Strategic Experimentation with Competition and Private Arrival of Information [*]

Kaustav Das [†]

September 19, 2015

Abstract

This paper considers a two-armed bandit problem with one safe arm and one risky arm. The risky arm if good, can potentially experience two kinds of arrivals. One is publicly observable and the other is private to the agent who experiences it. The safe arm experiences publicly observable arrivals according to a given intensity. Private arrivals yield no payoff. Only the first publicly observed arrival(in any of the arms) yields a payoff of 1 unit. Players start with a common prior about the quality of the risky arm. It has been shown that in a particular kind symmetric equilibrium, conditional on no arrival players tend to experiment too much along the risky arm if they start with too high a prior and experiment too less if they start with a low prior.

JEL Classification Numbers:C73, D83, O31.

Keywords: Two-armed Bandit, R&D competition, Duplication, Learning

[*] An earlier and preliminary version of this work was subsumed in the draft entitled *Competition, Duplication and Learning in R&D*. My greatest intellectual debt is to my supervisor Kalyan Chatterjee. Thanks are due to Ufuk Akcigit, Pathikrit Basu, Robert Evans, Sanjeev Goyal, Vijay Krishna, In-Uck Park, Sven Rady, Qingmin Liu, Hamid Sabourian, Caroline Thomas, Sonje Richie and Venky Venkateswaran for the helpful discussions with them. The remaining errors are my own. A substantial portion of the draft was prepared and revisions were made during my visit to Cambridge-INET, Faculty of Economics, University of Cambridge. 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

[†] Department of Economics, University of Exeter Business School; email: K.Das@exeter.ac.uk

1 Introduction

This paper addresses the non-cooperative behaviour of players in a game of strategic experimentation with two armed bandits when there is competition between the agents and private arrival of informations.

The trade-off between exploration and exploitation is faced by economic agents in many real life situations. The two-armed bandit model has been extensively used in the Economics literature to formally address this issue. This stylized model depicts the situation when an economic agent repeatedly chooses between alternative avenues(which are called arms in the formal analysis) to experiment along, with the ultimate objective to maximize the discounted expected payoff. In course of experimenting along an arm, the agent upgrades the likelihood it attributes to the arm being capable of generating rewards. In the present work, I study a variant of the standard exponential two-armed bandit model (with one safe and one risky arm) which has both informational externalities and competition and we have private learning along the risky arm. In this two-armed bandit model, an agent not only learns from his own experimentation, but also learns from the experimentation experiences of others. This gives rise to informational externalities. On the other hand, in the model considered in this paper, only the first player to experience a reward can successfully convert it into a meaningful payoff. In addition to these, the model of this paper has the property of private learning by agents along the risky arm. This means that when an agent experiments along a *good* risky arm, then apart from experiencing the reward(which is publicly observable), it also experiences private signals. A private signal does not yield any payoff, but it completely resolves the uncertainty to the player who experiences it. This is because private signals can be experienced only along a good risky arm. With these features in this model, I show that compared to a full information benchmark (a social planner's problem who observes everything and controls the actions of both the players), in a particular kind of non-cooperative equilibrium, conditional on no arrival there is too much experimentation along the risky arm if the players start with too high prior(probability that the risky arm is good) and too little experimentation if they start off with a low prior.

The setting is a modified version of the now-canonical two armed exponential bandit model of experimentation(Keller, Rady and Cripps (2005)[8]). We have two homogeneous players, both of whom can access a common two-armed exponential bandit in continuous time. One of the arms is safe(S) and the other one is risky(R). A risky arm can either be *good* or *bad*. Throughout the paper, we will be concerned with two kinds of arrivals. One, is the arrival of reward which is publicly observable. From now on we will call it as the publicly observable arrival. The other is an informational arrival, which is only observable to the player who experiences it. Suppose a player is experimenting along the risky arm. If the risky arm is good then the player can experience two kinds of arrivals. First is the publicly observable arrival which follows a Poisson process with intensity $\pi_2 > 0$. The other one is an informational arrival, which follows a Poisson process of intensity $\pi_1 > 0$. If the risky arm is bad, the player experiences no arrival. On the other hand, if a player is experimenting along the safe arm then he experiences publicly observable arrivals according to a Poisson process with intensity π_0 with $\pi_2 > \pi_0$. Only the first publicly observable arrival yields a positive payoff of 1 unit. Each player can observe the action of the other. They start of with a common prior p, the probability with which the risky arm is good, and update their beliefs as per their own private arrival and the publicly observable arrivals and the action of the opponents.

