Under this assumption, type $\overline{s}$'s expected payoff of the partial separating stage will be

$$u(p_t, q_i(p_t)) = q_i(p_t)(\underline{s} + \overline{s}) + (1 - q_i(p_t))2\overline{s} \tag{2}$$

and thus

$$\frac{du(p_t, q_i(p_t))}{dp_t} = (\underline{s} - \overline{s})q_i'(p_t). \tag{3}$$

The above equations hold because when $p_t \le p_s$ and one player reveals himself as type $\overline{s}$, the other player will reveal his type in the next moment. See section 3.4 and Appendix A.2 for the detailed reason.

Notice that $q_i(p_t) = \frac{q_0}{q_0 + (1-q_0)e^{-\int_{m=0}^{t} e_m dm}}$, so

$$e_t = \frac{q_i'(p_t)}{q_i(p_t)(1 - q_i(p_t))}\frac{dp_t}{dt}. \tag{4}$$

By plugging (2), (3), and (4) into (1) we get

$$2\overline{s} + q_i(p_t)(\underline{s} - \overline{s}) = p\lambda h + \frac{1}{r}\Big\{[p2\lambda(\lambda h - 2\overline{s} - q_i(p_t)(\underline{s} - \overline{s}))\Big\}, \tag{5}$$

which can give us the following:

$$q_i(p_t) = \frac{2\overline{s}r - [(r + 2\lambda)\lambda h - 4\lambda\overline{s}]p_t}{(\overline{s} - \underline{s})(r + 2\lambda p_t)}, \; p_c(\underline{s} + \overline{s}) < p_t \le p_s. \tag{6}$$

[11] Here dt to the power higher than 1 is omitted.

With the expression of $q_i(.)$ of $p_t \in (p_c(\underline{s}+\overline{s}), p_s]$, we can get the expression of e_t (or $e(p_t)$) by (4).

Moreover, $q_i(p_t)$ is decreasing in p_t, thus q_t^i, q_t^j are increasing in t. And at $p_t = p_c(\underline{s}+\overline{s})$, $q_i(p_t) = 1$, so the stopping of experimentation happens before p_t reaches $p_c(\underline{s}+\overline{s})$ with probability 1. In other words, if no stopping of experimentation is observed before p_t arrives at $p_c(\underline{s}+\overline{s})$, a player is sure that the other player is type $\underline{s}$. Once the stopping happens, the game becomes a one-sided private information game, which is discussed in section 3.4.

Cutoff point p_s is also decided by (6). It solves $q_i(p_t) = q_0$. We have $q_i(p_c(\underline{s}+\overline{s})) = 1$, $q_i(p_c(2\overline{s})) = 0$, $q_0 \in (0,1)$, and $q_i(p)$ decreases in p, so p_s is a uniquely determined value between $p_c(\underline{s}+\overline{s})$ and $p_c(2\overline{s})$. Assuming q_0 large enough can ensure that the partial separating stage happens late enough[12] to prevent players from deviating.

As for the intuition of having the partial separating stage, recall that for type $\overline{s}$, when the other player is type $\underline{s}$, the free-riding problem is mitigated and experimenting is beneficial; when the other player is type $\overline{s}$, then continuing to experiment can result in over-experimentation, and the player incurs a loss. When p_t is high, for type $\overline{s}$, the gain of experimenting is high enough to cover the loss of possible over-experimentation. However, as p_t goes down with time, the loss from over-experimentation becomes relatively large compared to the gain from other possibilities. When the gain can no longer cover the loss, type $\overline{s}$ begins to stop experimentation randomly. When this happens, the game proceeds as a one-sided private information problem. When the stopping of experimentation does not happen, q_t^i and q_t^j increase with time. Though the loss becomes relatively larger as time t increases, the probability of having a gain by experimenting (q_t^i) also increases. As a result, it can keep type $\overline{s}$ indifferent between mimicking type $\underline{s}$ by $k_t = 1$ and revealing his type by $k_t = 0$.

[12]We at least need p_s small enough so that the KRC strategy of $s = \underline{s}+\overline{s}$ is $k = 0$ at p_s, as mentioned before. The sufficient condition for q_0 to ensure the equilibrium is discussed in Appendix A.1.

3.4. One-sided private information problem

The one-sided private information problem after one player reveals his type is similar to the model in Dong (2021), so in this section I will omit the detailed algebra and focus on the main construction of the equilibrium where type $\overline{s}$ reveals his type. According to the equilibrium construction above, only type $\overline{s}$ can reveal his type before the full separating stage (i.e., q_t^i cannot jump to 1 when $p_t > p_c(\underline{s} + \overline{s})$), so I discuss only the equilibrium after type $\overline{s}$ reveals his type.

If q_0 is large enough, there can be an equilibrium of only two stages, a *pooling stage* and a *full separating stage*, with the cutoff point $p_c(2\overline{s})$. In this equilibrium, players achieves the efficient level of experimentation when the realization of s is $2\overline{s}$.

The detailed construction of this subgame equilibrium is in Appendix A.2. The construction shows that the full separating stage happens after p_t reaches $p_c(2\overline{s})$. However, in the equilibrium of the two-sided private information problem, revealing one's type (type $\overline{s}$ moves to the no-learning arm) happens only when $p_t \leq p_s$, where $p_s \in (p_c(\underline{s} + \overline{s}), p_c(2\overline{s}))$. As a result, in the outcome of the two-sided private information problem, when type $\overline{s}$ reveals himself, the full separation has happened in the continuing one-sided private information problem, so the other player also reveals his type.

Although there can be a multiplicity problem in this one-sided private information problem, I will assume that the players play the equilibrium above, where type $\overline{s}$ keeps his private information unrevealed as long as possible.

4. Welfare

4.1. Ex ante welfare

I first discuss the ex ante welfare of the equilibrium. The ex ante welfare is the expected total payoff of the two players before knowing their own types.

Before knowing his own type and that of the other player, one player's expected payoff is

$$EW(p_0, q_0) = q_0 W(p_0, q_0; \underline{s}) + (1 - q_0) W(p_0, q_0; \overline{s}),$$

where $W(p, q; s)$ is type s's initial expected payoff in my equilibrium, with $p_0 = p$, $q_0 = q$. The two players have the same ex ante expected payoff, so the ex ante welfare is

$$2EW(p_0, q_0) = 2q_0 W(p_0, q_0; \underline{s}) + 2(1 - q_0) W(p_0, q_0; \overline{s}).$$

I regard this ex ante welfare as the total welfare of my model and compare it to the ex ante welfare of the full-information benchmark, the KRC model. In the benchmark, denote the player's initial expected payoff as $W_b(p_0; s)$ when the initial belief about the state of the learning arm is p_0 and the flow payoff of the no-learning arm is s. Then compare the ex ante welfare of my model to

$$EW_b(p_0, q_0) = 2\Big[q_0^2 W_b(p_0, 2\underline{s}) + 2q_0(1 - q_0) W_b(p_0, \underline{s} + \overline{s}) + (1 - q_0)^2 W_b(p_0, 2\overline{s})\Big].$$

Proposition 2. *Two-sided private information improves the ex ante welfare compared to the full-information benchmark if the equilibrium in Proposition 1 exists.*

Figure 5 illustrates the ex ante welfare difference between the full-information benchmark and the two-sided private information model ($EW_b(p_0, q_0) - 2EW(p_0, q_0)$) under different initial beliefs (different p_0). When $p_s < p_0 < 1$, the two-sided private information always increases the ex ante payoff. Obviously, at $p_0 = 1$, players play $k = 1$ in both models, so they have the same ex ante welfare.

