Strategic Experimentation with Two-sided Uncertainty

Yihang Zhou*

February, 2023

Abstract

Two players have access to replicas of a two-arm bandit where both arms have unobserved information. By using the learning arm, players generate public information about the unknown state of this arm. By using the no-learning arm, players do not generate information. Instead, both players have private information about this arm. In this two-arm bandit model with two-sided uncertainty in the no-learning arm, I construct an equilibrium where the under-experimentation caused by free-riding is mitigated, and sometimes overturned to over-experimentation. Players experiment more due to the incentive of encouraging the other player to experiment and the ignorance towards the true state of the no-learning arm. Though there can be over-experimentation, two-sided uncertainty increases the total ex-ante welfare.

Keywords: Strategic experimentation, two-arm bandit, two-sided uncertainty

*Department of Economics, University of Texas at Austin, yhzhou@utexas.edu

1 Introduction

In the strategic experimentation problem where two players have access to replicas of a two-arm bandit, we have an under-experimentation problem due to free-riding (Keller, Rady and Cripps, 2005). But if both two players have private information about the bandit, both of them may want to experiment more to hide their private information and keep the other one experimenting. Furthermore, since no one knows all private information, they may over-experiment.

I study a two-player strategic experimentation problem with the following two-arm bandit: one arm is the *learning arm*, which is the same as the risky arm in Keller, Rady and Cripps (2005) and players can generate information by directly using this arm; the other arm is the *no-learning arm*, which generates constant flow payoff. However, no player observes the flow payoff and both players have some private information about the flow payoff. The payoff of the no-learning arm is realized in the far future, so players do not directly generate information by using this arm. Instead, they infer the other one's private information from past actions.

With the two-sided private information (two-sided uncertainty henceforth) on the no-learning arm, players experiment more than the full information benchmark, which has no private information[1]. There are two reasons for more experimentation. Firstly, experimenting is a signal of a lower flow payoff in the no-learning arm, which can encourage the other player to experiment more. Secondly, no one knows the no-learning arm for sure means that they may experiment more than needed. For example, one possible situation is that players will stop experimenting if they know the true flow payoff of the no-learning arm. However, with private information, both players think that the true flow payoff may be a lower level, so they may continue to experiment. The second reason not only can lead to more experimentation than the case without private information, but it can also cause an over-experimentation problem – more experimentation than the efficient level.

[1]Without private information, the model will be the same as in Keller, Rady and Cripps (2005).

Though we can face an over-experimentation problem with two-sided uncertainty, the ex-ante welfare increases compared to the benchmark. The ex-ante welfare is the total expected payoff of players before they observe their private information. Though in some outcomes, the over-experimentation problem reduces the welfare, in other possibilities, the increasing experimentation mitigates the under-experimentation problem in the benchmark without over-experimentation and thus increases the welfare. The loss from the over-experimentation will be covered by the gain of other possibilities. Furthermore, the intermediate expected payoffs of players right after knowing their types also exceed the expected full information payoffs. But if one player has the chance of verifiably revealing their private information at the beginning while the other cannot , the player with certain private information will want to do that.

The no-learning arm tries to capture the situation where the backup option of a risky investment is not completely safe, and both decision-makers do have some extra knowledge about this backup option. The knowledge is private information for decision-makers, so the uncertainty on the no-learning arm is two-sided.

We can use oil companies as an example. But now besides exploring a new site (the learning arm), companies can invest in some academic geography research. This long-run research gives no feedback immediately, so no companies will know exactly what the flow payoff is when investing. But different companies have some private accumulation of research, so their investment can reveal their private information. Another possible application is that companies face choices between a risky new project and a financial asset. The market is not perfectly competitive, so there is some uncertainty about the asset's payoff. And the asset only generates payoffs in the far future so companies do not learn about the uncertainty by investing in it. Instead, they can have some private information.

The organization of this paper is as follows. Section 2 is the setup of the model. Section 3 discusses the main messages. Section 4 presents the welfare discussions. Section 5 includes the multiplicity problem and other discussions. Section 6 is the concluding remark.

Related literature. – Bolton and Harris (1999) firstly introduced the strategic experimentation problem and characterized the unique symmetric MPE. Keller, Rady and Cripps (2005) introduced an exponential bandit, and constructed a unique symmetric MPE with under-experimentation, due to free-riding. Keller and Rady (2010) expanded the perfect good news model in Keller, Rady and Cripps (2005) to a Poisson bandit where there is no decisive information. Keller and Rady (2015) visited the risky arm whose events are breakdowns. These papers discussed the encouragement effect and free-riding effect in strategic experimentation and found that MPEs feature under-experimentation compared to the collaborative benchmark.

The classic analysis of strategic experimentation assumes that all information is public except for the true state of the learning arm. This paper follows the trend of allowing private information of players in strategic experimentation problems, which adds new dimensions of learning besides learning the unobserved state of the learning arm. Bonatti and Hörner (2011) discussed the effort invested in experimentation when it is unobserved by other players. Heidhues, Rady and Strack (2015) assumed unobserved realization of risky payoffs in the two-arm bandit model. In a more closely related paper, Dong (2018) allowed 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. In contrast, I assume both two players have private information about the no-learning arm. The two-sided uncertainty also has an encouraging effect and leads to more experimentation. However, over-experimentation will be a problem in my model due to ignorance about the true flow payoff of the no-learning arm. This paper explores the possibility of further mitigating the free-riding problem in strategic experimentation. It also points out the possibility of over-experimentation, which overturns the effect of free-riding.

