Over-and Under-Experimentation in a Patent Race with Private Learning *

Kaustav Das † Nicolas Klein‡

March 2, 2020

Abstract

This paper analyses a two-player game with two-armed exponential bandits. A player experiences publicly observable arrivals by pulling the safe arm. On the other hand, a player operating a good risky arm experiences publicly observable arrivals at an intensity greater than that in the safe arm. In addition, a player pulling the risky arm can also privately learn about its quality. With direct payoff externalities and private learning, we construct a symmetric Markov equilibrium where, depending on the initial optimism about the quality of the risky arm, we can have either *too much* or *too little* experimentation.

JEL Classification Numbers:C73, D83, O31.

Keywords: Two-Armed Bandit, Private Learning, Competition

*Previously circulated as *Strategic Experimentation with Competition and Private Arrival of Information.* Special note of thanks goes to Kalyan Chatterjee and Sven Rady for their continuous guidance. This paper has benefited from useful discussions with Martin Cripps, Robert Evans, Sanjeev Goyal, Vijay Krishna, Johannes Hörner, Nilanjan Roy, Larry Samuelson, Qingmin Liu, and Hamid Sabourian. Thanks are also due to the participants of Bristol GW4 Game Theory Workshop 2016, Lancaster Game Theory Conference 2017 and the Royal Economic Society Conference 2018, and seminar participants at the University of Cambridge, University of Bonn and the Indian Statistical Institute. The remaining errors are my own. A substantial portion of a very preliminary initial draft was prepared during my visit to Cambridge-INET, Faculty of Economics, University of Cambridge during the spring term of 2014. I thank Hamid Sabourian and Sanjeev Goyal for their warm hospitality and also the University of Exeter Business School for allowing me to spend this time in Cambridge

†University of Leicester; email: daskaustav84@gmail.com; daskaustav84@gmail.com

‡Université de Montréal and CIREQ, Email:kleinnic@yahoo.com

1 Introduction

In many real-life situations, information generated by one agent is payoff relevant to other agents as well. For example, consider two pharmaceutical companies trying to invent a drug for the same disease. One company's outcome of the clinical trials using a novel target can influence the decision-making of another company. Pharma companies operate in a competitive environment and carry out R&D activities to become the first inventor of a new drug, because the company inventing the drug first can apply for a patent and hence will get a disproportionately higher payoff than the later inventors. Such a game of both informational and payoff externalities can be analysed using a two-armed bandit model.

To date, the literature on strategic experimentation with two-armed bandit models has mostly considered cases in which all outcomes and actions are publicly observable. However, in some situations, one of the pharma companies can learn some outcome of its clinical trial process privately. Booth and Zemmel (2004) found that many pharma companies, for clinical trials incline to opt for the novel targets discovered from the human genome project and computational analysis methods. Hence, companies are shifting their discovery portfolio more toward these riskier alternatives, and are abandoning the risk-adjusted systematic project-development processes. One possible reason can be that the science behind these riskier candidates is novel and attractive. In the context of pharmaceutical rese arch, pharma companies often have many pre-clinical trials whose results are private to them. The influence of these private observations on the decision of a company to choose riskier alternatives is worth exploring. This paper provides a stylised model to analyse the impact of private learning in a game of strategic experimentation with two-armed exponential bandits under competition.

The key features of this paper's framework are as follows: (i) *Two players, each of whom has access to an identical two-armed bandit:* The safe arm, when pulled, generates *publicly observable* arrivals at random times, which are exponentially distributed with a given level of intensity. The risky arm can either be good or bad. When a good risky arm is pulled, it generates publicly observable arrivals. The arrival times are exponentially distributed with a level of intensity higher than that in the safe arm. A novel characteristic of the model is that the good risky arm also independently generates arrivals, which are only privately observable to the player who is pulling it.

(ii) *Initially, the type of the risky arm is not known:* Players share a common prior belief about it. (iii) *Payoff only from the first publicly observable arrival:* This feature of the model incorporates competition or payoff externalities. Private arrival does not yield any payoff. The game ends after the first publicly observable arrival. (iv) *All actions are publicly observable:* Based on these actions and the publicly observable arrivals, each player updates its beliefs. As evident, because of private learning, public and private beliefs may diverge. (v) *Switching between arms is almost costless:* If a player switches arms, the decision can be revoked in a finite time duration of arbitrarily small length.

The main findings of the paper are the following:

- *A symmetric Markov equilibrium exists:* If the intensity of private arrival is below a threshold, an equilibrium exists where, conditional on no arrival, each player pulls the risky arm as long as the private belief about the quality of the risky arm is higher than a threshold.

- *Excessive (Inadequate) use of the risky arm for high priors:* Compared to the benchmark (full-information optimal), the duration of the use of the risky arm is higher (lower) in the equilibrium constructed if the common prior is higher (lower) than a threshold.

In a competitive environment, researchers may not always have the incentive to reveal their interim private findings, This is because once it is revealed, nothing stops another researcher from using that interim result. The use of this intertim result might lead to the other researcher becoming the first to make the final innovation.[1] In this paper, we construct an equilibrium where, although players explicitly do not reveal their interim private findings, it is implicitly conveyed through the equilibrium behaviour, because a player receiving a private signal will never use the safe arm. Because actions are perfectly observable, when a player observes his opponent not switching arms when he is supposed to, he infers that the opponent has received a private signal. If a player (say, player 2) infers his opponent (say, player 1) has experienced a private arrival, he will react by immediately switching back to the risky arm. Because

[1]The incidence of private learning before the final success of innovation is also relevant in the academic world. In this context, the proof of Fermat's last theorem by Sir Andrew Wiles is worth mentioning. This theorem was proved over a span of seven years in complete secrecy, and in between, he did not reveal any of the interim findings.

private information is not directly payoff relevant, player 2 can reap the benefit of player 1's private breakthrough. This characteristic brings *free-riding* in an implicit manner. In the constructed equilibrium, players' actions are of the threshold type, and the threshold is common across the players. From the analysis of Keller, Rady and Cripps (2005), we know that because of free-riding, we do not get an equilibrium where players' actions are of the threshold (common) type. In the current model, because the implicit free-riding is due to private arrivals, to ensure the existence of the equilibrium, we need to keep the intensity of the private arrivals within a limit. This intuitively explains the condition for the existence of the equilibrium.

A novel feature of the results obtained in this paper is that in the equilibrium constructed, both *too much* or *too little* experimentation can occur, depending on the initial optimism about the quality of the risky arm. When a player is pulling the risky arm, his opponent always knows that with some probability, the player will experience a private arrival. This makes the movement of the public belief sluggish compared to the benchmark case. Therefore, too much experimentation tends to occur in equilibrium compared to the benchmark. On the other hand, an implicit free-riding effect tends to lower the extent of experimentation in the equilibrium. When the likelihood of the risky arm being good is high, the former effect dominates, and we have too much experimentation. Otherwise, the latter effect dominates, and we have too little experimentation.