Though two-sided private information leads to the possibility of over-experimentation, the ex ante welfare is increased compared to the full-information benchmark. The intuition of this increase is similar to the reason for over-experimentation. There is a possible welfare loss compared to the benchmark when over-experimentation happens. However, in other realizations of s where over-experimentation does not happen, there is a payoff gain: more experimentation can mitigate the

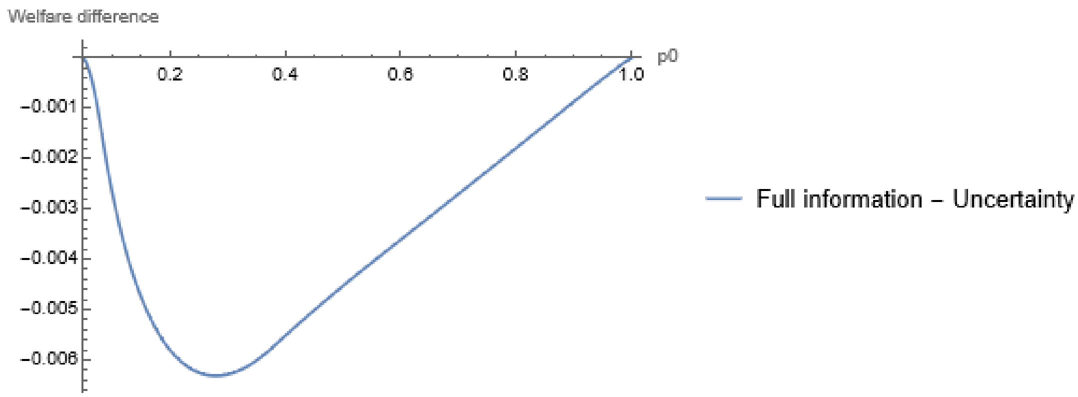


Figure 5: Welfare difference between the full-information benchmark and the two-sided private information model under different initial beliefs, $p_0 > p_s$

under-experimentation problem caused by free riding. When the researcher is considering the ex ante welfare, the welfare gain can cover the welfare loss.

As shown in Figure 5, there is a hump in $EW_b(p_0, q_0) - 2EW(p_0, q_0)$. The difference between my model and the benchmark comes from how players experiment under different private information conditions on the no-learning arm. However, the difference caused by the two-sided private information will be overwhelmed by the similarity if p_0 is very high or very low. Under my assumption, the players prefer the learning arm if its state is $\theta = 1$, and the no-learning arm if the learning arm's state is $\theta = 0$. When p_0 is high, the learning arm is very attractive, and the players will use it in both models. So, the difference caused by the different private information in the no-learning arm is overwhelmed by the similarity. The same happens for very low p_0. Therefore, the most significant difference for p_0 is in between the two cases.

4.2. Interim payoffs

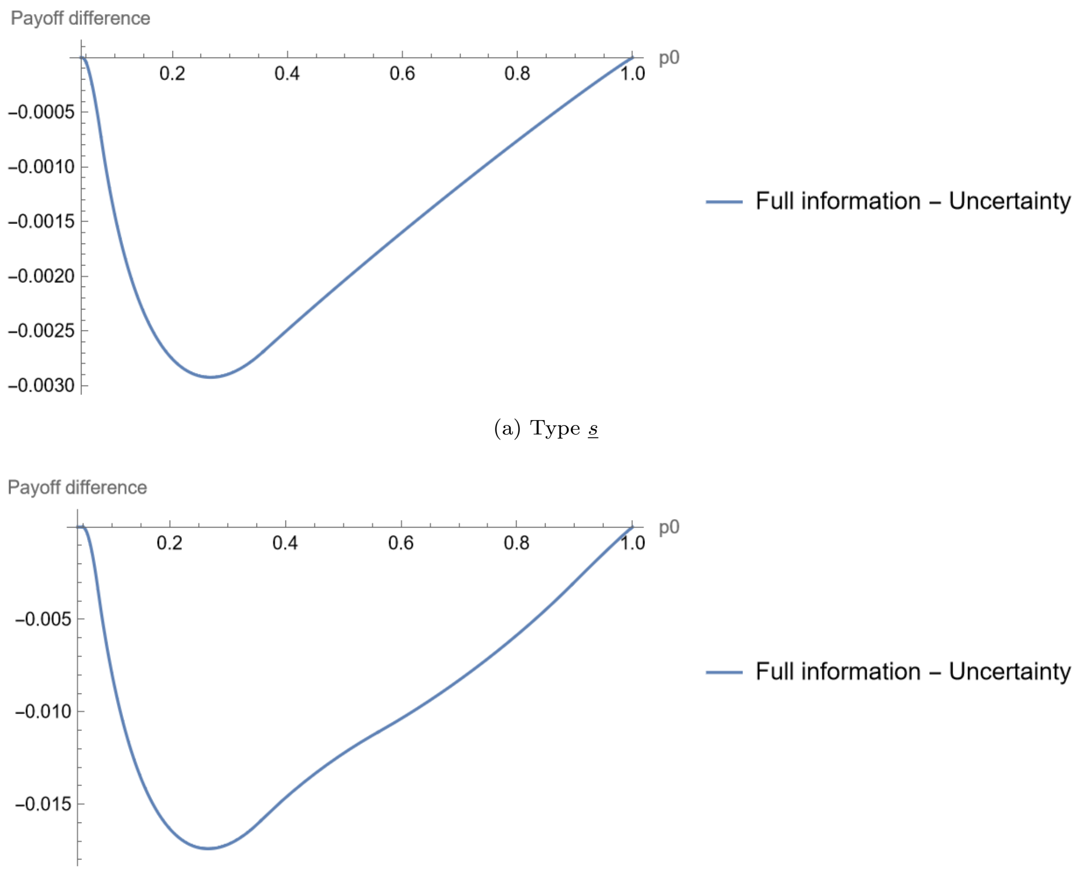
In this part, I discuss the interim payoffs, specifically, the expected payoff of players at the beginning of the game but after knowing their own types. Clearly, $W(p_0, q_0; s)$ is the interim payoff of type s. Again, $W(p_0, q_0; s)$ will be compared to the full-information benchmark, $q_0 W_b(p_0, s + \underline{s}) + (1 - q_0) W_b(p_0, s + \overline{s})$.

As shown in Figure 6, for most p_0 values, my model has a higher interim payoff than the full-information benchmark for both types. For type $\underline{s}$, having two-sided private information mitigates the under-experimentation problem without incurring over-experimentation. For type $\overline{s}$, having two-sided private information leads to more experimentation, and over-experimentation is possible. However, because it is an equilibrium and type $\overline{s}$ does not reveal his type in equilibrium, he prefers the current equilibrium payoff to the full-information benchmark. Therefore, players do not want to have full information even after knowing their own types.

Consider the situation in which one player has a chance of verifiably communicating his private information to the other player, and the problem becomes one-sided. Type $\overline{s}$ does not want to communicate his private information. Even without this communicating chance, type $\overline{s}$ can reveal his type by choosing $k_t < 1$. Since it is an equilibrium strategy that type $\overline{s}$ mimics type $\underline{s}$, type $\overline{s}$ does not want to use the communicating chance. It is not possible to get the same result for type $\underline{s}$ in the same way, because type $\underline{s}$ does not have the chance to reveal his type in the equilibrium.

Compare the equilibrium result of the two-sided private information problem and the one-sided private information problem for type $\underline{s}$. If both players are type $\underline{s}$, then in both problems, the equilibrium result is that the players play the KRC strategy of $s = 2\underline{s}$.