Two-sided uncertainty in my model possibly causes an undesired result – over-experimentation. The detrimental effect of two-sided uncertainty also happens in other situations. For example, in the bargaining model with two-sided uncertainty by Cramton (1992), two players may waste time bargaining with each other but end up figuring out that a transaction cannot be made since the seller values more than the buyer towards the item. 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 include papers like Kahn and Huberman (1988) and Banks (1993).

But the overall effect of two-sided uncertainty in my model is beneficial – the ex-ante welfare improves. This happens due to the mitigation of the free-riding problem. So the two-sided uncertainty can serve as a potential solution to the free-riding problem, added to other solutions like communication (Isaac and Walker, 1988), asymmetric information (Dong, 2018; Komai, Stegeman and Hermalin, 2007), multiple-order sanctions (Kiyonari and Barclay, 2008), etc.

2 Model

Time is continuous, $t \in [0, \infty)$. Two players (P_1 and P_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. The division of the resource at each date is public. And players discount future payoffs by the factor r.

2.1 Two-arm bandit

The learning arm's payoff depends on the state θ, which is not observed by players. When the state is $\theta = 1$, 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 $\theta = 0$, the learning arm generates no payoff. So it is a perfect good news model. Furthermore, two players' learning arms are governed by the same state θ, and the payoff arrivals of the learning arms are public, so players 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. $s_1, s_2 \in \{\underline{s}, \overline{s}\}$ are random variables drawn independently at the beginning of the game. Both of them have the probability of $q_0 \in (0,1)$ being $\underline{s}$, and the probability of $1 - q_0$ being $\overline{s}$. P_i only observes s_i, which is his private type. I also assume that the flow payoff s is not observed until date T, which is far in the future, so players do not learn by using this arm. Instead, they 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, the nature decides s_1 and s_2 and P_i observes the realization of s_i. At date t, P_i decides the resource $k_i \in [0,1]$ on the learning arm. He makes this decision based on his type, the past actions of two players, and the past payoff outcomes on the learning arms. The space of past actions is $[0,1]^{[0,t)}$, and the space of the past outcomes is $\{0,1\}^{[0,t)}$, where 1 represents payoff arrival. The action of P_i at date t is:

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

which is the resource on the learning arm.

Following the literature, I focus on the Markov Perfect Equilibrium. At date t, the public information (past actions and past outcomes) can be summarized in three state variables: p_t, the belief of the learning arm's state being $\theta = 1$; q_t^1, P_2's belief of P_1's type being $\underline{s}$; q_t^2, P_1's belief of P_2's type being $\underline{s}$[2]. We will write the Markov strategy of P_i as $k^i(p_t, q_t^i, q_t^j; s_i)$, where

[2]More rigorously, we will have two extra state variables, P_1's second order belief about P_2's belief of s_1 and P_2's second order belief about P_1's belief of s_2. But since the past actions are public information in this

s_i is the private type, q_t^i is the player's own belief about the other player's type, and q_t^j is the other player's belief about this player's type (and henceforth). I will focus on symmetric MPE – players will choose the same action if they have the same p_t, q_t^i, q_t^j, and s_i. Finally, I require the strategy to be left continuous in p_t to avoid further problems.

Besides the pure strategies, I also allow for a specific kind of mixed strategies: players can randomly stop and move to the no-learning arm, i.e. mixing stopping and continuing experimentation at a date.

An symmetric MPE consists of a strategy function $k^i(p_t, q_t^i, q_t^j; s_i)$ 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 the player i observes an action of k_j from the other player. At each date, the strategy $k^i(p_t, q_t^i, q_t^j; s_i)$ maximizes the expected payoff of player i

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

given the other player's strategy and belief updating. Besides, 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[3], players will play the unique symmetric equilibrium presented in Keller, Rady and Cripps (2005) (the KRC strategy henceforth). Though they also provided asymmetric equilibria, I will focus on the symmetric solution.

2.3 State variables

My model depart from the model in Keller, Rady and Cripps (2005) (the KRC model henceforth) and the variant on the no-learning arm does not affect the information accumulation on the learning arm. If there is payoff arrival on the learning arm, then two players

game, the second order beliefs are equal to the beliefs of the other player. So, I compress the state variables a bit for the simplicity.

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

know $\theta = 1$. If there is no payoff arrival, the update of p_t follows the law of motion:

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

where K_t is the total resource at date t on the learning arms by two players. The belief of the learning arm p_t decreases as time goes by.

The updates of q_t^1 and q_t^2 completely depend on the strategies in equilibrium. If the other player of type $\underline{s}$ and $\overline{s}$ chooses the same action (pooling), then q_t^i does not change. Otherwise they updates according to the Bayes rule. For example, suppose in the whole history, type $\underline{s}$ chooses $k_t = 1$, type $\overline{s}$ jump to $k_t = 0$ from 1 at an arrival rate e_t, then the belief of type $\underline{s}$ at time t (q_t^i) seeing $k = 1$ 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 the equilibrium.