Related Literature: This paper contributes to the currently nascent literature on private learning in models of experimentation. Some recent related papers on this topic are Akcigit and Liu (2015), Heidhues, Rady and Strack (2015), Guo and Roesler (2016), Bimpikis and Drakupoulos (2014), Dong (2017), and Thomas (2018).

Akcigit and Liu (2015) analyse a two-armed bandit model with one risky and one safe arm. The risky arm could lead to a dead end. Thus, private information is in the form of *bad news*. Inefficiency arises from the fact that wasteful dead-end replication and early abandonment of the risky project can occur. In the current paper, private information is in the form of *good news* about the risky arm. However, the present paper shows that early abandonment of the risky project can still happen if, to start with, players are *not* too optimistic about the quality of the risky line. Further, in the current paper, we have learning even in the absence of information asymmetry. Heidhues, Rady and Strack (2015) analyse a model of strategic experimentation with private payoffs. They take a two-armed bandit model with a risky arm and a safe arm. Players observe each other's behaviour but not the realised payoffs. They communicate with each other via cheap talk. The free riding problem can be reduced because of private payoffs, and conditions exist under which the cooperative solution can be supported as a perfect Bayesian equilibrium. The present paper differs from their work in the following ways. First, we have private arrivals of information only. Second, players are rivals.

Guo and Roesler (2016) study a collaboration model in continuous time. In both the good and the bad state, success can arrive at a positive rate. In the bad state, players may get a perfectly informative signal. Players have to exert costly effort to stay in the game, and at any time, they have the option to exit and secure a positive payoff from an outside option. Both the probability of success and private learning are directly related to the amount of effort exerted. An increase in the payoff from the outside option increases collaboration among agents. Bimpikis and Drakupoulos (2014)consider a setting in which agents experiment with an opportunity of an unknown value. Information generated by the experimentation of an agent can be credibly communicated to others. They identify an optimal time $T > 0$ such that if agents commit not to share any information up to time T and disclose all available information at time T, the extent of free-riding is reduced.

Dong (2017) studies a model of strategic experimentation with two symmetric players, where all actions and outcomes are public. However, one of the players is initially better informed about the state of nature. In the current paper, we have a competitive environment, and through a private outcome, both players can become privately informed. In Thomas (2018), each player can either learn privately along his exclusive risky arm or can compete for a single safe arm. Thus, unlike the present paper, no direct payoff externalities exist.[2]

This paper has important implications for the literature on patent-race games.

[2] Thomas (2017) analyses a model in which a decision-maker chooses a stopping time for a project, and she receives private information gradually over time about whether the project will succeed. Rosenberg, Salomon and Vieille (2013) also analyses a strategic experimentation game in which they look at the effect of varying the observability of the experimentation outcomes. The current paper differs from theirs in three ways. First, both observable and unobservable kinds of outcomes occur. Second, the environment is competitive. Lastly, in the current paper, switching between arms is not irrevocable. Murto and Välimäki (2011) also consider irrevocable switching with private outcomes of experimentation. In their model, both kinds of all risky arms are perfectly positively correlated, and nature initially chooses the proportion of good risky arms.

Chatterjee and Evans (2004) analyses a model in which firms compete to make a discovery, and it can be made through one and only one of the available two methods. A prori, firms only know the likelihood of each method being the correct one. The cost to search differs across research avenues. Depending on the prior, either *too much* or *too little* exploration of a research avenue occurs. The current paper yields a similar result but due to private learning. Further, we do not have a perfect negative correlation between the research avenues. Besanko and Wu (2013) adopt the framework of Keller, Rady and Cripps (2005) and study how the market structure affects an R&D race. Their paper is about *how much* to invest in R&D. On the other hand, in the current paper, we analyse how to allocate a given amount of resources among available R&D methods. Moscarini and Squintani (2010) analyse private learning in R&D races. In their model, the arrival rate of a firm's invention is its private information and shows that when the social planner is sufficiently impatient, firms' R&D activities are inefficiently low. In the current paper, we show that along with a commonly known arrival rate of invention, if a private signal occurs, we can have both inefficiently low and high R&D activities.

Finally, this paper contributes to the broad literature on strategic bandits. Some of the important papers[3] in this area are Bolton and Harris (1999), Keller, Rady and Cripps (2005), Keller and Rady (2010), Klein and Rady (2011), Klein (2013), Keller and Rady (2015). In most of these papers, under-experimentation exists due to free-riding, and all learning is public. Different variants of this problem have been studied in the literature. Bonatti and Hörner (2011), in an exponential bandit framework, study the problem of moral hazard in teams with private actions. In Rosenberg, Solan, Vieille (2007), a switch to the safe arm is irreversible and the role of the observability of outcomes and the correlation between risky-arm types is analysed. In the current paper, the framework modifies the setup of Keller, Rady and Cripps (2005) by introducing direct payoff externalities and having a purely informational arrival along the risky arm that only the player pulling it observes. We show that in the presence of private learning in a competitive environment, depending on the initial prior, both too much and too little experimentation can exist in the same model.[4] Another novel feature of the current paper is that it captures a setting

[3] The survey by Hörner and Skrzypacz (2016) gives a comprehensive picture of the current literature on this topic.

[4] Das (2018) shows that in a competitive environment with no private learning, we have too much experimentation.

that is competitive and simultaneously has free-riding opportunities; the free-riding opportunities arise from the communication of private signals through equilibrium actions.[5]

The rest of the paper is organised as follows. Section 2 discusses the environment and the full-information optimal solution. Section 3 discusses the non-cooperative game and the nature of inefficiency. Finally, section 4 concludes.

2 The Environment

Two players (1 and 2) face a replica of a two-armed bandit in continuous time. If the safe arm is pulled, publicly observable arrivals occur at the jumping times of a Poisson process with intensity $\pi_0 > 0$. The risky arm can either be good or bad. The quality of the risky arm of both players is the same. When the good risky arm is pulled, publicly observable arrivals occur at the jumping times of a Poisson process with intensity $\pi_2 > \pi_0$. In addition, a player pulling the good risky arm can experience privately observable arrivals according to a Poisson process with intensity $\pi_1 > 0$. Only the first publicly observable arrival (along any of the arms) yields a payoff of 1 unit.

Players start with a common prior p_0, the probability with which the risky arm is good. Players' actions are publicly observable. At each time point, players update their belief (private) using the public history (publicly observable arrivals and actions) and private history. We start our analysis by examining the benchmark case in which all information is public.

2.1 The planner's problem: The full-information optimal solution

In this sub-section, we discuss the optimisation problem of a benevolent social planner who can completely control the actions of the players and can observe all the arrivals they experience. This is intended to be the efficient benchmark of the model described above.

The action of the planner at time point t is defined by k_t ($k_t \in \{0, 1, 2\}$). k_t is the

[5] Wong (2017) also analyse competition and free-riding in the same environment. However, he does not have private learning.

number of players the planner makes to experiment. k_t ($t \geq 0$) is measurable with respect to the information available at time point t.