If the other player is type $\overline{s}$, the two problems have different experimentation levels. When q_0 is high at the beginning of the game, then the one-sided private information problem after type $\underline{s}$ communicates his private type can support an equilibrium with only two stages – pooling and full separating – and the cutoff point is $p_c(\underline{s} + \overline{s})$. So, in the one-sided problem, a player of type $\underline{s}$ achieves efficient experimentation when the other player is type $\overline{s}$. However, in the two-sided private information problem, the players may stop experimentation before $p_c(\underline{s} + \overline{s})$. Therefore, the one-sided problem gives a better equilibrium payoff, and type $\underline{s}$ wants to communicate if he has the chance.



(a) Type $\underline{s}$

(b) Type $\overline{s}$

Figure 6: Interim payoff difference between the model and the full-information benchmark, $p_0 > p_s$

5. Discussion

5.1. Multiplicity problem

As a signaling game, we will face the multiplicity problem. Similar problems also happen in Dong (2021). There are multiple equilibria for the game with two-sided private information and for the continuation game with one-sided private information (after one player reveals his private information). Even after both sides reveal their private information and the game degenerates to the full-information benchmark, there are multiple MPEs, but I focus on the symmetric MPE here.

To deal with the multiplicity problem, the first restriction (**restriction 1**) I use is the symmetric requirement. As discussed in Keller, Rady and Cripps (2005), requiring two players to play the same strategy will resolve the multiplicity problem of the continuation game after both sides reveal their private information. This restriction does not apply to the continuation game where only one player reveals his private signal, because the two players are asymmetric here. For the original game with two-sided private information, the symmetric requirement is that the two players should use the same strategy if they are of the same type.

I place the second restriction (**restriction 2**) on the one-sided private information problem after one player reveals his private type. The one-sided private information problem also suffers from the multiplicity problem. Following a similar idea as in Dong (2021), I assume that when the players are facing the one-sided private information problem, they play the equilibrium where the private information is held for the longest time (the equilibrium introduced in section 3.4).

The third restriction (**restriction 3**) I use is on the off-path beliefs. If the strategy of type $\underline{s}$ is $k(\underline{s})$, then any action $k^i < k(\underline{s})$ will lead to $q_t^i = 0$ and any action $k^i > k(\underline{s})$ will lead to $q_t^i = 1$ if q_t^i is not 0 before the jump. That is, any deviation from the on-path strategy has a deterministic effect on q_t^i. But for $q_t^i = 0$, deviating to $k > k(\underline{s})$ does not change q_t^i anymore. This restriction has a similar intuition as the D1 criterion but is stronger. For type $\underline{s}$, the opportunity cost of experimenting is small, so he tends to experiment more. On the contrary, type $\overline{s}$ tends to experiment less because of a higher opportunity cost. So, a larger action is more likely to be used by type $\underline{s}$, and I assume this consideration has a deterministic effect on the beliefs about the private type. This restriction also means that if there is full separation, the strategy of type $\underline{s}$ must be higher than that of type $\overline{s}$. Besides, this restriction also says there are only two ways to change q_t^j. The first way is to let it jump to 1 or 0, which is achieved by deviating for sure. The second way is to let it gradually change by mixing the on-path strategy and the deviation.[13] Then we have the following proposition.

Proposition 3. *Under the three restrictions above and the same parameter conditions as in Proposition 1, the on-path belief path* (q_t^1, q_t^2) *in a MPE is unique.*

The following is a heuristic proof idea of the proposition; the complete proof is relegated to Appendix A.4.

Firstly, if $p_t \leq p_c(\underline{s} + \overline{s})$, type $\overline{s}$ does not want to experiment, because his belief about the learning arm p_t is already below the efficient cutoffs, no matter if the other player is type $\underline{s}$ or $\overline{s}$. But type $\underline{s}$ can experiment with his full resource if the other player is also type $\underline{s}$, so there will be separating when $p_t \leq p_c(\underline{s}+\overline{s})$.

Secondly, if type $\overline{s}$ is indifferent between mimicking type $\underline{s}$ and revealing his type, the on-path q_t must satisfy $q_i(p_t)$ when $p_c(\underline{s}+\overline{s}) < p_t \leq p_s$. Furthermore, type $\underline{s}$ cannot be indifferent between mimicking type $\overline{s}$ and revealing his type at $p_t > p_c(\underline{s} + \overline{s})$ in equilibrium.

Thirdly, $q_t = 1$ or 0 cannot be the on-path belief when $p_c(\underline{s} + \overline{s}) < p_t \leq p_s$. If there is full separation, then type $\overline{s}$ will mimic type $\underline{s}$ if the other player reveals himself as type $\underline{s}$. By doing so, type $\overline{s}$ can get more experimentation when $p_t > p_c(\underline{s} + \overline{s})$.

[13]Of course, gradual change happens when no deviation happens, and q_t^j jumps to 0 or 1 when deviation happens.

Fourthly, pooling cannot happen when $p_t \leq p_s$. If the two types pool on $k < 1$, type $\underline{s}$ will deviate up to reveal his type and then get a better payoff, as discussed in section 4.2. If the two types pool on $k = 1$ for $p_t < p_s$, the beliefs q_t^i and q_t^j have to stay above q_0 for any $p_t > p_s$, which is a contradiction.

Finally, for $p_t > p_s$, q_t^i and q_t^j should remain unchanged at q_0 because there is no partial revealing from type $\underline{s}$.

Consequently, the beliefs q_t^i and q_t^j in any MPE should be the same as in the constructed equilibrium.

5.2. More players

A natural question is what would happen if there were more than two players. For example, in the setup with three players, I assume that the no-learning arm generates flow payoff $s = s_1 + s_2 + s_3$, where $s_1, s_2, s_3 \in \{\underline{s}, \overline{s}\}$. The distributions of the three variables are $q_0 \circ \underline{s} + (1 - q_0) \circ \overline{s}$. All other setups are the same as the original model.

Having more players leads to more free riding, encouragement, and private information.[14] As discussed, the free-riding problem makes players experiment less; players may want to experiment more to hide their private information from others to encourage them to experiment; the ignorance towards the true state can make players experiment beyond the efficient boundary.

More players do not change the result qualitatively. Take three players as an example; when $\overline{s}$ is large enough and $\underline{s}$ is small enough, in equilibrium, we still have three stages if q_0 and p_0 are large enough: the pooling stage, the partial separating stage, and the full separating stage. Two cutoffs govern this equilibrium: $p_c(\overline{s} + 2\underline{s})$ and p_s^3. For $p_t > p_s^3$, both types play $k = 1$; for $p_c(\overline{s} + 2\underline{s}) < p_t \leq p_s^3$, type $\underline{s}$ plays $k = 1$, and type $\overline{s}$ randomly stops from $k = 1$ to $k = 0$; for $p_t \leq p_c(\overline{s} + 2\underline{s})$, type $\overline{s}$ plays $k = 0$, and type $\underline{s}$ plays the KRC strategy of the true s. And when $p_t \in (p_c(\overline{s} + 2\underline{s}), p_s^3]$, the belief of the other

[14] This is because each player will have his own private information.

player being type $\underline{s}$ seeing no stopping happens is

$$q_i^3(p_t) = \frac{3\overline{s}r - p_t[\lambda(r+3\lambda) - 9\lambda\overline{s}]}{(\overline{s}-\underline{s})(2r+6\lambda p_t)},$$

while the belief jumps to 0 when stopping happens. And p_s^3 solves $q_i^3(p_t) = q_0$.

The full separating happens earlier than in the original model ($p_c(\overline{s}+2\underline{s}) > p_c(\overline{s}+\underline{s})$) because the flow payoff of the no-learning arm – the opportunity cost of experimentation – is larger than in the two-player case. If I make the highest and lowest realization of s the same in the three-player and two-player cases, i.e., $s_1, s_2, s_3 \in \{\frac{2}{3}\underline{s}, \frac{2}{3}\overline{s}\}$, in the three-player case the full separation happens later than in the original model ($p_c(\frac{2}{3}\overline{s} + \frac{4}{3}\underline{s}) < p_c(\overline{s}+\underline{s})$). So, after the effect of the changing flow payoff of the no-learning arm is removed, the overall effect of increasing the number of players to three encourages experimentation. Thus, the pro-experimentation effects of encouragement and ignorance overwhelm the deteriorating free-riding problem.