3 Equilibria Results

3.1 The cooperative solution and benchmark

When two players work cooperatively, they will share their private information, so the model degenerates into the KRC model. As a result, the cooperative solution of my model is the same as their model. Players will invest all resource into the learning arm when $p_t > p_c(s)$ and into the no learning arm when $p_t \leqslant p_c(s)$, 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 it will be the benchmark. To be more specific, I

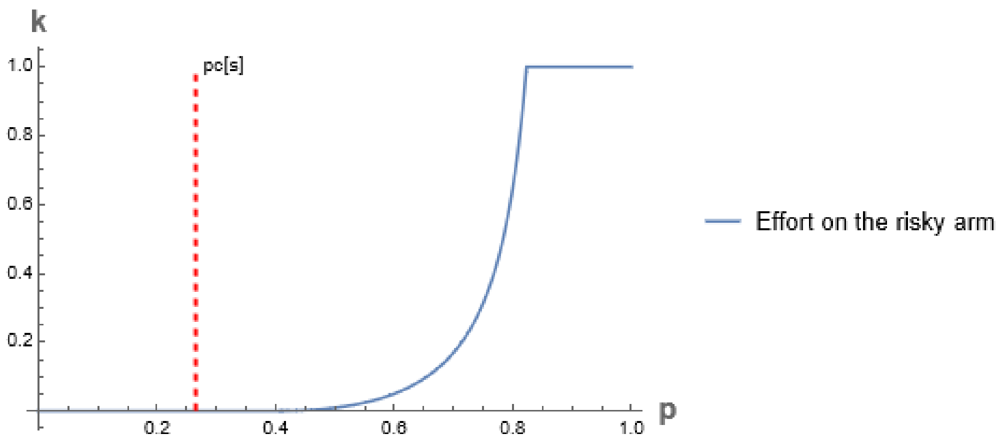


Figure 1: The KRC strategy and the cooperative threshold

will use the unique symmetric MPE in the KRC model as the benchmark, since I also look at a symmetric MPE. The free-riding leads to under-experimentation in this solution.

For the purpose of illustration, I present the KRC strategy[4] (the blue line) and the cooperative threshold $p_c(s)$ (the red dotted line) in Figure 1.

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. $\overline{s}$ 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 $s = \underline{s} + \overline{s}$ needs to be $k_t = 1$, as shown in the Figure 2.

Clearly, if the learning arm's payoff arrives, players know $\theta = 1$ and will put all resource into the learning arm. Then 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 $p_c(\underline{s} + \overline{s}) < p_s < p_c(2\overline{s})$.

Notice that the following strategies are conditional on no payment arrival of the learning arm and the beliefs about the other player's type (q_t^1, q_t^2) staying on-path (discussed later).

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

[4]The strategy is presented as a function of p_t

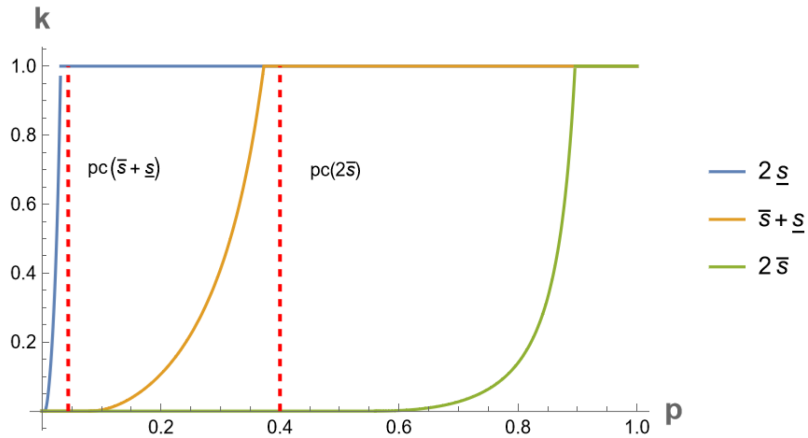


Figure 2: The KRC strategies of three possible realizations of s

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

When $p_c(\underline{s}+\overline{s}) < p_t \leqslant p_s$, type $\underline{s}$ experiments with resource $k = 1$ while type $\overline{s}$ randomly stops, which are partial separating strategies. To be more specific, at the time interval $(t, t+dt)$, type $\overline{s}$ stops experimenting with probability $e_t dt$[5], i.e. the stopping happens with an arrival rate of e_t, which may change with time, see the next subsection for a detailed construction. If the stopping happens, the other player knows that the type is $\overline{s}$; if the stopping does not happen, q_t^j increases as time goes by (also decreases in p_t, since p_t decreases in time). This stage is called *the partial separating stage.*

When $p_t \leqslant p_c(\underline{s}+\overline{s})$, there will be full 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. If we have (q_t^i, q_t^j) where $q_t^j < 1$, he plays $k = 1$ to reveal himself; if $(q_t^i, q_t^j) = (0,1)$, he plays $k = 0$, which is the KRC strategy of $s = \underline{s}+\overline{s}$; if we have (q_t^i, q_t^j) where $q_t^j = 1$, $0 < q_t^i < 1$, he plays $k = 1$ to keep $q_t^j = 1$ and wait for the self-disclosure of the other player. Type $\overline{s}$ does not experiment in this stage. It can be regarded as full separation, since type $\overline{s}$ always play $k = 0$ and type $\underline{s}$ plays $k = 1$ when the other one does not know his type[6]. This stage is called *the full separating stage.* Notice that the action $k = 1$ for (q_t^i, q_t^j) where $q_t^j = 1$, $0 < q_t^i < 1$ is simply to keep q_t^j at 1,

[5] $(dt)^n$ with a power no less than 2 will be omitted

[6] When p_t is very small, both players do not experiment regardless of the other player's type, so there is no need to distinguish this situation from full separation.