Let p_t be the prior at time point t. If no arrival occurs over the time interval $\Delta > 0$, it must be the case that none of the players who were experimenting experienced any arrival, because the planner can observe all arrivals. As $\Delta \to 0$, as in Keller, Rady and Cripps (2005), we get

$$dp_t = -k(\pi_1 + \pi_2)p_t(1 - p_t).$$

The longer the planner has players experimenting without any arrival, the more pessimistic they become about the likelihood of the risky arm being good. As soon as the planner observes any arrival at the risky arm, the uncertainty is resolved. If it is an arrival that would have been publicly observable in the non-cooperative game, the game ends. For the other kind of arrival, the planner gets to know for sure that the good risky arm is good, and makes both the players experiment until the first publicly observable arrival.

Let $v(p)$ be the value function of the planner. Then, along with k, it should satisfy the following Bellman equation:

$$rv = \max_{k \in \{0,1,2\}} \{2\pi_0(1-v)]+k\big[\{(\pi_1+\pi_2)p(\frac{\pi_2}{(\pi_1 + \pi_2)}+\frac{\pi_1}{(\pi_1 + \pi_2)}\frac{2\pi_2}{r + 2\pi_2}-v-(1-p)v^{'})\}-\pi_0(1-v)\big]\big\}. \tag{1}$$

The solution to the planner's problem is summarised in the following proposition.

Proposition 1 *A exists a threshold belief* $p^* = \frac{\pi_0}{\pi_2 + \frac{2\pi_1\{\pi_2 - \pi_0\}}{r+2\pi_2}}$ *exists such that if the belief* p *at any point is strictly greater than* p^**, the planner makes both the players experiment, and if the belief is less than or equal to* p^**, the planner makes both the players exploit the safe arm. The planner's value function is*

$$v(p) = \begin{cases} \frac{2\pi_2}{r+2\pi_2}p + [\frac{2\pi_0}{r+2\pi_0} - \frac{2\pi_2}{r+2\pi_2}p^*]u_0(p) & \text{if } p > p^* \\ \frac{2\pi_0}{r+2\pi_0} & \text{if } p \leq p^*, \end{cases} \tag{2}$$

where $u_0(p) = (1-p)(\frac{1-p}{p})^{\frac{r}{2(\pi_1+\pi_2)}}$. *The ODE satisfied by the planner's problem is displayed in Appendix B*

Proof.

The proof follows from a standard verification argument. Please refer to Appendix B

■

From the expression of p^*, we can see that in the absence of any informational arrival ($\pi_1 = 0$), the threshold is identical to the myopic belief.[6] A higher rate of informational arrival increases the incentive to experiment, and hence the belief up to which the planner makes players experiment goes down. This effect also depends on the difference between the public arrival rates across the safe arm and a good risky arm, which reflects the fact that any meaningful payoff is obtained only through public arrival.

The following section describes the non-cooperative game and constructs a particular symmetric equilibrium.

3 The non-cooperative game

In this section, we consider the non-cooperative game between the players. We restrict ourselves to Markovian strategies. In the current context, a Markov strategy for player i ($i = 1, 2$) is defined as a left-continuous function $k_i : [0,1] \to \{0,1\}$, $p_i \mapsto k_i(p_i)$. $p_i \in [0,1]$ is the private belief of player i. Players can observe each others' actions. The informational arrival is privately observable to the player who experiences it. This characteristic of the model implies that, conditional on no arrival, the private belief of player i evolves according to

$$dp_{i,t} = -[\pi_2(k_{i,t} + k_{j,t}) + \pi_1 k_{i,t}]p_{i,t}(1 - p_{i,t})\,dt.$$

If both players pull the risky arm for a particular time interval at a common private belief, the above equation of motion implies that, conditional on no arrival, the private posteriors during this interval will be identical across the players. We exploit this fact to construct a particular equilibrium in the following subsection.

Only the first publicly observable arrival yields a payoff of 1 unit to the player who experiences it. In the current model, inertia is present in the players' decision to switch between arms or change the decision to switch arms. Formally, this means

[6] For $\pi_1 = 0$, we get back the result of the planner's problem with homogeneous players in Das (2018).

that if a player shifts from one arm to another, then if he wants to shift back, he cannot do so before a time interval $\eta > 0$. Also, if a player decides not to switch to another arm, then to reconsider his decision, he needs a time of $\eta > 0$. In the next subsection, we construct a particular symmetric equilibrium in Markovian strategies when the inertia goes to zero ($\eta \to 0$).

3.1 Equilibrium

In this subsection, we construct a particular symmetric Markov perfect equilibrium of the game when the inertia η goes to zero. The features of the symmetric equilibrium we intend to construct are as follows: (i) Players start with a common private belief about the quality of the risky arm. (ii) Conditional on no arrival, each player pulls the risky arm as long as the private belief is greater than a particular threshold. Note that in the non-cooperative game, a player can observe only the public arrivals and his private arrival only. (iii) If a player experiences a private arrival, he keeps pulling the risky arm until a public arrival.

In this conjectured equilibrium, players' strategies are pure and symmetric. Because actions are publicly observable, on the equilibrium path, conditional on no arrival, players' private beliefs will be identical. This approach is similar to the one followed by Bergemann and Hege (2005). We first assume the existence of the conjectured equilibrium and determine the common threshold belief where players switch to the safe arm from the risky arm. The following lemma does this.

Lemma 1 *Suppose the symmetric Markov perfect equilibrium as conjectured above exists. Let $p^{*N}(p)$ be the threshold for switching for the prior p. If the private belief is higher than this threshold, the player pulls the risky arm. Conditional on no arrival, a player switches to the safe arm when the belief hits $p^{*N}(p)$. We have*

$$p^{*N}(p) = \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2 - \pi_0)\frac{r}{r+\pi_0}}$$

as the inertia $\eta \to 0$

Proof. Players start with a common private belief p. Let $p^{*N}(p)$ be the threshold belief where, conditional on no arrival, both players switch to the safe arm. Given the common prior, conditional on not experiencing any arrival, a player can always determine the probability with which the opponent has experienced a private arrival. Let this probability be q_p. At $p^{*N}(p)$, given the other player's strategy, each player is indifferent between staying on the risky arm switching to the safe arm. When both players are on the safe arm, each receives a payoff of $v_s = \frac{\pi_0}{r+2\pi_0}$.[7] Similarly, if it is known with certainty that the risky arm is good and both players experiment, each gets $v_{rg} = \frac{\pi_2}{r+2\pi_2}$. Suppose a player does not experience any arrival and switches to the safe arm as the private belief hits $p^{*N}(p)$. If the opponent experiences a private arrival, he keeps pulling the safe arm. The former player observes this and infers the risky arm is good. He then reverts to the risky arm after the time interval η.