Furthermore, with three players, both the realizations $s = 3\overline{s}$ and $s = \underline{s} + 2\overline{s}$ can have over-experimentation. Though there are more possibilities of over-experimentation with three players, welfare results are still similar to the two-player case: the ex ante welfare is increased because of the private information. The benefit of mitigating the free-riding problem covers the loss from possible over-experimentation.

6. Conclusion

This paper departs from the two-arm bandit problem of Keller, Rady and Cripps (2005) by changing the previous safe arm to the no-learning arm. The no-learning arm generates an unobserved constant flow payoff about which both players have private information. The two-sided private information in the no-learning arm increases the experimentation level compared to the benchmark without private information. A player experiments more because he (1) has the incentive to hide his private information and encourage the other player to experiment more and (2) does not know the true flow payoff of the no-learning arm. Only the second force can lead to over-experimentation. For example, players should stop experimentation early when the true flow payoff is high. However, in my model, both players may think the flow payoff is possibly low and thus continue to experiment beyond the efficient threshold. The two-sided private information can increase the experimentation level and even overturn the under-experimentation problem caused by information free riding in Keller, Rady and Cripps (2005).

Though the possibility of over-experimentation exists, if the ex ante welfare of two players is considered, two-sided private information is still beneficial. More experimentation in my model has the benefit of mitigating the free-riding problem and the harm from possible over-experimentation. The former benefit can counterbalance the latter harm and increase the ex ante welfare of players.

The current model evaluates what happens when the backup of a risky investment is not completely safe, by assuming a no-learning arm. Based on a similar idea, possible extensions may lay in other unsafe backup options. For example, people may hire an expert to manage their money as a backup for investing in a new and risky project, which leads to a delegating problem. Modeling these backups may capture richer situations.

References

Banks, Jeffrey S. 1993. "Two-sided uncertainty in the monopoly agenda setter model." *Journal of Public Economics*, 50(3): 429–444.

Bolton, Patrick, and Christopher Harris. 1999. "Strategic experimentation." *Econometrica*, 67(2): 349–374.

Bonatti, Alessandro, and Johannes Hörner. 2011. "Collaborating." *American Economic Review*, 101(2): 632–63.

Cho, In-Koo. 1990. "Uncertainty and delay in bargaining." *The Review of Economic Studies*, 57(4): 575–595.

Cramton, Peter C. 1992. "Strategic delay in bargaining with two-sided uncertainty." *The Review of Economic Studies*, 59(1): 205–225.

Cronshaw, Mark B, and James Alm. 1995. "Tax compliance with two-sided uncertainty." *Public Finance Quarterly*, 23(2): 139–166.

Dong, Miaomiao. 2021. "Strategic experimentation with asymmetric information." *Unpublished Paper, Pennsylvania State University.[1027].*

Heidhues, Paul, Sven Rady, and Philipp Strack. 2015. "Strategic experimentation with private payoffs." *Journal of Economic Theory*, 159: 531–551.

Kahn, Charles, and Gur Huberman. 1988. "Two-sided uncertainty and" up-or-out" contracts." *Journal of Labor Economics*, 6(4): 423–444.

Keller, Godfrey, and Sven Rady. 2010. "Strategic experimentation with Poisson bandits." *Theoretical Economics*, 5(2): 275–311.

Keller, Godfrey, and Sven Rady. 2015. "Breakdowns." *Theoretical Economics*, 10(1): 175–202.

Keller, Godfrey, Sven Rady, and Martin Cripps. 2005. "Strategic experimentation with exponential bandits." *Econometrica*, 73(1): 39–68.

Appendix A. Appendix

Appendix A.1. Equilibrium results

In this part I verify that there is no profitable deviation for any player in all three stages, which is the proof for Proposition 1.

Appendix A.1.1. The full separating stage ($p_t \leq p_c(\underline{s} + \overline{s})$)

Since the belief about the learning arm is already below the efficient cutoff point of $s = \underline{s} + \overline{s}$, type $\overline{s}$ will play $k = 0$, no matter which type the other player is. So type $\overline{s}$ does not want to deviate.

For type $\underline{s}$, since the game starts with $p_0 > p_s$, when p_t enters the full separating stage, the true types of players will be revealed. There is no uncertainty in the no-learning arm now, so type $\underline{s}$ does not want to deviate from the KRC strategy of the true s.

If $q_t^i \in (0,1)$, as shown in the strategy, type $\underline{s}$ will play $k=1$ for an instance. After that moment, the types are revealed and then players follow the KRC strategy of the true s. Clearly, he will not deviate after the type is revealed. And he will also not deviate from a moment of $k=1$, since that will make the other player think that he is type $\overline{s}$ and play $k=0$ after that moment. This will be no better than the on-path payoff: if the other player is type $\underline{s}$, the deviation hurts him since it reduces experimentation; if the other player is type $\overline{s}$, the deviation gives the same payoff as on-path strategies (in both cases, they play $k=0$ afterwards).

Appendix A.1.2. The partial separating stage ($p_c(\underline{s}+\overline{s}) < p_t \leq p_s$)

As discussed in section 3.3, type $\overline{s}$ does not want to deviate since he is indifferent between mimicking type $\underline{s}$ and revealing himself.

For type $\underline{s}$, if he deviates from $k=1$, q_t^i will jump to 0. Since we have assumed that q_0 is large enough so that the KRC strategy of $s=\underline{s}+\overline{s}$ is $k=0$ at p_s, the other player plays $k=0$ then. But stay on-path will make the other player experiments more – type $\underline{s}$ plays $k=1$, and type $\overline{s}$ mixes between $k=1$ and $k=0$. This is better for type $\underline{s}$ since $p_t > p_c(\underline{s}+\overline{s})$.

Appendix A.1.3. The pooling stage ($p_t > p_s$)

Firstly, for type $\underline{s}$, deviating from $k=1$ makes q_t^i jumps to 0 and makes the other player experiment less, so he does not want to deviate.

Then we consider type $\overline{s}$. Firstly let us suppose that the cutoff point where the KRC strategy of $s=\underline{s}+\overline{s}$ changes to 0 is $\hat{p}_1$, the cutoff point where the KRC strategy of $s=\underline{s}+\overline{s}$ reach 1 is $\hat{p}_2$.

(1) $p_s < p_t \leq \hat{p}_1$

Since $p_s < \hat{p}_1$, the payoff after revealing type $\overline{s}$ is $q_0(\underline{s}+\overline{s})+(1-q_0)(2\overline{s})$ at p_s, and the right derivative of the payoff after revealing at p_s is 0.

Since before p_s, two players play $k=1$, so the on-path payoff $u(p)$ of type $\overline{s}$

should satisfy the following ODE.

$$2\lambda p(1-p)u'(p) + (r+2\lambda p)u(p) = (r+2\lambda)\lambda hp \tag{A.1}$$

Recall our equation (2), type $\overline{s}$'s payoff at p_s will be

$$u(p_s) = q_0(\underline{s}+\overline{s}) + (1-q_0)2\overline{s} \tag{A.2}$$

Plug (A.2) into (A.1) with $p = p_s$, and use equation (5) at $p_t = p_s$, we get

$$2\lambda p_s(1-p_s)u'(p_s) = 0$$

where $u'(p_s)$ is the right derivative, since (A.1) satisfied only for $p > p_s$.