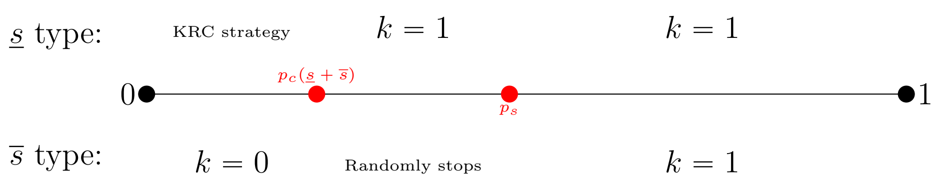


Figure 3: On-path strategies

since according to the strategies, q_t^i will jump to either 0 or 1 at the next moment due to the revealing of the other player. If there is no concerns about q_t^j, type $\underline{s}$'s action is arbitrary at this moment, so he just chooses $k=1$ to keep the reputation of type $\underline{s}$.

Figure 3 summarizes these strategies.

We can have the on-path beliefs q_t^i when no stopping of the experimentation happens and no payoff arrives on the learning arm as a function of p_t by Bayes rule:

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

which is also presented in Figure 4. When stop happens in the partial revealing stage, q_t^i jumps to 0. We will see in the next subsection that $q_i(p_t)$ reaches 1 when p_t arrives at $p_c(\underline{s}+\overline{s})$. In other words, if no stopping is observed before p_t arrives at $p_c(\underline{s}+\overline{s})$, the player is sure that the other player is type $\underline{s}$.

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 side reveals his private signal, the problem becomes a one-sided uncertainty problem similar to Dong (2018). A detailed construction of the equilibrium of this subgame is in section 3.4 and Appendix A.2.

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

In this equilibrium, over-experimentation may happen. When the true state is $s = 2\overline{s}$,

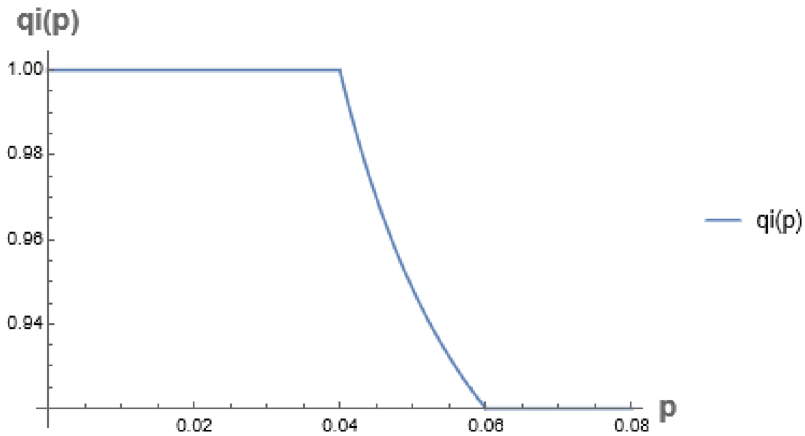


Figure 4: On-path belief

the outcome of this equilibrium is[7]:

Suppose the initial prior about the learning arm $p_0 > p_s$, then at the beginning, both players play $k_t = 1$. As p_t goes down and reach p_s, both players (they are type $\overline{s}$) randomly stop at an arrival rate e_t. With probability 1, the stop happens before p_t reach $p_c(\underline{s}+\overline{s})$. After the stop happens, since $p_t \leqslant p_s < p_c(2\overline{s})$, both players use the no-learning arm according to the equilibrium in the one-sided uncertainty problem.

We can see that both players will experiment at $p_t < p_c(2\overline{s})$ since $p_s < p_c(2\overline{s})$. But when they work cooperatively, they will stop at $p_c(2\overline{s})$. We have over-experimentation here. We have more experimentation than the benchmark due to two reasons. Firstly, experimenting is a signal of a lower opportunity cost of experimentation, and thus leads to more experimentation from the other player. This encouragement effect mitigates the free-riding problem and cause more experimentation, but it does not lead to over-experimentation. Another reason for more experimentation is the ignorance towards the true s. In this example, no player knows that the true state is $s = 2\overline{s}$. Instead, both of them believe that the other player might be type $\underline{s}$. Since they think the opportunity cost of experimenting might be low, they will experiment more than needed. They are aware of that they have over-experimentation if the state is $2\overline{s}$, but the expected gain from other possible realizations of s can cover the

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

loss from over-experimentation and they continue to experiment, even if they might pass the efficient threshold. In this equilibrium, the selfishness will cause more experimentation, but only with the ignorance towards the true s, the experimentation level will exceed the efficient level.

The equilibrium above is summarized in the following proposition.

Proposition 1 *$\exists\, c_1(r,\lambda,h)$, $c_2(r,\lambda,h)$ s.t. for any $\overline{s} \in (c_1(r,\lambda,h), \frac{\lambda h}{2})$ and $\underline{s} \in (0, c_2(r,\lambda,h))$, there 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 $p_0 > p_s$[8], we can 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 $p_t > p_c(\underline{s}+\overline{s})$; plays the KRC strategy of the true s if $p_t \leqslant p_c(\underline{s}+\overline{s})$

(2) Type $\overline{s}$ plays $k_t = 1$ if $p_t > p_s$; plays $k_t = 0$ if $p_t \leqslant p_c(\underline{s}+\overline{s})$; randomly stops to the no-learning arm if $p_s < p_t \leqslant p_c(\underline{s}+\overline{s})$

(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 $\underline{s}$ use the learning arm with full resource; type $\overline{s}$ randomly stops at an arrival rate of e_t.

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

For type $\overline{s}$, since he randomly stops, he needs to be indifferent between continuing on $k_t = 1$ and revealing himself by $k_t = 0$. Assuming type $\overline{s}$'s expected payoff with no one revealing himself in this stage to be $u(p_t, q_i(p_t))$, it should satisfy the following HJB equation:

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

[8] As discussed above, p_s depends on q_0.