If a player decides to stay on the risky arm for an additional duration of η, the payoff is

$$\theta_r = (1-q_p)\{\pi_2 p\eta + \pi_1 p\eta(1-r\eta)(1-\pi_0\eta)\frac{\pi_2}{r+2\pi_2}$$

$$+(1-r\eta)[1-(\pi_1+\pi_2)p\eta-\pi_0\eta][\pi_0\eta+(1-r\eta)(1-\pi_0\eta-\pi_1 p\eta-\pi_2 p\eta)v_s+\pi_1 p\eta(1-r\eta)(1-\pi_0\eta)\frac{\pi_2}{r+2\pi_2}]\}$$

$$+q_p\frac{\pi_2}{r+2\pi_2}.$$

If the player instead switches to the safe arm, then for the duration η, his payoff will be

$$\theta_s = (1-q_p)\{\pi_0\eta+(1-r\eta)(1-2\pi_0\eta)v_s\}+q_p\{\pi_0\eta+(1-r\eta)(1-(\pi_0+\pi_2)\eta)\frac{\pi_2}{r+2\pi_2}\}.$$

At $p = p^{*N}(p)$, we have $\theta_r = \theta_s$. We are constructing the equilibrium for $\eta \to 0$. Because compared to v_s, we can ignore terms involving η in a multiplicative manner when $\eta \to 0$, we have $\pi_0\eta + (1-r\eta)(1-\pi_0\eta-\pi_1 p\eta-\pi_2 p\eta)v_s + \pi_1 p\eta(1-r\eta)(1-\pi_0\eta)\frac{\pi_2}{r+2\pi_2} \approx v_s$. Similarly, we have $\pi_0\eta+(1-r\eta)(1-(\pi_0+\pi_2)\eta)\frac{\pi_2}{r+2\pi_2} \approx \frac{\pi_2}{r+2\pi_2}$. After ignoring the terms of the order η^2, we get

$$(1-q_p)\{\pi_2 p^{*N}\eta+\pi_1 p^{*N}\eta\frac{\pi_2}{r+2\pi_2}-(\pi_1+\pi_2)p^*\eta v_s-\pi_0\eta v_s+v_s-r\eta v_s\}+q_p\frac{\pi_2}{r+2\pi_2}$$

$$=(1-q_p)\{\pi_0\eta-2\pi_0\eta v_s+v_s-r\eta v_s\}+q_p\{\frac{\pi_2}{r+2\pi_2}\}$$

[7] In this case, each player's payoff is calculated as $v_s = \int_0^\infty e^{-rs}\frac{\pi_0}{2\pi_0}\frac{d(1-e^{-2\pi_0 s})}{ds}\,ds = \frac{\pi_0}{r+2\pi_0}$.

$$\Rightarrow p^{*N}\{\pi_2 + \pi_1 \frac{\pi_2}{r+2\pi_2} - (\pi_1+\pi_2)v_s\} = \pi_0 - \pi_0 v_s$$

Substituting the value of v_s, we get

$$p^{*N} = \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2-\pi_0)\frac{r}{r+\pi_0}}.$$

This concludes the proof of the lemma.

■

The above lemma shows that if the conjectured equilibrium exists, the common threshold belief is given by p^{*N}. We now formally state the equilibrium strategies along with a system of beliefs (both on and off the equilibrium path) and then describe the Bellman equations of the players. Later, we establish that the proposed equilibrium strategies and the system of beliefs together indeed constitute an equilibrium.

Player i's equilibrium strategy is given by

$$k_i = 1 \text{ for } p > p^{*N} \text{ and } k_i = 0 \text{ for } p \leq p^{*N}.$$

We will now describe how beliefs evolve, both on and off the equilibrium path.

Let A be the set of possible action profiles of the game. We have

$$A = \{(R,R);(S,R);(R,S)\}.$$

In any action profile, the i^{th} element denotes the action of player i ($i=1,2$). Note that we leave aside the action profile (S,S), because under this, no learning occurs. For player i, we define z_i such that $z_i \in \{0,1\}$. $z_i = 1(0)$ denotes that player i has (not) experienced a private arrival. Consider player i ($i=1,2$). Suppose at an instant, his private belief is p. Based on the public observations and private observation, player i updates his beliefs, and this is given by the function $p^u(p,a,z_i)$. As soon as a player experiences a private arrival, all uncertainty is resolved to him. This implies that as z_i becomes equal to 1, the belief of player i jumps to 1 and stays there. Formally, this is expressed as

$$p^u(p,a,z_i=1) = 1$$

$\forall a \in A$ and $\forall p$.

First, consider the range of beliefs $p \in (p^{*N}, 1)$. Because players are homogeneous and we are considering a symmetric equilibrium, without loss of generality, consider player 1. On the equilibrium path, the action profile is $a = (R, R)$. Conditional on no private and public arrival, $p^u(p, a, z_1)$ satisfies

$$dp = -p(1-p)(2\pi_2 + \pi_1)\, dt.$$

The above equation of motion implies a player *on the equilibrium path* while updating beliefs takes into account the public history and his private observations. For both the off-the-equilibrium path action profiles (S, R) and (R, S), conditional on no public or private breakthrough, private beliefs of both the players evolve according to

$$dp = -p(1-p)(\pi_1 + \pi_2)\, dt.$$

Thus, if a player deviates and pulls the safe arm for a particular range of beliefs, he updates his beliefs according to how his opponent does, conditional on no public or private arrivals. This updating ensures that if the players start with a common prior that is greater than p^{*N}, and if one of the players deviates for some time interval, the private beliefs remain the same across the players.

Next, consider the range of beliefs $p \leq p^{*N}$ and, as before, consider player 1. For $a = (S, R)$ and $a = (R, R)$, we have

$$p^u(p, a, z_1 = 0, d_1 = 0) = 1.$$

The above equation implies if a player had not experienced any private arrival until the time point when his private belief became equal to p^{*N}, and he observes his opponent pulling the risky arm, he interprets this as the opponent receiving private information, and hence the belief jumps to 1.

Suppose players 1's belief jumps to 1 by observing the opponent not shifting to the safe arm when he is supposed to. If at any point after that time, player 1 observes his opponent pulling the safe arm, he thinks his opponent must have committed a mistake before, and the updated belief of player 1 becomes 0. We formally express this updating as

$$p^u(1, a, z_1 = 0, d_1) = 0$$

for $a = (R, S)$ and $a = (S, S)$.