So, the right derivative of type $\overline{s}$'s on-path payoff is also 0. Then we know that type $\overline{s}$ does not want to stop when $p_s < p_t \le \hat{p}_1$: the on-path payoff $u(p_t)$ is increasing when $p_t \in (p_s, \hat{p}_1]$; the deviation payoff is constant at $q_0(\underline{s}+\overline{s}) + (1-q_0)2\overline{s}$ when $p_t \in (p_s, \hat{p}_1]$; the on-path payoff $u(p_t)$ is equal to the deviation payoff at $p_t = p_s$.

(2) $\hat{p}_1 < p_t \le p_c(2\overline{s})]$[15].

When both players choose $k=1$ for $p_t > p_s$ and $k=0$ for $p_t \le p_s$, let the payoff type $\overline{s}$ gets be $v_l(p_t)$ if the other one is type $\underline{s}$, be $v_h(p_t)$ if the other one is type $\overline{s}$. Then type $\overline{s}$'s payoff in the pooling stage will be $v(p_t) = q_0 v_l(p_t) + (1-q_0)v_h(p_t)$[16].

Type $\overline{s}$'s payoff of revealing himself is $q_0 w_l(p_t) + (1-q_0)w_h(p_t)$, where $w_l(p_t)$ is the payoff in the KRC model with $s = \overline{s}+\underline{s}$, $w_h(p_t)$ is the efficient payoff with $s = 2\overline{s}$, i.e. the payoff when both players play $k=0$ for $p_t \le p_c(2\overline{s})$ and play $k=1$ for $p_t > p_c(2\overline{s})$[17].

To make revealing himself not profitable for type $\overline{s}$, we need

$$q_0 v_l(p_t) + (1-q_t)v_h(p_t) \ge q_0 w_l(p_t) + (1-q_0)w_h(p_t) \Leftrightarrow \frac{q_0}{1-q_0} \ge \frac{w_h(p_t)-v_h(p_t)}{v_l(p_t)-w_l(p_t)}$$

---

[15] Recall that according to our assumption on $\overline{s}$ and $\underline{s}$, we have $\hat{p}_1 < \hat{p}_2 < p_c(2\overline{s})$.

[16] The on-path payoff at $p_t > p_s$ in my equilibrium satisfies the ODE $2\lambda p(1-p)v'(p) + (r+2\lambda p)v(p) = (r+2\lambda)\lambda hp$, and has payoff of $q_0(\underline{s}+\overline{s}) + (1-q_0)2\overline{s}$ at the point p_s, so the on-path payoff will be the same as the payoff if two players choose $k=1$ for $p_t > p_s$ and $k=0$ for $p_t \le p_s$.

[17] According to the constructed equilibrium in the one-sided private information problem.

notice that $v_l(p_t) - w_l(p_t) > 0$ since $p_c(\underline{s}+\overline{s}) < p_s < \hat{p}_1$.

Clearly, $w_h(p_t) - v_h(p_t)$ is bounded above, since $w_h(p_t)$ and $v_h(p_t)$ are surely between 0 and λh.

Now we only consider q_0 large enough such that $p_s \leq \hat{p}_1 - \epsilon$, where ϵ is a small positive number. Notice that $v_l(p_t)$ changes with p_s.

Lemma 1. *For any* $p_s \in [p_c(\underline{s}+\overline{s}), \hat{p}_1 - \epsilon]$ *and* $p_t \in (\hat{p}_1, p_c(2\overline{s})]$, $v_l(p_t) - w_l(p_t)$ *is bounded away from* 0, *i.e.* $\exists\, \tau > 0$ *s.t.* $v_l(p_t) - w_l(p_t) > \tau$, $\forall p_s \in [p_c(\underline{s}+\overline{s}), \hat{p}_1 - \epsilon]$, $p_t \in (\hat{p}_1, p_c(2\overline{s})]$.

Proof. Firstly, Let consider $p_t \in (\hat{p}_1, \hat{p}_2]$.

As discussed in Keller, Rady and Cripps (2005), $w_l(p)$ satisfies the ODE

$$\lambda p(1-p)w_l'(p) + \lambda p w_l(p) = (r+\lambda)\lambda hp - r(\underline{s}+\overline{s})$$

and thus

$$w_l(p) = \underline{s} + \overline{s} + (\frac{r}{\lambda} + 1)(\lambda h - (\underline{s}+\overline{s})) + \frac{r(\underline{s}+\overline{s})(1-p)}{\lambda} \ln \frac{1-p}{p} + C_1(1-p)$$

where C_1 is a constant depending on the cutoff point $\hat{p}_1$ and the payoff at that point $\underline{s}+\overline{s}$.

Take derivative twice we get

$$w_l''(p) = \frac{r(\underline{s}+\overline{s})}{\lambda} \frac{1}{(1-p)p^2} > 0$$

Also we have $v_l(p)$ satisfies the ODE

$$2\lambda p(1-p)v_l'(p) + (r + 2\lambda p)v_l(p) = (r+2\lambda)\lambda hp$$

with the initial condition $v_l(p_s) = \underline{s} + \overline{s}$.

Consequently,

$$v_l(p) = \lambda hp + C_2(1-p)(\frac{1-p}{p})^{\frac{r}{2\lambda}}$$

where $C_2 = (\underline{s} + \overline{s} - \lambda h p_s)\frac{1}{1-p_s}(\frac{p_s}{1-p_s})^{\frac{r}{2\lambda}}$.

Then we have

$$\frac{v_l''(p)}{w_l''(p)} = \frac{C_2}{2(\underline{s}+\overline{s})}(1 + \frac{r}{2\lambda})(\frac{1-p}{p})^{\frac{r}{2\lambda}} > 0$$

which is decreasing in p.

So, $v_l''(p) - w_l''(p)$ can cross 0 at most once (from above to below).

Notice that $v_l'(\hat{p}_1) > 0$ and $w_l'(\hat{p}_1) = 0$, we have $v_l'(\hat{p}_1) - w_l'(\hat{p}_1) > 0$. Then $v_l'(p) - w_l'(p)$ can cross 0 at most once (from above to below) on $(\hat{p}_1, \hat{p}_2]$. So the minimum of $v_l(p) - w_l(p)$ on $(\hat{p}_1, \hat{p}_2]$ is reached at either $\hat{p}_1$ or $\hat{p}_2$.

Notice that as p_s increase, $w_l(p)$ stays unchanged and $v_l(p)$ decreases (since $p_s > p_c(\underline{s} + \overline{s})$). Then we only need to check $v_l(p) - w_l(p)$ at $p_s = \hat{p}_1 - \epsilon$ to explore the minimum. It is easy to see that when $p_s = \hat{p_1} - \epsilon$, $v_l(p) - w_l(p)$ is positive at both $p = \hat{p}_1$ or $\hat{p}_2$ (it will be a positive number depending on ϵ).

Then I consider $p_t \in (\hat{p}_2, p_c(2\overline{s})]$.