We first obtain the efficiency benchmark or the full information optimal of this model, i.e when both the players are controlled by a social planner, who can observe all arrivals experienced. Hence, both the players and the planner share a common belief about the state of the risky arm. The planner at each instant allocates each player to an arm. As soon as there is a publicly observable arrival, the experimentation ends. If any of the players experiences an informational arrival, then all uncertainties are resolved and both the players thereon are allocated to the risky arm(which, in fact is now found to be good). The solution is of threshold type. There exists a threshold belief p^* such that conditional on no arrival, both players are allocated to the risky arm if $p > p^*$ and to the safe arm otherwise.

Next, we turn our attention to the non-cooperative game. We restrict ourselves to symmetric markovian equilibria. This implies that on the equilibrium path, given same information, actions will be identical across players. Hence, if the players start with a common prior, then on the equilibrium path both players would be experimenting along the risky arm if the prior exceeds a threshold p^{*N}. If initially a player starts experimenting along the risky arm then conditional on observing nothing, it switches to the safe arm if the posterior is less than or equal to p^{*N}. Since the players are homogeneous and their actions are identical on the equilibrium path, players' posterior, although private, will be identical across them. If a player, while experimenting along the risky arm experiences an informational arrival, then it keeps on experimenting along the risky arm as long the game continues. As stated, if initially a player starts experimenting along the risky arm and gets no arrival till the belief hits p^{*N}, then it switches to the safe arm. However, if it observes that its competitor has not switched, then it reverts back to the risky arm again. This is because the action of the competitor gives the player a signal that an informational arrival has been experienced at the risky arm and thus it is good. If such an event occurs and the competitor switches to the safe arm after some time, then the player who had reverted back to the risky arm would also follow suit, conditional on experiencing no informational arrival in between. This actually deters a player to not to switch to the safe arm when it is supposed to. We establish the existence of a unique equilibrium as described.

Having described the full information optimal and a non-cooperative equilibrium, we try to analyse the nature of inefficiency. We observe that $p^{*N} > p^*$. However, this will not help us to determine the nature of inefficiency in the non-cooperative interaction, if there is any. This is because in the benchmark case, the beliefs are *public* and in the non-cooperative case the beliefs are *private*. Moreover, the belief updating processes are different. In the non-cooperative game, movement of beliefs are sluggish. Hence, to determine the nature of inefficiency, we adopt a different method as follows

First of all, for each initial prior, at which the planner would have allocated both the players to the risky arm, we try to calculate the duration for which the players are made to experiment along the risky arm, conditional on no observation. Then we compare this with the duration for which the firms would be in the risky arm in the equilibrium described above for the non-cooperative game, given the same prior.

It is trivially true that if the prior is in the range (p^*, p^{*N}), in the non cooperative game, the duration for which the players experiment along the risky arm (which is actually 0) is less than that a planner would have wanted. Then we establish the existence of a threshold belief $p_{0*} \in (p^{*N}, 1)$, such that if the initial prior is higher (lower) than this threshold, then the duration for which the players experiment along the risky arm in the equilibrium of the non-cooperative game is higher (lower) than that a planner would have wanted. Hence, too much optimism results in excessive experimentation along the risky arm.

To cite an example which would motivate this research, consider the world of academia. Often two researchers try to solve the same problem independently. Whoever solves the problem first, gets a disproportionately higher payoff (say a very good publication) than the subsequent researcher solving the problem. In this situation, it is very likely that one of them may get an interim result earlier. This individual now has two options: Either to reveal this interim discovery or to conceal it. Revealing might give an instantaneous payoff(say a publication in a relatively low ranked journal). However, this also increases the probability of the competing researcher solving the final problem earlier. This shows that a researcher will not always have incentive to reveal his interim success. In particular, in the absence of any interim payoff a researcher will never reveal any interim result. In the present paper we consider an environment where there is no payoff from revealing the interim result.