Finally, for any $p \leq p^{*N}$, for the action profile $a = (R, S)$, beliefs of player 1 evolve according to

$$dp = -(\pi_1 + \pi_2)p(1-p)\, dt.$$

This completes the description of the system of beliefs. We now compute the payoffs of the players in the conjectured symmetric equilibrium. Both players start with a common prior. On the equilibrium path, the actions of the players are symmetric. For $p > p^{*N}$, given k_j $(j \neq i)$, player i's value should satisfy

$$v_i = \max_{k_i \in \{1,0\}} \big\{(1-k_i)\pi_0\, dt + k_i p\big[\pi_2\, dt + \pi_1 \frac{\pi_2}{r+2\pi_2}\, dt\big] +$$

$$(1-r\, dt)\big[1-(2-k_1-k_2)\pi_0\, dt - [k_1(\pi_1+\pi_2)p\, dt + k_2\pi_2 p\, dt]\big]\big[v_i - v_i^{'}p(1-p)[k_i(\pi_1+\pi_2)+k_j\pi_2]\, dt\big]$$

$$+k_j p\, dt \pi_1 \frac{\pi_2}{r+2\pi_2}\big\}.$$

To player i, k_j is given. By ignoring the terms of the order $0(\,dt)$ and rearranging the above, we can infer that v_i along with k_i satisfies the following Bellman equation:

$$rv_i = \max_{k_i \in \{0,1\}} \big\{(\big[\pi_0(1-v_i)\big] + k_i\big[p(\pi_1+\pi_2)\{(\frac{\pi_2}{\pi_1+\pi_2} + \frac{\pi_1}{\pi_1+\pi_2}\frac{\pi_2}{r+2\pi_2}) - v_i - v_i^{'}(1-p)\} - \pi_0(1-v_i)\big]\big\}$$

$$-(1-k_j)\pi_0 v_i - k_j\big[p\pi_2 v_i + \pi_2 p(1-p)v_i^{'}\big] + k_j p \pi_1 \frac{\pi_2}{r+2\pi_2}. \tag{3}$$

For $p > p^{*N}$, both players experiment. Putting $k_i = k_j = 1$ in (3), we get

$$v_i^{'} + \frac{v_i[r+(\pi_1+2\pi_2)p]}{p(1-p)(\pi_1+2\pi_2)} = \frac{\pi_2 p[r+2\pi_1+2\pi_2]}{(r+2\pi_2)p(1-p)(\pi_1+2\pi_2)}. \tag{4}$$

Solving this O.D.E., we obtain

$$v_i = \frac{\pi_2}{r+2\pi_2}[\frac{r+2\pi_1+2\pi_2}{r+\pi_1+2\pi_2}]p + C^n_{rr}(1-p)[\Lambda(p)]^{\frac{r}{\pi_1+2\pi_2}}$$

$$\Rightarrow v_i = \frac{\pi_2}{r+2\pi_2}[1+\frac{\pi_1}{r+\pi_1+2\pi_2}]p + C^n_{rr}(1-p)[\Lambda(p)]^{\frac{r}{\pi_1+2\pi_2}}. \tag{5}$$

where C^n_{rr} is an integration constant and $\Lambda(.)$ is as defined before. The integration constant is given by

$$C_{rr}^{n} = \frac{\frac{\pi_0}{r+2\pi_0} - \frac{\pi_2}{r+2\pi_2}[\frac{r+2\pi_1+2\pi_2}{r+\pi_1+2\pi_2}]p^{*N}}{(1-p^{*N})[\Lambda(p)]^{\frac{r}{\pi_1+2\pi_2}}}.$$

For $p \leq p^{*N}$, both players get a payoff of $v_s = \frac{\pi_0}{r+2\pi_0}$. We now establish that the conjectured equilibrium is indeed an equilibrium through the following proposition.

Proposition 2 *A threshold $\pi_1^* > 0$ exists such that if $\pi_1 < \pi_1^*$, the conjectured equilibrium is indeed an equilibrium.*

Proof.

We can establish a $\pi_1^{*1} > 0$ exists such that for all $\pi_1 < \pi_1^{*1}$, the payoff for each player is increasing and convex for $p > p^{*N}$. Please refer to Appendix (C) for a detailed proof.

Next, for each $p > p^{*N}$, from (3), we can infer that no player has any incentive to unilaterally deviate if $p(\pi_1+\pi_2)\{(\frac{\pi_2}{\pi_1+\pi_2} + \frac{\pi_1}{\pi_1+\pi_2}\frac{\pi_2}{r+2\pi_2}) - v_i - v_i^{'}(1-p)\} - \pi_0(1-v_i) \geq 0$. Because for $p > p^{*N}$ both $k_i = k_j = 1$, the implication is that

$$[r + \pi_0 + \frac{\pi_0\pi_2}{(\pi_1+\pi_2)}]v_i \geq \pi_0 + \frac{\pi_0\pi_2}{(\pi_1+\pi_2)} - \frac{\pi_2^2(r+\pi_1+2\pi_2)p}{(\pi_1+\pi_2)(r+2\pi_2)}$$

$$\Rightarrow v_i \geq \frac{\pi_0 + \frac{\pi_0\pi_2}{(\pi_1+\pi_2)} - \frac{\pi_2^2(r+\pi_1+2\pi_2)p}{(\pi_1+\pi_2)(r+2\pi_2)}}{[r+\pi_0+\frac{\pi_0\pi_2}{(\pi_1+\pi_2)}]}. \tag{6}$$

The R.H.S. of (6) evaluated at $p = \frac{\pi_0}{\pi_2}$ is equal to $\frac{\pi_0 - \frac{\pi_1\pi_2\pi_0}{(\pi_1+\pi_2)(r+2\pi_2)}}{[r+\pi_0+\frac{\pi_0\pi_2}{(\pi_1+\pi_2)}]}$.

Because $\pi_2 > \pi_0$, we have

$$\frac{\pi_0}{r+2\pi_0} - [\frac{\pi_0 - \frac{\pi_1\pi_2\pi_0}{(\pi_1+\pi_2)(r+2\pi_2)}}{[r+\pi_0+\frac{\pi_0\pi_2}{(\pi_1+\pi_2)}]}] = \frac{\frac{r\pi_1\pi_0}{(\pi_1+\pi_2)(r+2\pi_2)}(\pi_2-\pi_0)}{(r+2\pi_0)[r+\pi_0+\frac{\pi_0\pi_2}{(\pi_1+\pi_2)}]} > 0. \tag{7}$$

We know $v_i(p^{*N}) = \frac{\pi_0}{r+2\pi_0}$. At $p = \frac{\pi_0}{\pi_2}$, for any $\pi_1 > 0$, we have

$$\frac{\pi_0}{r+2\pi_0} > \frac{\pi_0 + \frac{\pi_0\pi_2}{(\pi_1+\pi_2)} - \frac{\pi_2^2(r+\pi_1+2\pi_2)p}{(\pi_1+\pi_2)(r+2\pi_2)}}{[r+\pi_0+\frac{\pi_0\pi_2}{(\pi_1+\pi_2)}]}.$$

As $\pi_1 \to 0$, $p^{*N} \uparrow \frac{\pi_0}{\pi_2}$. Therefore, a $\pi_1^{*2} > 0$, such that for all $\pi_1 < \pi_1^{*2}$, we have

$$v_i(p^{*N}) = \frac{\pi_0}{r+2\pi_0} \geq \frac{\pi_0 + \frac{\pi_0\pi_2}{(\pi_1+\pi_2)} - \frac{\pi_2^2(r+\pi_1+2\pi_2)p^{*N}}{(\pi_1+\pi_2)(r+2\pi_2)}}{[r+\pi_0+\frac{\pi_0\pi_2}{(\pi_1+\pi_2)}]}. \tag{8}$$