In this interval, both $v_l(p)$ and $w_l(p)$ satisfy the ODE

$$2\lambda p(1-p)y'(p) + (r + 2\lambda p)y(p) = (r + 2\lambda)\lambda hp$$

and thus

$$w_l(p) = \lambda hp + (w_l(\hat{p}_2) - \lambda h\hat{p}_2)\frac{1}{1-\hat{p}_2}(\frac{\hat{p}_2}{1-\hat{p}_2})^{\frac{r}{2\lambda}}(1-p)(\frac{1-p}{p})^{\frac{r}{2\lambda}}$$

$$v_l(p) = \lambda hp + (v_l(\hat{p}_2) - \lambda h\hat{p}_2)\frac{1}{1-\hat{p}_2}(\frac{\hat{p}_2}{1-\hat{p}_2})^{\frac{r}{2\lambda}}(1-p)(\frac{1-p}{p})^{\frac{r}{2\lambda}}$$

At $p_t = \hat{p}_2$, we have $v_l(\hat{p}_2) > w_l(\hat{p}_2)$ already, so

$$v_l'(p) - w_l'(p) < 0, p \in (\hat{p}_2, p_c(2\overline{s})]$$

Consequently, the minimum among $p_t \in (\hat{p}_2, p_c(2\overline{s})]$ is reached at $p_c(2\overline{s})$. Again, since $v_l(p)$ decreases in p_s, the minimum in $p_t \in (\hat{p}_2, p_c(2\overline{s})]$ and $p_s \in [p_c(\underline{s} + \overline{s}), \hat{p}_1 - \epsilon]$ is reached at $p_s = \hat{p}_2 - \epsilon$ and $p_t = p_c(2\overline{s})$.

Notice that $v_l(p) - w_l(p)$ is decreasing in p and $v_l(1) - w_l(1) = 0$, so the minimum which is reached at $p_s = \hat{p}_2 - \epsilon$ and $p_t = p_c(2\overline{s})$ is positive, which depends on ϵ.

Combining two positive lower bounds for $p_t \in (\hat{p}_1, \hat{p}_2]$ and $p_t \in (\hat{p}_2, p_c(2\overline{s})]$, we can conclude that there will be a positive τ such that $v_l(p) - w_l(p) \geq \tau > 0$ for any $p_s \in [p_c(\underline{s} + \overline{s}), \hat{p}_1 - \epsilon]$, $p_t \in (\hat{p}_1, p_c(2\overline{s})]$. ■

With the lemma above, we can see that $\frac{w_h(p_t)-v_h(p_t)}{v_l(p_t)-w_l(p_t)}$ is bounded above for any $p_s \in [p_c(\underline{s}+\overline{s}), \hat{p}_1-\epsilon]$, $p_t \in (\hat{p}_1, p_c(2\overline{s})]$[18]. So we know that for q_0 larger than a threshold, $\frac{q_0}{1-q_0} \geq \frac{w_h(p_t)-v_h(p_t)}{v_l(p_t)-w_l(p_t)}$ with any $p_s \in [p_c(\underline{s}+\overline{s}), \hat{p}_1-\epsilon]$, $p_t \in (\hat{p}_1, p_c(2\overline{s})]$.

Notice that to ensure $p_s \leq \hat{p}_1 - \epsilon$, we need q_0 to be larger than another threshold. Consequently, if q_0 is larger than both two thresholds discussed above, we have that $p_s \in [p_c(\underline{s}+\overline{s}), \hat{p}_1-\epsilon]$ and type $\overline{s}$ does not want to deviate to reveal himself when $p_t \in (\hat{p}_1, p_c(2\overline{s})]$.

(3) $p_t > p_c(2\overline{s})$

Similar to above, both the on-path payoff $v(p)$ and the revealing payoff $w(p)$ satisfy the ODE

$$2\lambda p(1-p)y'(p) + (r+2\lambda p)y(p) = (r+2\lambda)\lambda hp$$

since two players use $k=1$ for $p_t > p_c(2\overline{s})$ both before and after revealing.

Then we have

$$w(p) = \lambda hp + (w(p_c(2\overline{s})) - \lambda hp_c(2\overline{s}))\frac{1}{1-p_c(2\overline{s})}(\frac{p_c(2\overline{s})}{1-p_c(2\overline{s})})^{\frac{r}{2\lambda}}(1-p)(\frac{1-p}{p})^{\frac{r}{2\lambda}}$$

$$v(p) = \lambda hp + (v(p_c(2\overline{s})) - \lambda hp_c(2\overline{s}))\frac{1}{1-p_c(2\overline{s})}(\frac{p_c(2\overline{s})}{1-p_c(2\overline{s})})^{\frac{r}{2\lambda}}(1-p)(\frac{1-p}{p})^{\frac{r}{2\lambda}}$$

As discussed in part (2), with q_0 large enough, we have $v(p) \geq w(p)$ with $p \in (\hat{p}_2, p_c(2\overline{s})]$, so $v(p_c(2\overline{s})) \geq w(p_c(2\overline{s}))$. Then we can conclude that $v(p) \geq w(p)$ for $p > p_c(2\overline{s})$, and thus deviating to revealing himself is not profitable for type $\overline{s}$.

Appendix A.2. One-sided private information problem

Appendix A.2.1. Construction of the equilibrium

In this part I provide a detailed construction of the equilibrium for the one-sided private information problem after type $\overline{s}$ revealing himself, i.e. the one who has revealed himself is of type $\overline{s}$.

[18] Notice that $v_l(p_t) - w_l(p_t)$ is also bounded above

When $p_t \leq p_c(2\overline{s})$, the player who has not revealed himself (who knows the true realization of s) plays the KRC strategy of $s = \underline{s} + \overline{s}$ if he is type $\underline{s}$, and plays $k = 0$ if he is type $\overline{s}$. For the player who has revealed himself (who is still unsure about the other player's type), he plays the KRC strategy of $s = \underline{s} + \overline{s}$ if $q_t^i = 1$, and plays $k = 0$[19] if $q_t^i = 0$.

When $p_t > p_c(2\overline{s})$, the player who has not revealed himself plays $k = 1$ for both types. The player who has revealed himself plays $k = 1$ as well.

Clearly, $p_t > p_c(2\overline{s})$ is the pooling stage and $p_t \leq p_c(2\overline{s})$ is the full separating stage. The on-path q_t[20] is q_0 with $p_t > p_c(2\overline{s})$, and 1 or 0 with $p_t \leq p_c(2\overline{s})$.

As for the off-path belief, any k smaller than the strategy of type $\underline{s}$ makes q_t^i jump to 0.

Appendix A.2.2. Proof of the equilibrium

(1) $p_t \leq p_c(2\overline{s})$

For the player who has not revealed himself, if he is type $\overline{s}$, he does not want to deviate to $k = 1$, because p_t is already below the efficient cutoff point for $s = 2\overline{s}$, and thus making both players experiment more is not beneficial for him. If he is type $\underline{s}$, he does not want to deviate down, since that reduces both players' experimentation[21]. And he does not want to deviate up either. Since he already has the full reputation of type $\underline{s}$ ($q_t = 1$), deviating up does not affect q_t. Then since KRC strategy is an equilibrium in the full information benchmark, type $\underline{s}$ does not deviate up due to no effect on q_t.

For the player who has revealed himself, his action change has no effect on q_t^i. Since now he knows the true realization of s and the KRC strategies is an equilibrium in the full information benchmark, this player has no incentive to deviate.

(2) $p_t > p_c(2\overline{s})$

[19] This is the KRC strategy of $s = 2\overline{s}$ at $p_t \leq p_c(2\overline{s})$

[20] I remove the superscript since there is only one belief about the other player's type in the one-sided private information problem.

[21] Obviously, when on-path strategy is $k > 0$ for type $\underline{s}$, $p_t > \hat{p}_1$.

For the player who has not revealed himself, no matter what his type is, if he deviates to $k < 1$, q_t jumps to 0 and the other player experiment less. It will not be profitable since $p_t > p_c(2\overline{s}) > p_c(\underline{s}+\overline{s})$, which means p_t is still above the efficient cutoff point, no matter $s = 2\overline{s}$ or $\underline{s} + \overline{s}$.