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 (dt to the power higher than 1 is omitted).

Furthermore, the expected payoff should be equal to the expected payoff of revealing by $k = 0$, since the player is indifferent between mimicking and revealing. We 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 large enough (discussed later). And under this assumption, the 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 equation holds because that when $p_t \leqslant p_s$ and one player reveals himself as type $\overline{s}$, the other player will reveal himself 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 we have:

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

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:

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

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, which also says that 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 stop happens before p_t reaching $p_c(\underline{s}+\overline{s})$ with probability 1. After the stop happens, the game becomes a one-sided uncertainty game which is discussed in the next subsection.

p_s is also decided by (6). It solves $q_i(p_t) = q_0$. And 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 $q_i(p_c(2\overline{s}))$. By assuming q_0 large enough, we can ensure that the partial separating stage happens late enough[9] 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 be over-experimentation and incur 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. But as p_t goes down with time, the loss of over-experimentation becomes relatively larger compared to the gain of other possibilities. When the gain can no longer cover the loss, type $\overline{s}$ begins to randomly stop. When stop happens, the game proceeds as a one-sided uncertainty problem. When stop does not happens, q_t^i and q_t^j increase with time. Though as time t increases, the loss becomes relatively larger, 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 $k = 1$ and revealing himself by $k = 0$.

3.4 One-sided uncertainty problem

The one-sided uncertainty problem after one player revealing himself is quite similar to the model in Dong (2018), so I will omit the detailed algebra and focus the main construction of the equilibrium where type $\overline{s}$ reveals himself in this section[10]. Notice that according to

[9] 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

[10] Again, I present the equilibrium without payoff arrival on the learning arm, which is the non-trivial part.

the equilibrium construction above, before the full separating stage, only type $\overline{s}$ is able to reveal himself, i.e. q_t^i can only jumps to 0. So I only discuss the equilibrium after type $\overline{s}$ revealing himself.

If q_0 is large enough, we can have an equilibrium of only two stages, one *pooling stage* and one *full separating stage*, with the cutoff point $p_c(2\overline{s})$. Notice that in this equilibrium players achieve the efficient level of experimentation when the other player is also type $\overline{s}$.

The detailed construction of this subgame equilibrium is in Appendix A.2.

From the construction, we can see that the full separating stage happens after p_t reach $p_c(2\overline{s})$. But in the equilibrium of the two-sided uncertainty problem, revealing (type $\overline{s}$ move to the no-learning arm) only happens when $p_t \leqslant p_s$, 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 uncertainty problem, when type $\overline{s}$ reveals himself, the full separating happens in the continuing one-sided uncertainty problem, so the other player also reveals himself after the revealing.

Notice that we can suffer from the multiplicity problem in this one-sided uncertainty problem, but I will assume that players play the equilibrium above, where type $\overline{s}$ keeps his private information as long as possible.

4 Welfare

4.1 Ex-ante welfare

Firstly, I discuss the ex-ante welfare of the equilibrium. The ex-ante welfare is the expected total payoff of two players before knowing the types.

Before knowing the type, 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 $p_0 = p$, $q_0 = q$.

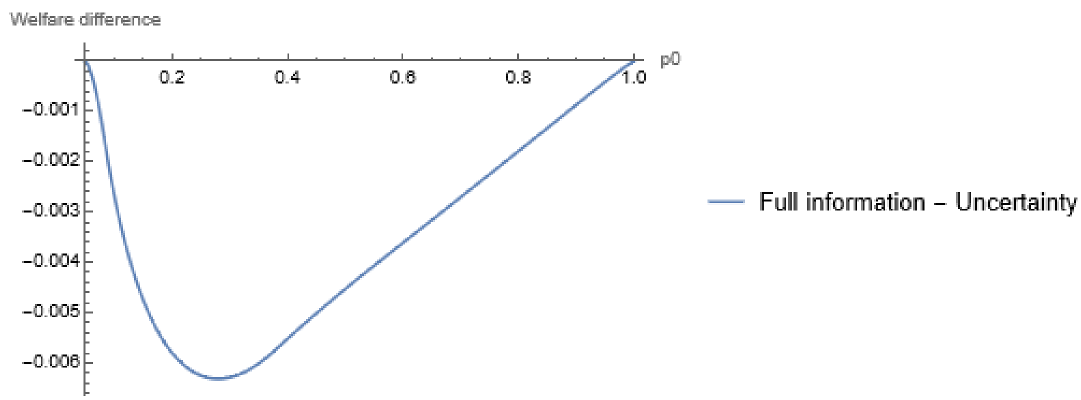


Figure 5: Welfare difference, $p_0 > p_s$

Notice that 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. Suppose the player in the benchmark with flow payoff s has a initial expected payoff of $W_b(p_0; s)$ when the initial belief about the state of the learning arm is p_0. Then the ex-ante welfare of my model will be compared 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 uncertainty improves the ex-ante welfare compared to the full information benchmark, if the equilibrium in Proposition 1 exists.*

Figure 5 illustrate the ex-ante welfare difference between the full information benchmark and the two-sided uncertainty model $(EW_b(p_0, q_0) - 2EW(p_0, q_0))$ under different initial belief p_0. When $p_s < p_0 < 1$, the two-sided uncertainty 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.