Define $\pi_1^* = \min\{\pi_1^{*1}, \pi_1^{*2}\}$. For any $\pi_1 < \pi_1^*$, we have v_i to be strictly convex and increasing for all $p > p^{*N}$ and condition (8) also holds. Because the R.H.S. of (6) is decreasing in p, we can infer that for all $p > p^{*N}$, (6) holds. This shows that no player has any incentive to deviate for any $p > p^{*N}$. Consider $p \leq p^{*N}$. We need to show that, conditional on no arrival, if the opponent (player j) chooses the safe arm, doing so constitutes a best response for player i. On the equilibrium path, for this range of beliefs, $v_i = \frac{\pi_0}{r+2\pi_0}$. Player i is playing a best response as long as $p(\pi_1 + \pi_2)\{(\frac{\pi_2}{\pi_1+\pi_2} + \frac{\pi_1}{\pi_1+\pi_2}\frac{\pi_2}{r+2\pi_2}) - v_i - v_i'(1-p)\} - \pi_0(1-v_i) \leq 0$. From direct computation, it can be seen that this is satisfied for $p \leq p^{*N}$. Finally, given the way p^{*N} is calculated, it follows that for $p \leq p^{*N}$, player i will have no incentive to deceive the opponent by experimenting a little longer.

This concludes the proof that the conjectured equilibrium is indeed an equilibrium if π_1 is lower than a threshold. ■

The above result can be interpreted intuitively. The equilibrium we construct involves both players playing a threshold-type strategy. In our model, we have private learning along the risky arm, which reveals the risky arm is good. In the equilibrium constructed, on the equilibrium path, any private signal received by a player is communicated to the other player through his equilibrium action. Therefore, a player can reap the benefit of an arrival his competitor experiences. This fact implicitly brings in an aspect of free-riding. Thus, although our setting is competitive, it contains an aspect of implicit free-riding. From the existing literature, we know that in a model of strategic experimentation, free-riding is a hindrance to the existence of an equilibrium in which both players use threshold-type strategies. In the current model, the higher the value of π_1, the higher the implicit free-riding effect. This explains why we require the intensity of the arrival of private information to be less than a threshold for the constructed equilibrium to exist.

3.2 Inefficiency in Equilibrium

In this subsection, we discuss the possible distortions that might arise in the equilibrium constructed compared to the benchmark case. We begin this subsection by observing that $p^* < p^{*N}$. However, because of private learning, we cannot infer whether we have too much or too little experimentation in the non-cooperative equilibrium solely by comparing the threshold probabilities of switching (p^* in the planner's case and p^{*N} in the non-cooperative case). The reason is that in the non-cooperative equilibrium, the beliefs are private (although the same across individuals), and in the benchmark case, it is public. In the non-cooperative equilibrium, the informational arrival along the good risky arm is only privately observable, and hence if the prior is greater than p^{*N}, the same action profile will give rise to a different system of beliefs. In the current paper, we determine the nature of inefficiency in the following manner.

For each prior, we first determine the duration of experimentation, conditional on no arrival for both the benchmark case and the non-cooperative equilibrium. Then, we say too much (too little) experimentation occurs in the non-cooperative equilibrium if, starting from a prior, conditional on no arrival, the duration of experimentation is higher (lower) in the non-cooperative equilibrium.

The following proposition describes the nature of inefficiency in the non-cooperative equilibrium.

Proposition 3 *The non-cooperative equilibrium involves inefficiency. A $p_0^* \in (p^{*N}, 1)$ exists such that if the prior $p_0 > p_0^*$, then, conditional on no arrival, we have too much experimentation, and for $p_0 < p_0^*$, we have too little experimentation. By too much (too-little) experimentation, we mean that starting from a prior, the duration for which players operate along the risky arm is higher (lower) than the duration a planner would have liked.*

Proof. Let $t^n_{p_0}$ be the duration of experimentation by the firms in the non-cooperative equilibrium described above when they start from the prior p_0. From the description of the equilibrium, we know that if the firms start out from the prior p_0, then they will keep experimenting until the posterior reaches p^{*N}. Thus, we have

$$t^n_{p_0} = -\frac{1}{(\pi_1 + 2\pi_2)} \int_{p_0}^{p^{*N}} [\frac{1}{p_t} + \frac{1}{(1-p_t)}]\, dp_t.$$

The dynamics of the posterior in equilibrium is

$$dp_t = -(\pi_1 + 2\pi_2) p_t (1-p_t)\, dt \Rightarrow\, dt = -\frac{1}{(\pi_1 + 2\pi_2)} \frac{1}{p_t(1-p_t)}\, dp_t,$$

which implies we have

$$t^n_{p_0} = \frac{1}{(\pi_1 + 2\pi_2)} [\log[\Lambda(p^{*N})] - \log[\Lambda(p_0)]]. \tag{9}$$

Let $t^p_{p_0}$ be the duration of experimentation a planner would have wanted if the firms started out from the prior p_0. Then, from the equation of motion of p_t in the planner's problem, we have

$$dp_t = -2(\pi_1 + \pi_2)p_t(1 - p_t)\, dt \Rightarrow dt = -\frac{1}{2(\pi_1 + \pi_2)}\frac{1}{p_t(1 - p_t)}\, dt,$$

which gives us

$$t^p_{p_0} = \frac{1}{(2\pi_1 + 2\pi_2)}[\log[\Lambda(p^*)] - \log[\Lambda(p_0)]]. \tag{10}$$

From 9 and 10, we can infer that excessive experimentation occurs if $t^n_{p_0} > t^p_{p_0}$, which happens when

$$\frac{1}{(\pi_1 + 2\pi_2)}[\log[\Lambda(p^{*N})] - \log[\Lambda(p_0)]] > \frac{1}{(2\pi_1 + 2\pi_2)}[\log[\Lambda(p^*)] - \log[\Lambda(p_0)]]$$

$$\Rightarrow \pi_1 \log[\Lambda(p_0)] < 2(\pi_1 + \pi_2)\log[\Lambda(p^{*N})] - (\pi_1 + 2\pi_2)\log[\Lambda(p^*)].$$

Let $\pi_1 \log[\Lambda(p_0)] \equiv \tau(p)$. Because logarithm is a monotonically increasing function and $\Lambda(p)$ is monotonically decreasing in p, $\tau(p)$ is monotonically decreasing in p.

First, observe that at $\tau(1) = -\infty$.

The R.H.S. can be written as

$$\pi_1 \log[\Lambda(p^{*N})] - (\pi_1 + 2\pi_2)[\log[\Lambda(p^*)] - \log[\Lambda(p^{*N})]]$$

. Because $[\log[\Lambda(p^*)] - \log[\Lambda(p^{*N})]] > 0$, we have

$$\text{R.H.S. } < \pi_1 \log[\Lambda(p^{*N})] = \tau(p^{*N}).$$

Because $p^{*N} \in (0, 1)$ and $\log[\Lambda(p^*)]$ is finite, we have the R.H.S. satisfying

$$2(\pi_1{+}\pi_2)\log[\Lambda(p^{*N})]-(\pi_1{+}2\pi_2)\log[\Lambda(p^*)] > 2(\pi_1{+}\pi_2)\log[\Lambda(1)]-(\pi_1{+}2\pi_2)\log[\Lambda(p^*)] = -\infty.$$

Thus, we have

$$\tau(1) < \text{ R.H.S. and } \tau(p^{*N}) > \text{ R.H.S.}$$

Hence, $\exists$ a $p^*_0 \in (p^{*N}, 1)$ such that for $p_0 > p^*_0$, $\tau(p_0) < R.H.S.$ and for $p_0 < p^*_0$, $\tau(p_0) > R.H.S.$ Hence, if the prior exceeds p^*_0, too much experimentation occurs, and if it is below the threshold, too little experimentation occurs in the non-cooperative equilibrium.