For the player who has revealed himself, his action has no effect on q_i. His on-path payoff $u(p)$ should satisfy the following Hamilton–Jacobi–Bellman equation

$$\begin{aligned} u(p) - Es = \{\lambda hp - Es + \frac{1}{r}[\lambda p(\lambda h - u(p)) - \lambda p(1-p)u'(p)]\} \\ + \frac{1}{r}[\lambda p(\lambda h - u(p)) - \lambda p(1-p)u'(p)] \end{aligned}$$

where $Es = q_0(\underline{s} + \overline{s}) + (1 - q_0)2\overline{s}$.

Using the results of Keller, Rady and Cripps (2005), to make $k = 1$ the optimal strategy of the player, we need

$$S(p) = u(p) - 2Es + \lambda hp \geq 0 \tag{A.3}$$

Actually, $u(p)$ can be written as $u(p) = q_0 u_l(p) + (1 - q_0)u_h(p)$, where $u_l(p)$ and $u_h(p)$ are the payoff if the other player is type $\underline{s}$ and $\overline{s}$ respectively, conditional on on-path strategies. Then we have two Hamilton–Jacobi–Bellman equations

$$\begin{aligned} u_l(p) - (\underline{s} + \overline{s}) = \{\lambda hp - (\underline{s} + \overline{s}) + \frac{1}{r}[\lambda p(\lambda h - u_l(p)) - \lambda p(1-p)u_l'(p)]\} \\ + \frac{1}{r}[\lambda p(\lambda h - u_l(p)) - \lambda p(1-p)u_l'(p)] \end{aligned}$$

$$\begin{aligned} u_h(p) - 2\overline{s} = \{\lambda hp - 2\overline{s} + \frac{1}{r}[\lambda p(\lambda h - u_h(p)) - \lambda p(1-p)u_h'(p)]\} \\ + \frac{1}{r}[\lambda p(\lambda h - u_h(p)) - \lambda p(1-p)u_h'(p)] \end{aligned}$$

Let

$$A(p) = u_l(p) - 2(\underline{s} + \overline{s}) + \lambda hp$$

$$B(p) = u_h(p) - 4\overline{s} + \lambda hp$$

Clearly, $S(p) = q_0 A(p) + (1 - q_0)B(p)$.

Notice that if the other player is type $\underline{s}$, the on-path outcome is exactly the same as the outcome of the symmetric MPE in the KRC model with $s = \underline{s} + \overline{s}$. So for $p_t > p_c(2\overline{s})$[22]

$$A(p) = u_l(p) - 2(\underline{s} + \overline{s}) + \lambda h p > 0$$

It is easy to see $A(p)$ is increasing in p for $p > \hat{p}_2$, so $A(p) > A(p_c(2\overline{s}))$. Recall our assumption on $\overline{s}$ and $\underline{s}$ s.t. $\hat{p}_2 < p_c(2\overline{s})$ and the results in Keller, Rady and Cripps (2005), we have $A(\hat{p}_2) = 0$ and $A(p_c(2\overline{s})) > 0$. So, fixing $\underline{s}$ and $\overline{s}$, for $p > p_c(2\overline{s})$, $\frac{-B(p)}{A(p)}$ will be bounded above, where the positive upper bound is determined by $\overline{s}$ and $\underline{s}$.

Then we conclude that by choosing q_0 large enough (the threshold related to $\underline{s}$ and $\overline{s}$), we can have inequality (A.3) satisfied for $p_t > p_c(2\overline{s})$. In other words, the player who has revealed himself does not deviate from $k = 1$ at $p_t > p_c(2\overline{s})$.

Similarly, we can have q_0 large enough such that the one-sided private information problem where type $\underline{s}$ reveals himself has an equilibrium with the pooling stage and the full separating stage, cut by point $p_c(\underline{s} + \overline{s})$.

Appendix A.3. Welfare results

This section provides a proof for Proposition 2.

As discussed in the main context, I will compare the ex-ante welfare of my model

$$2EW(p_0, q_0) = 2q_0 W(p_0, q_0; \underline{s}) + 2(1 - q_0)W(p_0, q_0; \overline{s})$$

to the ex-ante welfare of the full information benchmark

$$EW_b(p_0, q_0) = 2\Big[q_0^2 W_b(p_0, 2\underline{s}) + 2q_0(1 - q_0)W_b(p_0, \underline{s} + \overline{s}) + (1 - q_0)^2 W_b(p_0, 2\overline{s})\Big]$$

which is the same as the KRC model.

[22] We have put assumptions on $\overline{s}$ and $\underline{s}$ so that the KRC strategy of $s = \underline{s} + \overline{s}$ is $k = 1$ at $p_c(2\overline{s})$, i.e. $\hat{p}_2 < p_c(2\overline{s})$.

It easy to see that

$$
\begin{aligned}
2EW(p_0,q_0)-&EW_b(p_0,q_0) = \\
&2\Big\{q_0\Big[W(p_0,q_0;\underline{s}) - q_0W_b(p_0,2\underline{s}) - (1-q_0)W_b(p_0,\underline{s}+\overline{s})\Big] \\
&+(1-q_0)\Big[W(p_0,q_0;\overline{s}) - q_0W_b(p_0,\underline{s}+\overline{s}) - (1-q_0)W_b(p_0,2\overline{s})\Big]\Big\}
\end{aligned}
$$

Firstly, for type $\underline{s}$, according to the equilibrium construction, if the other player is type $\underline{s}$, the on-path strategies give the same payoff as in the KRC model of $s = 2\underline{s}$: $W_b(p_0, 2\underline{s})$. If the other player is $\overline{s}$, the on-path strategy is both players playing $k = 1$ or mixing both players playing $k = 1$ and $k = 0$, which gives more experimentation than the under-experimentation situation in the KRC model of $s = \underline{s} + \overline{s}$, without experimenting beyond the efficient threshold $p_c(\underline{s} + \overline{s})$. Consequently, the on-path strategies give a higher payoff than the payoff in the KRC model of $s = \underline{s} + \overline{s}$: $W_b(p_0, \underline{s} + \overline{s})$. So we have

$$W(p_0,q_0;\underline{s}) - q_0W_b(p_0,2\underline{s}) - (1-q_0)W_b(p_0,\underline{s}+\overline{s}) \geq 0$$

For type $\overline{s}$, in Appendix A.1, we have shown that the on-path payoff is no worse than the payoff after he revealing himself and entering a one-sided private information problem. Recall that in the equilibrium of the one-sided private information problem, if the other player is type $\underline{s}$, the on-path strategies give the same payoff as the KRC model of $s = \underline{s} + \overline{s}$; if the other player is type $\overline{s}$, the on-path strategies are efficient in experimentation. So, the payoff from revealing and entering the one-sided private information problem is larger than $q_0W_b(p_0, \underline{s} + \overline{s}) + (1 - q_0)W_b(p_0, 2\overline{s})$. Then we have

$$W(p_0,q_0;\overline{s}) - q_0W_b(p_0,\underline{s}+\overline{s}) - (1-q_0)W_b(p_0,2\overline{s}) \geq 0$$

Then we conclude

$$2EW(p_0,q_0) - EW_b(p_0,q_0) \geq 0$$

It is straightforward to check that the equality only happens at $p_0 = 1$.

The proof above also shows that the interim expected payoff of a player after knowing his type exceeds the full information benchmark.

Appendix A.4. Multiplicity

This section provides a proof for Proposition 3.

Recall restriction 1, since players start at the same q_0, q_t^i and q_t^j will only be different when someone reveals his type. So I will use q_t instead of q_t^i, q_t^j when there is no confusion.

Borrowing some ideas from Dong (2021), I prove the solution in a backward induction manner.

(1) $p_t \leq p_c(\underline{s} + \overline{s})$

Suppose now we have $0 < q_t^i, q_t^j < 1$, since the type is not revealed yet, on-path strategies of both types must be the same with a positive probability. Let this possibly partial pooling strategy be k_p.