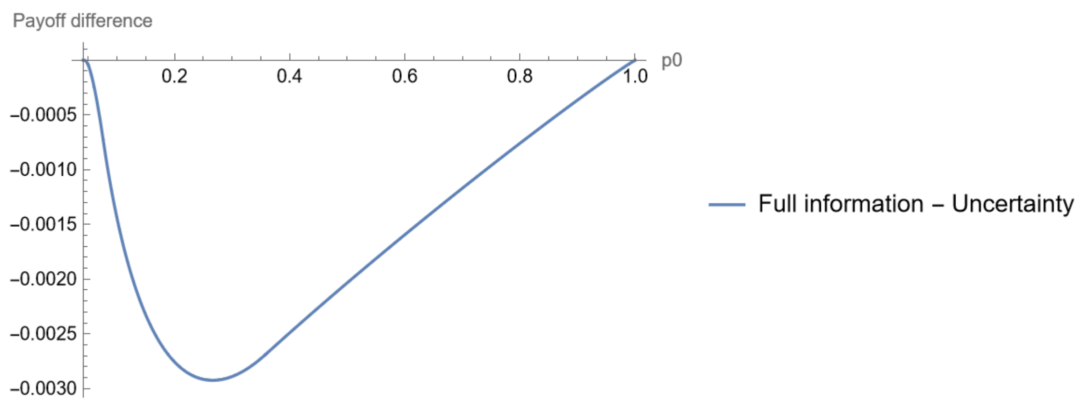
Though two-sided private information leads to the possibility of over-experimentation, when compared to the full information benchmark, the ex-ante welfare is increased. The intuition of this increase is similar to the reason of having over-experimentation. There is a possible welfare loss compared to the benchmark when over-experimentation happens, but in other realizations of s where over-experimentation does not happen, there is a payoff gain – more experimentation can mitigate the under-experimentation problem caused by free-riding. When we consider the ex-ante welfare, the welfare gain can cover the welfare loss.

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

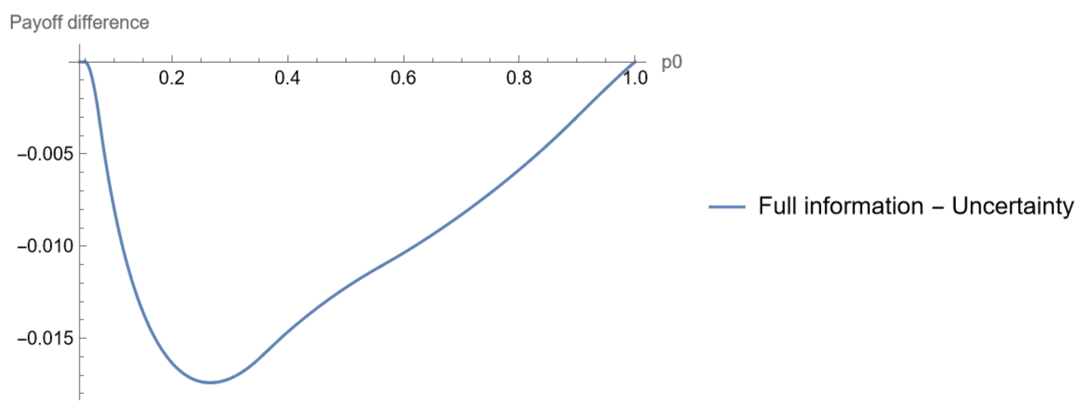
4.2 Intermediate payoffs

In this part I discuss the intermediate payoffs – to be more specific, the expected payoff of players at the beginning but after knowing their types. Clearly, $W(p_0, q_0; s)$ is the intermediate 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, my model has a higher intermediate payoff than the full benchmark for both types. For type $\underline{s}$, having uncertainty mitigates the under-experimentation problem without incurring over-experimentation. But for type $\overline{s}$, having uncertainty leads to more experimentation, and over-experimentation is possible. Even after knowing their types, players do not want to go full information.



(a) Type $\underline{s}$



(b) Type $\overline{s}$

Figure 6: Intermediate payoff difference, $p_0 > p_s$

We may also consider the situation that one players have a chance of verifiably communicating their private information to the other one and making the problem a one-sided one. Type $\overline{s}$ does not want to communicate his private information. Notice that even without this communicating chance, type $\overline{s}$ can revealing himself 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. We cannot get the same result for type $\underline{s}$ by the same way, since type $\underline{s}$ does not have the chance to reveal himself in the equilibrium.

We can compare the equilibrium result of the two-sided uncertainty problem and the one-sided uncertainty problem for type $\underline{s}$. If the other player is also type $\underline{s}$, then in both one-sided and two-sided uncertainty problem, the equilibrium result is that both players play the KRC strategy of $s = 2\underline{s}$. If the other player is type $\overline{s}$, we need to compare the experimentation level of two problems.

When q_0 is high at the beginning, then the one-sided uncertainty 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, players achieve the efficient level of experimentation when the other player is type $\overline{s}$. But in the two-sided uncertainty problem, players may stop experimentation before $p_c(\underline{s} + \overline{s})$, so the one-sided problem gives a better equilibrium payoff and type $\underline{s}$ wants to communicate if he has the chance.

5 Discussion

5.1 Multiplicity problem

As a signaling game, we will face the multiplicity problem. Similar problems also happen in Dong (2018). There are multiple equilibria for the game with two-sided uncertainty as well as for the subgame with one-sided uncertainty (after one player reveals his private information). Actually, even after both sides reveal private information and the game degenerates to the full information benchmark, we have multiple MPEs, but I restrict the attention to 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 play the same strategy will resolve the multiplicity problem of the subgame after both sides reveal the private information. For the subgame where only one player reveals his private signal, this restriction does not apply, since two players are themselves asymmetric here. For the original game with two-sided uncertainty, the symmetric requirement is that two player should use the same strategy if they are of the same type, as mentioned in section 2.2.