Related Literature: This paper contributes to the Strategic Bandit literature. Most of the literature on two-armed bandit, have considered models where all arrivals are publicly observable. In most of them there is absence of payoff externalities and they in general have obtained the result that non-cooperative equilibrium is inefficient and inefficiency is in form of too little experimentation due to free-riding ([8], [9], [11], [10]). In this paper we have payoff externalities between the players in form of competition. Other works which have considered payoff externalities in the strategic bandit literature are [3], [4] and [14].

One of the key features of the present paper is that there is private arrival of information along the good risky arm. To the best of my knowledge, there are only two papers which have analysed this issue.

The first one is the work by Akcigit and Liu ([1]). They analyse a two-armed bandit model with one risky and one safe arm. The risky arm could potentially lead to a dead end. Inefficiency arises from the fact that there is wasteful dead-end replication and an early abandonment of the risky project. The present work also incorporates the issue of private arrival of information. The private information is in the form of *good news* about the risky arm, unlike their work where private information is in the form of *bad news*. However, the present work shows that there can still be early abandonment of the risky project, if to start with players are *not* too much optimistic about the quality of the risky line. Further, in the present work we have learning even when there is no information asymmetry.

The other work is the one by Heidhues, Rady and Strack ([7]). They analyse a model of strategic experimentation where there are 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 present paper differs from their work in the following ways. Firstly, we have private arrivals of information only. Secondly, players are rivals against each other.

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

2 Environment

Two players (1 and 2) face a common continuous time two-armed exponential bandit. Both players can access each of the arms. One of the arms is *safe*(S) and the other one is *risky*(R). A player experimenting along a safe arm experiences publicly observed arrivals according to a Poisson process with commonly known intensity $\pi_0 > 0$. A risky arm can either be *good* or *bad*. A player experimenting along a good risky arm can experience two kinds of arrivals. One of these is publicly observable and it arrives according to a Poisson process with intensity $\pi_2 > \pi_0$. The other kind of arrival is only privately observable to the player who experiences it. It arrives according to a Poisson process with intensity $\pi_1 > 0$. Only the first public arrival (along any of the arms) yields a payoff of 1 unit to the player who experiences it.

Players start with a common prior p^0, which is the likelihood they attribute to the risky arm being good. Players can observe each other's actions. Hence at each time point players update their beliefs on the basis of the public history (publicly observable arrivals and the actions of the players).

We start our analysis with the benchmark case, the social planner's problem. The planner is benevolent and can observe all the arrivals experienced by the players. Hence this can also be called the full information optimal.

2.1 The planner's problem: The full information optimal

In this sub-section we discuss the optimisation problem of a benevolent social planner who can complete control the actions of the players and can observe all the arrivals experienced by them. This is intended to be the efficient benchmark of the model described above. Before we move on to the formal analysis, we demonstrate the process of belief updating in this situation.

The action of the planner at time point t is defined by $k_t(k_t = 0, 1, 2)$. k_t is the number of players the planner makes to experiment along the risky arm. $k_t(t \geq 0)$ is measurable with respect to the information available at the time point t.

Let p_t be the prior at the time point t. Then if there is no arrival over the time interval $\Delta > 0$, it must be the case that none of the players who were experimenting along the risky arm experienced any arrival. This is because the planner can observe all arrivals experienced by the players. Hence using Bayes' rule we can posit that the posterior $p_{t+\Delta}$ at the time point $(t + \Delta)$ will be given by

$$p_{t+\Delta} = \frac{p_t \exp^{-k_t(\pi_1+\pi_2)}}{p_t \exp^{-k_t(\pi_1+\pi_2)} + 1 - p_t}$$

The above expression is decreasing in both Δ and k. Longer the planner has players experimenting along the risky arm without any arrival, more pessimistic they become about the likelihood of the risky arm being good. Also, higher is the number of players experimenting along the risky arm without any arrival, higher is the extent to which the belief is updated downwards.