This concludes the proof of this proposition ■

In the non-cooperative equilibrium, distortion arises from two sources. First is what we call the *implicit* free-riding effect. This effect comes from the fact that if a player experiences a private arrival of information, the competing player also reaps the benefit. This scenario is possible here because we construct the equilibrium when the inertia goes to zero. In fact, if information arrival to firms were public, the non-cooperative equilibrium would always involve free-riding. This follows directly from (Keller, Rady and Cripps (2005)). Thus, this implicit free-riding effect tends to reduce the duration of experimentation.

The other kind of distortion arises from the fact that information arrival is private, and while the players are experimenting, the probability that the opponent has experienced a private arrival is directly proportional to the belief that the risky arm is good. Conditional on no observation, this makes the movement of the belief sluggish, resulting in an increase in the duration of experimentation. The effect of distortion from the second (first) source dominates if the common prior to start with is higher(lower). This intuitively explains the result of the above proposition.

4 Conclusion

This paper analysed a tractable model to explore the situation in which private arrival of information as well as public arrival of final invention can occur. We show a non-cooperative equilibrium can exist where, depending on the prior, both too much and too little experimentation can take place. The equilibrium is derived under the assumption that the inertia of players' action goes to zero. How the results change if a player, after switching to the safe arm, is unable to revert back to the risky arm immediately (fixed positive inertia) would be interesting to see. In addition, once we introduce the payoff from revealing informational arrival, situations might exist in which a player would have the incentive to reveal a private observation. I plan to address these issues in my future research.

References

Akcigit, U., Liu, Q., 2015 "The Role of Information in Competitive Experimentation. ", *Journal of the European Economic Association* 14, 828 − 870 .

Bergemann, D., Hege, U., 2005 "The Financing of Innovation: Learning and Stopping ", *The RAND Journal of Economics* 36, 719 − 752

Besanko, D., Wu, J., 2013: "The impact of market structure and learning on the tradeoff between R & D competition and cooperation ", *The Journal of Industrial Economics LXI*, 166 − 201

Bimpikis, K., Drakupoulos, K., 2014 "Disclosing Information in Strategic Experimentation ", *Working paper*

Bolton, P., Harris, C., 1999 "Strategic Experimentation ", *Econometrica* 67, 349−374.

Bonatti, A., Hörner, J., 2011: "Collaborating ", *American Economic Review* 101, 632 − 663

Booth, B., Zemmel, R., 2004: "Prospoects for Productivity ", *Nature Reviews Drug Discovery* 3

Chatterjee, K., Evans, R., 2004 "Rivals' Search for Buried Treasure: Competition and Duplication in R&D. ", *Rand Journal of Economics* 35, 160 − 183.

Das, K., 2018 "Competition, Learning and Duplicative Search in a Patent Race Game ", *mimeo, University of Exeter*

Dong, M., 2017 "Strategic Experimentation with Asymmetric Information ", *mimeo, Bocconi University*

Fershtman, C., Rubinstein, A., 1997 "A Simple Model of Equilibrium in Search Procedures. ",*Journal of Economic Theory* 72, 432 − 441.

Guo, Y., Roesler, A., 2016 "Private Learning and Exit Decisions in Collaboration ", *Working paper*

Heidhues, P., Rady, S., Strack, P., 2015 "Strategic Experimentation with Private Payoffs ", *Journal of Economic Theory* 159, 531 − 551

Hörner, J., Skrzypacz, A., 2016: "Learning, Experimentation and Information Design ", *Working paper, Stanford University.*

Keller, G., Rady, S., Cripps, M., 2005 "Strategic Experimentation with Exponential Bandits ", *Econometrica* 73, $39 - 68$.

Keller, G., Rady, S., 2010 "Strategic Experimentation with Poisson Bandits ", *Theoretical Economics* 5, $275 - 311$.

Keller, G., Rady, S., 2015 "Breakdowns ", *Theoretical Economics* 10, $175 - 202$

Klein, N., 2013: "Strategic Learning in Teams ", *forthcoming in Games and Economic Behavior*

Klein, N., Rady, S., 2011 "Negatively Correlated Bandits ", *The Review of Economic Studies* 78 $693 - 792$.

Moscarini, G., Squintani, F., 2010 "Competitive Experimentation with Private Information: The Survivor's Curse ", *Journal of Economic Theory* 145, $639 - 660$.

Murto, P., Välimäki, J., 2011: "Learning and Information Aggregation in an Exit Game ", *Review of Economic Studies* 78, $1426 - 1461$.

Presman, E.L., 1990 "Poisson Version of the Two-Armed Bandit Problem with Discounting ", *Theory of Probability and its Applications*

Rosenberg, D., Salomon, A., Vieille, N., 2013 "On games of strategic experimentation ", *Games and economic Behavior* $8231 - 51$

Rosenberg, D., Solan, E., Vielle, N., 2007: "Social Learning in One-arm Bandit Problems ", *Econometrica* 75, $1511 - 1611$.

Stokey,N.L., 2009 "The Economics of Inaction ", *Princeton University Press.*

Thomas, C., 2017 "Career Concerns and Policy Persistence: A Dynamic Signalling Model ", *mimeo, University of Texas Austin*

Thomas, C., 2018 "Stopping with Congestion and Private Payoffs ", *mimeo, University of Texas Austin*

Wong, T., 2017 "Free Riding and Duplication in R&D ", *mimeo, Aalto University*

Appendix

A Ordinary Differential Equations

A.1 ODE for planner's problem

When both firms are made to experiment, the planner's value function satisfies the following ODE:

$$2(\pi_1+\pi_2)p(1-p)v^{'}+[r+2(\pi_1+\pi_2)p]v = 2(\pi_1+\pi_2)p[\frac{\pi_2}{\pi_1+\pi_2}+\frac{\pi_1}{\pi_1+\pi_2}\frac{2\pi_2}{r+2\pi_2}]. \quad (11)$$