The second restriction (**restriction 2**) is on the one-sided uncertainty problem after one player reveals his private type. The one-sided uncertainty problem also suffers from the multiplicity problem. Following the similar idea as Dong (2018), I assume that facing the one-sided uncertainty problem, players play the equilibrium where the private information is held for the longest time. And this equilibrium is the one 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 $k < k(\underline{s})$ will lead to $q_t^j = 0$ and any action $k > k(\underline{s})$ will lead to $q_t^j = 1$ if q_t^j is not 0 before jump, i.e. any deviation from the on-path strategy has deterministic effect on q_t^j. But for $q_t^j = 0$, deviating to $k > k(\underline{s})$ does not change q_t^j anymore. This restriction has the similar intuition as D1 criterion but is stronger. For type $\underline{s}$, the opportunity cost of experimenting is small, so he tends to experiment more. Similar for type $\overline{s}$. So, a larger action is more likely to be used by type $\underline{s}$, and I assume this consideration has full effect on the beliefs about the private type. This restriction also says that if there is full separation, the strategy of type $\underline{s}$ must be higher than type $\overline{s}$. Besides, this restriction also says that 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[11].

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 beliefs* (q_t^1, q_t^2) *in a MPE is unique.*

Here is a heuristic proof idea of the proposition, please refer to Appendix A.4 for a complete proof.

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

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

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

Fourthly, pooling cannot happen $p_t \leqslant p_s$. If they pools on $k < 1$, type $\underline{s}$ will deviate up to reveal himself, and then gets a better payoff as discussed in section 4.2. If they pools on $k = 1$ for $p_t < p_s$, the belief 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 $p_t > p_s$, we will need q_t^i and q_t^j to remain unchanged at q_0 due to no partial revealing from type $\underline{s}$.

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

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

5.2 More players

A natural question is what would happen if there are more than two players. For example, in the setup of 3 players, I assume that the no-learning arm generates flow payoff $s = s_1 + s_2 + s_3$. $s_1, s_2, s_3 \in \{\underline{s}, \overline{s}\}$. And the distribution of all three variables are $q_0 \circ \underline{s} + (1-q_0) \circ \overline{s}$. All other setups are the same as the original model.

More players lead to more free-riding, more encouragement, and even more uncertainty[12]. As discussed, the free-riding problem makes players experiment less; players may want to experiment more to hide private information from others to encourage them to experiment more; the ignorance towards the true state can make players experiment beyond the efficient boundary.

More players do not change the result qualitatively. Take 3 players as an example, when $\overline{s}$ large enough and $\underline{s}$ small enough, in equilibrium, we still have 3 stages if q_0 and p_0 large enough: the pooling stage, the partial separating stage and the full separating stage. This equilibrium is governed by two cutoffs: $p_c(\overline{s} + 2\underline{s})$ and p_s^3. For $p_t > p_s^3$, both types play $k = 1$; for $p_c(\overline{s} + 2\underline{s}) < p_t \leqslant p_s^3$, type $\underline{s}$ plays $k = 1$, type $\overline{s}$ randomly stops from $k = 1$ to $k = 0$; for $p_{\leqslant} p_c(\overline{s} + 2\underline{s})$, type $\overline{s}$ plays $k = 0$, 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 other players 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 $q_i^3(p_t) = q_0$.

Though the full separating happens earlier than the original model ($p_c(\overline{s}+2\underline{s}) > p_c(\overline{s}+\underline{s})$), but this is because that the flow payoff of the no-learning arm – the opportunity cost of experimentation – is larger. If we make the highest and lowest realization of s the same in three-player and two-player cases, i.e. $s_1, s_2, s_3 \in \{\frac{2}{3}\underline{s}, \frac{2}{3}\overline{s}\}$, the three-player case's full separating happens later than the original model ($p_c(\frac{2}{3}\overline{s} + \frac{4}{3}\underline{s}) < p_c(\overline{s} + \underline{s})$). So after re-

[12] Because each player will have private information

moving the effect of the changing flow payoff of the no-learning arm, the overall effect of increasing the number of players to 3 encourages experimentation. So we can see that 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 due to the uncertainty. The benefit of mitigating the free-riding problem covers the loss of possible over-experimentation.

6 Conclusion

This paper departs from the two-arm bandit problem of Keller, Rady and Cripps (2005), and changes the previous safe arm to the no-learning arm. The no-learning arm generates unobserved constant flow payoff and both players have private information about this payoff. The two-sided uncertainty in the no-learning arm increases the experimentation level compared to the benchmark without private information. Players experiment more due to (1) the incentive of hiding private information and encouraging the other player to experiment more; (2) the ignorance towards the true flow payoff of the no-learning arm. Only the second force can lead to over-experimentation. For example, when the true flow payoff is high, players should stop experimentation early. But in my model both players may think that the flow payoff is possibly low, and thus continue to experimenting beyond the efficient threshold. The two-sided uncertainty can increase the experimentation level and even overturn the under-experimentation problem caused by information free-riding in Keller, Rady and Cripps (2005).

Though we have the possibility of over-experimentation, if we consider the ex-ante welfare of two players, two-sided uncertainty is still beneficial. More experimentation in my model has the benefit of mitigating the free-riding problem and the harm of possible over-experimentation. The former benefit can cover 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 the similar idea, possible extensions may lay in other unsafe backup options. For example, people may hire other experts to manage the money as a backup of investing in the new and risky project. 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. 2018. "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.