Let $dt = \Delta$. As $\Delta \rightarrow 0$, the law of motion followed by the belief will be given as (we do away with the time subscript from now on):

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

As soon as the planner observes any arrival at the risky arm, the uncertainty is resolved. If it is an arrival which would have been publicly observable in the non-cooperative game, then the game ends. For the other kind of arrival, the planner gets to know for sure that it is a good risky arm and makes both the players to experiment along it then on, until any first kind of arrival is observed.

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

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

$$+\big(1-r\,dt\big)\big(1-(2-k)\pi_0\,dt - kp(\pi_1+\pi_2)\,dt\big)\big(v(p) - v^{'}(p)kp(1-p)(\pi_1+\pi_2)\,dt\big)\big\}$$

By ignoring the terms of the order $o(\,dt)$ and rearranging the above we obtain the following Bellman equation

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

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

Lemma 1 *There exists a threshold belief* $p^* = \frac{\pi_0}{\pi_2 + \frac{2\pi_1\{\pi_2-\pi_0\}}{r+2\pi_2}}$*, such that if the belief* p *at any point is strictly greater than* p^**, the planner makes both the players to experiment along the risky arm and if the belief is less than or equal to* p^**, the planner makes both the players to experiment along the safe arm.*

Proof of Lemma.

The Bellman equation given by (1) is linear in k. Hence we can posit that at the optimal either $k=0$ or $k=2$. If $k=0$, then $v = \frac{2\pi_0}{r+2\pi_0}$. If $k=2$ then v satisfies the following first order O.D.E:

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

This is derived from (1) by putting $k=2$. The solution to this O.D.E is

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

where C is the integration constant and $\Lambda(p) = \frac{(1-p)}{p}$.

Let p^* be the belief at which the planner makes both players to switch to the safe arm from the risky arm. For $p=1$, the planner will make both players to experiment along the risky arm. For any $p\in(0,1)$, belief can change in leftward direction only. Thus left continuity of $v(p)$ can always be assumed. This implies that for p in the $\epsilon-$ neighbourhood of 1 the planner will still make both players to experiment along the risky arm. We need to determine the threshold belief p^* at which the planner will switch both the players to the safe arm. This is obtained by solving the optimal stopping problem of the planner.

Since at $p = 0$, the planner makes both players to experiment along the safe arm, we must have $p^* \in (0, 1)$. Thus $v(p)$ at p^* should satisfy the value matching and smooth pasting condition.

From the value matching condition at p^* we have

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

Smooth pasting condition at p^* implies $v^{'}(p^*+) = 0$. From (2) we have

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

Substituting the value of C and imposing the smooth pasting condition at p^*, we obtain

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

$$\Rightarrow p^* = \frac{\pi_0}{\pi_2 + \frac{2\pi_1\{(\pi_2-\pi_0)\}}{(r+2\pi_2)}} \tag{3}$$

This completes the proof of the lemma.

■

Das(2013)([4]) solves similar social planner's problem with the absence of any arrival of information. In terms of the parameters of the present model, the threshold belief in that case is $\frac{\pi_0}{\pi_2}$. Hence by comparing this threshold with the one obtained in the present model we can conclude that arrival of information induces the planner to experiment along the risky arm for larger range of belief, which confirms our intuition.

The next section describes the non-cooperative game and a symmetric equilibrium.

3 The non-cooperative game

In this section we consider the non-cooperative game between the players in the environment specified above. Players can observe each others' actions. The informational arrival is only privately observable to the player who experiences it. The other kind of arrival is publicly observable. Only the first publicly observable arrival yields a payoff of 1 unit to the player who experiences it. We assume that players start with a common prior p^0. Since in the present model not all arrivals are publicly observable, belief of each individual will be private to him. Each player chooses a posterior such that if their private belief exceeds that then they choose to experiment along the risky arm, else they experiment along the safe arm. In the present work, we discuss a particular kind of equilibrium as described in the following subsection.