This ODE is solved by

$$v(p) = \frac{2\pi_2}{r+2\pi_2}p + Cu_0(p). \quad (12)$$

B Proof of proposition 1

The planner's payoff satisfies the value-matching condition at $p = p^*$. From the relevant ODE (see equation 11 in Appendix A), we can conclude that $v^{'}(p^*) = 0$. Because $[\frac{2\pi_0}{r+2\pi_0} - \frac{2\pi_2}{r+2\pi_2}p^*] > 0$, from 12, we can infer that for all $p > p^*$, v is strictly increasing. From the conjectured solution and the ODE 11, it follows that for $p > p^*$, $[(\pi_1+\pi_2)p\{\frac{\pi_2}{\pi_1+\pi_2}+\frac{\pi_1}{\pi_1+\pi_2}\frac{2\pi_2}{r+2\pi_2} - v - (1-p)v^{'}\}] = \frac{rv}{2}$. At $p = p^*$, $\frac{rv}{2} = \pi_0(1-v)$. Because v is strictly increasing for $p > p^*$, we have $[(\pi_1+\pi_2)p\{\frac{\pi_2}{\pi_1+\pi_2}+\frac{\pi_1}{\pi_1+\pi_2}\frac{2\pi_2}{r+2\pi_2} - v - (1-p)v^{'}\}] > \pi_0(1-v)$. On the other hand, for $p \leq p^*$, we have $v^{'} = 0$, and thus $[(\pi_1+\pi_2)p\{\frac{\pi_2}{\pi_1+\pi_2}+\frac{\pi_1}{\pi_1+\pi_2}\frac{2\pi_2}{r+2\pi_2} - v - (1-p)v^{'}\}] \leq \pi_0(1-v)$. Thus, the proposed policy satisfies the Bellman equation 1.

C

First, we show convexity. From (5), we know v_i is convex if $C^n_{rr} > 0$. C^n_{rr} is obtained from the value-matching condition at p^{*N}, which implies

$$C^n_{rr} = \frac{\frac{\pi_0}{r+2\pi_0} - \frac{\pi_2}{r+2\pi_2}[1+\frac{\pi_1}{r+\pi_1+2\pi_2}]p^{*N}}{(1-p^{*N})[\Lambda(p^{*N})^{\frac{r}{\pi_1+2\pi_2}}}.$$

As $\pi_1 \to 0$, $p^{*N} \to \frac{\pi_0}{\pi_2}$ and $\frac{\pi_2}{r+2\pi_2}[1+\frac{\pi_1}{r+\pi_1+2\pi_2}] \to \frac{\pi_2}{r+2\pi_2}$. Hence, $C_{rr}^n \to \frac{\frac{\pi_0}{r+2\pi_0} - \frac{\pi_2}{r+2\pi_2}p^{*N}}{(1-p^{*N})[\Lambda(p^{*N})^{\frac{r}{\pi_1+2\pi_2}}} > 0$. From this, we can infer a $\pi_1^{*1} > 0$ exists such that for all $\pi_1 < \pi_1^{*1}$, $C_{rr}^n > 0$.

Next, we show $v_i^{'}(p^{*N}) > 0$ as long as $p^{*N} > \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}[2\pi_2 - \pi_0] + \frac{2\pi_0\pi_1\pi_2}{(r+2\pi_2)r}}$.

Suppose both players are experimenting when $p > \bar{p}$. Hence, v_1 will be given by (5), and we have

$$v_1^{'} = \frac{\pi_2(r+2\pi_1+2\pi_2)}{(r+2\pi_2)(r+\pi_1+2\pi_2)} - C_{rr}^n[\Lambda(p)]^{\frac{r}{\pi_1+2\pi_2}}[1 + \frac{r}{\pi_1+2\pi_2}\frac{1}{p}].$$

Substituting the value of C_{rr}^n, we have

$$v_1^{'} = \frac{\pi_2}{r+2\pi_2}[\frac{r+2\pi_1+2\pi_2}{r+\pi_1+2\pi_2}] - [\frac{\frac{\pi_0}{r+2\pi_0} - \frac{\pi_2}{r+2\pi_2}[\frac{r+2\pi_1+2\pi_2}{r+\pi_1+2\pi_2}]\bar{p}}{(1-\bar{p})}][1 + \frac{r}{\pi_1+2\pi_2}\frac{1}{\bar{p}}]$$

$$= \frac{\frac{\pi_2}{r+2\pi_2}[\frac{r+2\pi_1+2\pi_2}{r+\pi_1+2\pi_2}](1-\bar{p}) - [\frac{\pi_0}{r+2\pi_0} - \frac{\pi_2}{r+2\pi_2}[\frac{r+2\pi_1+2\pi_2}{r+\pi_1+2\pi_2}]\bar{p}][1 + \frac{r}{\pi_1+2\pi_2}\frac{1}{\bar{p}}]}{(1-\bar{p})}.$$

The numerator of the above term is

$$\frac{\pi_2(r+2\pi_1+2\pi_2)}{(r+2\pi_2)(r+\pi_1+2\pi_2)}(1-\bar{p}) - \frac{\pi_0}{r+2\pi_0} + \frac{\pi_2(r+2\pi_1+2\pi_2)}{(r+2\pi_2)(r+\pi_1+2\pi_2)}\bar{p}$$

$$-\frac{\pi_0 r}{(r+2\pi_0)(\pi_1+2\pi_2)}\frac{1}{\bar{p}} + \frac{\pi_2(r+2\pi_1+2\pi_2)}{(r+2\pi_2)(r+\pi_1+2\pi_2)}\frac{r}{(\pi_1+2\pi_2)}$$

$$= \frac{\pi_2(r+2\pi_1+2\pi_2)}{(r+2\pi_2)(\pi_1+2\pi_2)} - \frac{\pi_0}{(r+2\pi_0)}[\frac{r+(\pi_1+2\pi_2)\bar{p}}{(\pi_1+2\pi_2)\bar{p}}].$$

$v_1^{'}(\bar{p})$ is positive if

$$\frac{\pi_2(r+2\pi_1+2\pi_2)}{(r+2\pi_2)(\pi_1+2\pi_2)} - \frac{\pi_0}{(r+2\pi_0)}[\frac{r+(\pi_1+2\pi_2)\bar{p}}{(\pi_1+2\pi_2)\bar{p}}] > 0$$

$$\Rightarrow \bar{p}[\frac{\pi_2(r+2\pi_1+2\pi_2)}{(r+2\pi_2)} - \frac{\pi_0}{r+2\pi_0}(\pi_1+2\pi_2)] > \frac{r\pi_0}{(r+2\pi_0)}$$

$$\Rightarrow \bar{p}[\frac{\pi_2(r+2\pi_1+2\pi_2)(r+2\pi_0) - \pi_0(\pi_1+2\pi_2)(r+2\pi_2)}{(r+2\pi_2)(r+2\pi_0)}] > \frac{r\pi_0}{(r+2\pi_0)}$$

$$\Rightarrow \bar{p}[\frac{r\pi_2(r+2\pi_2) + r\pi_1(2\pi_2-\pi_0) + 2\pi_0\pi_1\pi_2}{(r+2\pi_2)}] > r\pi_0$$

$$\Rightarrow \bar{p} > \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}[2\pi_2 - \pi_0] + \frac{2\pi_0\pi_1\pi_2}{(r+2\pi_2)r}} \equiv p^{'}.$$

Because $p^{*N} > p^{'}$, we have $v_i^{'}$ to be strictly positive for all $p > p^{*N}$. This concludes the proof.