Isaac, R Mark, and James M Walker. 1988. "Communication and free-riding behavior: The voluntary contribution mechanism." *Economic inquiry*, 26(4): 585–608.

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.

Kiyonari, Toko, and Pat Barclay. 2008. "Cooperation in social dilemmas: free riding may be thwarted by second-order reward rather than by punishment." *Journal of personality and social psychology*, 95(4): 826.

Komai, Mana, Mark Stegeman, and Benjamin E Hermalin. 2007. "Leadership and information." *American Economic Review*, 97(3): 944–947.

A Mathematical 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 a proof for Proposition 1.

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

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).

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

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^j 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 $p_t > p_c(\underline{s} + \overline{s})$.

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

Firstly, for type $\underline{s}$, deviating from $k = 1$ makes q_t^j 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) $p_s < p_t \leqslant \hat{p}_1$

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.

$$2\lambda p(1-p)u'(p) + (r + 2\lambda p)u(p) = (r + 2\lambda)\lambda hp \tag{7}$$

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

$$u(p_s) = q_0(\underline{s} + \overline{s}) + (1 - q_0)2\overline{s} \tag{8}$$

Plug (8) into (7) 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 (7) 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 $p_s < p_t \leqslant \hat{p}_1$.

(2) $\hat{p}_1 < p_t \leqslant p_c(2\overline{s})]$[13].

When both players choose $k = 1$ for $p_t > p_s$ and $k = 0$ for $p_t \leqslant p_s$, let the payoff type $\overline{s}$

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

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 $v(p_t) = q_0 v_l(p_t) + (1-q_0)v_h(p_t)$[14].

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 $p_t \leqslant p_c(2\overline{s})$ and play $k = 1$ for $p_t > p_c(2\overline{s})$[15].

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

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

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 $p_s \leqslant \hat{p}_1 - \epsilon$, 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

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

and thus

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

[14]The on-path payoff at $p_t > p_s$ in my equilibrium satisfies the ODE $2\lambda p(1-p)v'(p) + (r + 2\lambda p)v(p) = (r+2\lambda)\lambda h p$, 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 $p_t \leqslant p_s$.

[15]According to the constructed equilibrium in the one-sided uncertainty problem.

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

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 $v_l(p) - w_l(p) \geqslant \tau > 0$ 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 $p_s \in [p_c(\underline{s}+\overline{s}), \hat{p}_1 - \epsilon]$, $p_t \in (\hat{p}_1, p_c(2\overline{s})]$[16]. So we know that for q_0 larger than a threshold, $\frac{q_0}{1-q_0} \geqslant \frac{w_h(p_t)-v_h(p_t)}{v_l(p_t)-w_l(p_t)}$ 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 $p_s \leqslant \hat{p}_1 - \epsilon$, 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) $p_t > p_c(2\overline{s})$

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

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

A.2 One-sided uncertainty problem

A.2.1 Construction of the equilibrium

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

When $p_t \leqslant p_c(2\overline{s})$, 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 $k = 0$[17] 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 $p_t \leqslant p_c(2\overline{s})$ is the full separating stage. The

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

on-path q_t[18] is q_0 with $p_t > p_c(2\overline{s})$, and 1 or 0 with $p_t \leqslant p_c(2\overline{s})$.

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

A.2.2 Proof of the equilibrium

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

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[19]. 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) $p_t > p_c(2\overline{s})$

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 HJB equation

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

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

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

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

$$S(p) = u(p) - 2Es + \lambda hp \geqslant 0 \tag{9}$$

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 HJB 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 $p_t > p_c(2\overline{s})$[20]

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

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

[20]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. $\hat{p}_2 < p_c(2\overline{s})$.

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 (9) 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 uncertainty problem where type $\underline{s}$ reveals himself has an equilibrium with the pooling stage and the full separating stage, cut by point $p_c(\underline{s} + \overline{s})$.

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.

It easy to see that

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

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

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

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 uncertainty problem. Recall that in the equilibrium of the one-sided uncertainty 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 uncertainty 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

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

Then we conclude

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

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

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 (2018), I prove the solution in a backward induction manner.

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

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[21].

If $k_p = 0$, type $\underline{s}$ can deviate to $k = 1$ for a moment to reveal himself and enters an one-sided uncertainty 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) $p_c(\underline{s} + \overline{s}) < p_t \leqslant p_s$

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.*

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 uncertainty 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 $k < 1$[22]. ■

[21] 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 $\overline{s}$, $p_t \leqslant p_c(\underline{s} + \overline{s})$.

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

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

$$\begin{aligned} u(p) - Es =& k_0(\lambda hp - Es) + \frac{1}{r}\Big\{(1-q(p))e_t[2\overline{s} - u(p)] \\ &+ [2k_0 p\lambda(\lambda h - u(p)) + \frac{du(p)}{dp}\frac{dp}{dt}]\Big\} \end{aligned} \tag{10}$$

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 (10) 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.*

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 $p_t \leqslant p_s < \hat{p}_1 < p_c(2\overline{s})$. ■

discussed above.

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[23]. 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 uncertainty 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 – in any larger p_t than the current pooling interval, type $\overline{s}$ will not partially reveal himself by random 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 $p_c(\underline{s}+\overline{s}) < p_t \leqslant p_s$, we can only have $q_t = q_i(p_t)$.

(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, which again requires random revealing from type $\underline{s}$. But it is not achievable.

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

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