If $k_p > 0$, once any player reveals himself, type $\overline{s}$ will be better off since $p_t < p_c(\underline{s} + \overline{s})$ and both players turn to $k = 0$ after revealing (according to restriction 2). So, type $\overline{s}$ wants to deviate to $k = 0$ and reveal himself[23].

If $k_p = 0$, type $\underline{s}$ can deviate to $k = 1$ for a moment to reveal himself and enters an one-sided private information problem. As discussed in Appendix A.2, he gets the KRC payoff of $s = 2\underline{s}$ if the other player is type $\underline{s}$; gets the efficient payoff if the other player is type $\overline{s}$ (restriction 2). Consequently, the deviation will be better than keeping at $k = 0$.

As a result, we need to have full separating in this part.

(2) $p_c(\underline{s} + \overline{s}) < p_t \leq p_s$

We have the following lemmas, if we have the same parameter conditions as in Proposition 1.

Lemma 2. *With 3 restrictions, at $p_t > p_c(\underline{s} + \overline{s})$, type $\underline{s}$ cannot do random revealing in an equilibrium, i.e. he cannot mix pooling with type $\overline{s}$ and revealing himself.*

[23]If he continues, he either continues with both players playing $k > 0$ or end up with someone revealing himself and $k = 0$ (this happens if revealing is possible in equilibrium). This continuation payoff will be worse than directly deviating, since keeping both players at $k > 0$ is worse than both players at $k = 0$ for type $\overline{s}$, $p_t \leq p_c(\underline{s} + \overline{s})$.

Proof. Suppose type $\underline{s}$ do random revealing in an equilibrium. Since here type $\underline{s}$ can reveal himself, then the partial pooling strategy needs to be $k < 1$ according to restriction 3.

When $p_t > p_c(\underline{s} + \overline{s})$, type $\underline{s}$ can deviate up to $k = 1$ to reveal himself (restriction 3) and enter an one-sided private information problem. Then according to restriction 2, the other player will choose $k = 1$ until $p_t = p_c(\underline{s} + \overline{s})$. After $p_t < p_c(\underline{s} + \overline{s})$, the other player plays the KRC strategy of $s = 2\underline{s}$ if he is type $\underline{s}$, and plays $k = 0$ if he is type $\overline{s}$. This will be better for type $\underline{s}$ than keeping at $k < 1$, no matter what the other player's type is. Consequently, type $\underline{s}$ cannot be indifferent between revealing himself and keeping at $k < 1$[24]. ■

Lemma 3. *With 3 restrictions, when type $\overline{s}$ is doing random revealing at $p_t \in (p_c(\underline{s} + \overline{s}), p_s]$, we must have $q_t = q_i(p_t)$.*

Proof. Since type $\underline{s}$ cannot do random revealing here, the only possibility here is that both types will choose some $k_0 > 0$, while type $\overline{s}$ has some rate of deviating down to reveal himself. Similar to Section 3.3, type $\overline{s}$'s payoff $u(p)$ satisfy the ODE

$$\begin{aligned} u(p) - Es =& k_0(\lambda hp - Es) + \frac{1}{r}\Big\{(1 - q(p))e_t[2\overline{s} - u(p)] \\ &+ [2k_0 p\lambda(\lambda h - u(p)) + \frac{du(p)}{dp}\frac{dp}{dt}]\Big\} \end{aligned} \tag{A.4}$$

where $Es = q(p)(\underline{s} + \overline{s}) + (1 - q(p))2\overline{s}$, e_t is the arrival rate of revealing.

Since we need indifferent between continuing and revealing, so $u(p) = Es$. Then combine $u(p) = Es$, (4) and (A.4) we get

$$Es - \lambda hp = \frac{1}{r}(2p\lambda(\lambda h - Es)) \Rightarrow q(p) = \frac{2\overline{s}r - p[\lambda h(2\lambda + r) - 4\lambda\overline{s}]}{(\overline{s} - \underline{s})(2p\lambda + r)} = q_i(p)$$

■

Lemma 4. *With 3 restrictions, when $p_t \in (p_c(\underline{s}+\overline{s}), p_s]$, full separation cannot happen in an equilibrium.*

[24] Actually, two strategies give the same payoff at $p_c(\underline{s} + \overline{s})$, since full separating happens after that, as discussed above.

Proof. Suppose full separation happens in an equilibrium and type $\overline{s}$'s on-path strategy is k', type $\underline{s}$'s on-path strategy is k''. According to restriction 3, $k'' > k'$. Then what type $\overline{s}$ will do is that he chooses k'' for a moment and wait for the other player reveals himself. After that, type $\overline{s}$ continues at k'' until $p_c(\underline{s}+\overline{s})$ if the other player is type $\underline{s}$; chooses $k = 0$ (and thus reveals himself as type $\overline{s}$) if the other player is type $\overline{s}$. By this way, he is better off than staying on-path if the other player is type $\underline{s}$, since $p_t > p_c(\underline{s} + \overline{s})$; he is no worse than staying on-path if the other player is type $\overline{s}$, since $p_t \leq p_s < \hat{p}_1 < p_c(2\overline{s})$. ■

Then use backward induction, suppose we have q_t where no revealing happens stay on-path for $p_t \in [p_c(\underline{s}+\overline{s}), p]$. Then for the interval $[p, p+dp]$, according to Lemma 2 and Lemma 4, we will not have full separation and type $\underline{s}$ randomly revealing. If we have $q_t < q_i(p_t)$ in $[p, p + dp]$, then to make q_t match $q_i(p_t)$ again at p, we need type $\overline{s}$ randomly revealing to drive up q_t. But according to Lemma 3, we then need to have $q_t = q_i(p_t)$, which is a contradiction.

If we have $q_t > q_i(p_t)$ in $[p, p + dp]$, since there is no full separation and no random revealing from type $\underline{s}$, the only possibility is that two types use the pooling strategy[25]. Furthermore, the pooling strategy cannot be $k < 1$, since that means type $\underline{s}$ will want to deviate to $k = 1$ to reveal himself and enter an one-sided private information problem, due to the similar reason as in the proof for Lemma 2. But if two types pool at $k = 1$, type $\overline{s}$ will not want to mix revealing and mimicking for any larger p_t. Recall the proof in Appendix A.1.3, q_0 is large enough to ensure that mimicking is better for type $\overline{s}$ with any $p_t > p_s$. Here we have two types pooling at $q_t > q_i(p_t) > q_0$. Consequently, the result in Appendix A.1.3 still holds here – for any larger p_t than the current pooling interval, type $\overline{s}$ is better off with mimicking (using $k = 1$) and cannot do partial revealing. But the game starts with q_0, so without the random revealing of type $\overline{s}$, we cannot get our q_t here, which is between q_0 and 1. Now we have a contradiction for having a pooling interval.

So, for $p_c(\underline{s} + \overline{s}) < p_t \leq p_s$, we can only have $q_t = q_i(p_t)$.

[25]Notice that by using type $\overline{s}$ randomly revealing we cannot get $q_t > q_i(p_t)$.

(3) $p_t > p_s$

In this part the on-path q_t is q_0. Suppose we already have q_t on-path for $p_t \in [p_s, p]$. Then let us consider the interval $[p, p + dp]$. If we have $q_t > q_0$, to let q_t match q_0 again at $p_t = p$, we need a gradual decreasing in q_t as time goes by, which means we need random revealing from type $\underline{s}$. But it cannot be achieved due to Lemma 2.

If we have $q_t < q_0$, then before this moment, we must have a gradual decreasing of q_t in time to make the initial belief q_0 go down to this q_t, which again requires random revealing from type $\underline{s}$. But it is not achievable.

So, we can conclude that for $p_t > p_s$, q_t needs to stay at q_0.