3.1 Equilibrium

In this subsection we discuss the nature of the equilibrium we intend to describe for the non-cooperative game described above. It is assumed that players start with a common prior. The equilibrium we describe is of *symmetric markovian* kind. In the current set-up *markovian* implies that given the common prior, each player first chooses to experiment along an arm. If a player chooses the risky arm initially then he also chooses a threshold such that conditional on no observation the player would switch to the safe arm if the posterior(which is private in this case) is less than or equal to this threshold. By *symmetric* we mean that this threshold is same for both the players. Hence although the beliefs are private, conditional on no arrival players will always have the same posterior on the equilibrium path.

Assuming the existence of an equilibrium as described above, suppose p^{*N} is the common threshold where both players switch to the safe arm, conditional on no arrival[1]. Then, consider a situation when both players have a common belief $p > p^{*N}$. Then according to the conjectured equilibrium, both players should be experimenting along the risky arm. If over a time interval $\Delta > 0$, a player does not observe anything then he infers that none of the players have experienced any publicly observable arrival and he himself has not experienced any informational arrival. Thus the private posterior at $t + \Delta$ will be (since players are symmetric, this will also be the private posterior of the other player) given by

$$p_{t+\Delta} = \frac{p_t e^{-(\pi_1+2\pi_2)\Delta}}{p_t e^{-(\pi_1+2\pi_2)\Delta} + (1 - p_t)}$$

This is because during the time interval $[t, t + \Delta]$, conditional on the risky arm being good, probability that a player does not experience any informational arrival or publicly observable arrival is $e^{-(\pi_1+\pi_2)\Delta}$ and the probability that the opponent does

[1] or observation by any of the players, since any arrival is observed by either of the players

not experience any publicly observable arrival is $e^{-(\pi_2)\Delta}$. Hence probability that the risky arm is good and the player does not observe anything is $pe^{-(\pi_1+2\pi_2)\Delta}$. Then, by applying Bayes' rule we obtain the above expression.

As $\Delta \to 0$, in the proposed equilibrium players' common posterior for $p \geq p^{*N}$ satisfies the following law of motion

$$dp_t = -(\pi_1 + 2\pi_2)p_t(1 - p_t)\,dt$$

Suppose the common prior the players start with is strictly greater than p^{*N}. According to our conjectured equilibrium, both players will choose to experiment along the risky arm. Conditional on no observation, a player would switch to the safe arm as the belief hits the point p^{*N}. If a player at p^{*N} observes that the opponent has not switched to the safe arm, then it instantaneously switches back to the risky arm and follows his opponent then on, conditional on not experiencing any informational arrival. This is because on the equilibrium path, at the switching point if a player is not switching, then it must be the case that he has experienced an informational arrival. We make an assumption that it is possible for a player to instantaneously(costless) switch between the arms. Further, If a player has deviated by not switching to the safe arm at the belief p^{*N}, then conditional on not experiencing an informational arrival, it immediately switches to the safe arm as soon as the belief is less than p^{*N}. We will explain the significance of these off the equilibrium path behaviour later, after describing the equilibrium.

The following lemma establishes that if an equilibrium as described above exists, then the common belief p^{*N} where both players would switch to the safe arm should never be greater than a particular threshold.

Lemma 2 *If an equilibrium as described above exists, then the common threshold belief for players to switch to the safe arm from the risky arm should satisfy*

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

Proof of Lemma.

Suppose there exists a symmetric markovian equilibrium as conjectured above and p^{*N} is the common belief where conditional on no observation players switch to the safe arm from the risky arm. Let the action of player i ($i = 1, 2$) be denoted by k_i.

$k_i \in \{0,1\}$. $k_i = 0(1)$ implies that the player is choosing to experiment along the safe(risky) arm. On the equilibrium path, for $p > p^{*N}$ both players experiment along the risky arm. Let v_1 be the optimal value function of player i $(i = 1,2)$. Hence given k_2, v_1 along with k_1 should satisfy

$$v_1 = \max_{k_1 \in \{1,0\}} \big\{(1-k_1)\pi_0\,dt + k_1 p\big[\pi_2\,dt + \pi_1\frac{\pi_2}{r+2\pi_2}\,dt\big] +$$

$$(1-r\,dt)\big[1-(2-k_1-k_2)\pi_0\,dt - [k_1(\pi_1+\pi_2)p\,dt + k_2\pi_2 p\,dt]\big]\big[v_1 - v_1^{'}p(1-p)[k_1(\pi_1+\pi_2)+k_2\pi_2]\,dt\big]$$

$$+k_2 p\,dt\pi_1\frac{\pi_2}{r+2\pi_2}\big\}$$

Since to player 1 k_2 is given, by ignoring the term of the order $0(\,dt)$ and rearranging the above we can say that v_1 along with k_1 satisfies the following Bellman equation

$$rv_1 = \max_{k_1 \in \{0,1\}} \big\{(1-k_1)\big[\pi_0(1-v_1)\big] + k_1 p\big[(\frac{\pi_2(r+\pi_1+2\pi_2)}{r+2\pi_2}) - (\pi_1+\pi_2)v_1 - v_1^{'}(1-p)(\pi_1+\pi_2)\big]\big\}$$

$$-(1-k_2)\pi_0 v_1 - k_2\big[p\pi_2 v_1 + \pi_2 p(1-p)v_1^{'}\big] + k_2 p\pi_1\frac{\pi_2}{r+2\pi_2} \tag{4}$$

Define $B_s(p)$ and $B_r(p)$ as

$$B_s(p) = \big[\pi_0(1-v_1)\big] \tag{5}$$

$$B_r(p) = p\big[(\frac{\pi_2(r+\pi_1+2\pi_2)}{r+2\pi_2}) - (\pi_1+\pi_2)v_1 - v_1^{'}(1-p)(\pi_1+\pi_2)\big] \tag{6}$$

Thus $B_s(p)$ $(B_r(p))$ is the benefit of experimenting along the safe arm (risky arm) at the belief p. From (4) it is clear that if at a particular p it is optimal for player 1 to experiment along the risky (safe) arm, then we shall have $B_r(p) \geq (\leq) B_s(p)$.

According to the conjectured equilibrium, given player 2's strategy, player 1 finds it optimal to switch to the safe arm at $p = p^{*N}$. Hence at $p = p^{*N}$ we must have

$$B_s(p) \geq B_r(p) \Rightarrow \pi_0\big(1-v_1\big) \geq p\big[(\frac{\pi_2(r+\pi_1+2\pi_2)}{r+2\pi_2}) - (\pi_1+\pi_2)v_1 - v_1^{'}(1-p)(\pi_1+\pi_2)\big]$$

In the conjectured equilibrium, both players switch to S at $p = p^{*N}$. This implies that in equilibrium, the left derivative of v_1 at p^{*N} is zero. Given k_2, if player 1 remains at the risky arm at $p = p^{*N}$, then conditional on there being no arrival, belief can change only in the leftward direction. Hence at $p = p^{*N}$, if player 1 decides to deviate and not switch to the safe arm, then the left derivative will be relevant. This implies

$$\pi_0\big(1 - v_1(p^{*N})\big) \geq p^{*N}\big[(\frac{\pi_2(r + \pi_1 + 2\pi_2)}{r + 2\pi_2}) - (\pi_1 + \pi_2)v_1(p^{*N})\big]$$

Since belief can change only in the leftward direction, left continuity of v_1 can always be assumed. Hence v_1 would satisfy value matching condition at p^{*N}. This implies $v_1(p^{*N}) = \frac{\pi_0}{r+2\pi_0}$. Thus we shall have

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

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

This concludes the proof of the lemma ■

In equilibrium, both players experiment along the risky arm for $p > p^{*N}$. Thus v_1 should satisfy the following O.D.E

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

This is obtained by putting $k_1 = 1$ and $k_2 = 1$ in (4). Solving this O.D.E we obtain

$$v_1 = \frac{\pi_2}{r + 2\pi_2}[\frac{r + 2\pi_1 + 2\pi_2}{r + \pi_1 + 2\pi_2}]p + C(1 - p)[\Lambda(p)]^{\frac{r}{\pi_1+2\pi_2}} \tag{7}$$

where C and $\Lambda(.)$ are as defined before. The derivative of (7) with respect to p is then given by

$$v_1^{'} = \frac{\pi_2}{r + 2\pi_2}[\frac{r + 2\pi_1 + 2\pi_2}{r + \pi_1 + 2\pi_2}] - C[\Lambda(p)]^{\frac{r}{\pi_1+2\pi_2}}[1 + \frac{r}{\pi_1 + 2\pi_2}\frac{1}{p}] \tag{8}$$

We are now in a position to prove that if an equilibrium as conjectured above exists then it is unique, that is the common belief p^{*N} can take a single value. The following lemma describes this

Lemma 3 *If an equilibrium as conjectured above exists, then it must be unique with*

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

Proof of Lemma. We begin the proof of this lemma by first proving the following claim.

Claim. If an equilibrium as conjectured exists with a common switching point p^{*N}, then it is never possible to have $v_1^{'}(p) < 0$ for $p = p^{*N} + \epsilon$, $\epsilon > 0$ and ϵ is arbitrarily small. ■

Proof of the claim. From (6) we have

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

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

From the O.D.E which v_1 satisfies when $p > p^{*N}$, we can conclude that

$$B_r(p) = rv_1 + \pi_2 p v_1 - \frac{\pi_1\pi_2}{r + 2\pi_2}p + \pi_2 p(1-p)v_1^{'} \tag{9}$$

whenever $p > p^{*N}$.

Consider $p = p^{*N} + \epsilon$, such that $\epsilon > 0$ and ϵ is arbitrarily small. Since v_1 is left continuous and satisfies the value matching condition at p^{*N}, $v_1 \approx \frac{\pi_0}{r+2\pi_0}$. This implies

$$B_s(p) \approx r\frac{\pi_0}{r + 2\pi_0} + \frac{\pi_0^2}{r + 2\pi_0}$$

Now suppose $v_1^{'}(p^{*N}+) < 0$. Then

$$B_r(p) < rv_1 + \pi_2 p v_1 \approx r\frac{\pi_0}{r + 2\pi_0} + \pi_2 p\frac{\pi_0}{r + 2\pi_0}$$

Since $p^{*N} \leq \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2 - \pi_0)\frac{r}{r+\pi_0}} < \frac{\pi_0}{\pi_2}$, we can conclude that

$$r\frac{\pi_0}{r + 2\pi_0} + \pi_2 p\frac{\pi_0}{r + 2\pi_0} < r\frac{\pi_0}{r + 2\pi_0} + \frac{\pi_0^2}{r + 2\pi_0} = B_s(p)$$

Thus we have $B_r(p) < B_s(p)$. This is not possible in equilibrium. Hence our supposition that $v_1^{'}(p) < 0$ leads us to contradiction. Hence $v_1^{'}(p) \geq 0$. In fact we can say from the above that we should have $v_1^{'}(p) > 0$ for $p = p^{*N} + \epsilon$, $\epsilon > 0$ and ϵ arbitrarily small. This proves the claim. ■

Using the above claim, we shall now show that $p^{*N} = \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2 - \pi_0)\frac{r}{r+\pi_0}}$.

Suppose it is the case that $p^{*N} < \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2 - \pi_0)\frac{r}{r+\pi_0}}$. Then from our previous analysis we know that $B_r(p^{*N}) < B_s(p^{*N})$. Since both players switch to the safe arm from the risky arm at $p = p^{*N}$, The left derivative of v_1 at p^{*N} should be equal to 0. Since we have $B_s(p^{*N}) > B_r(p^{*N})$, this implies

$$\pi_0(1 - v_1) > p^{*N}[\frac{\pi_2(r + \pi_1 + 2\pi_2)}{r + 2\pi_2} - (\pi_1 + \pi_2)v_1]$$

Since v_1 is left continuous, the above inequality will hold strictly for $p = p^{*N} + \epsilon$, $\epsilon > 0$ and ϵ arbitrarily small. We have already proved that $v_1^{'}(p^{*N} + \epsilon) > 0$. Hence

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

$$\Rightarrow B_s(p) > B_r(p)$$

This is not possible in equilibrium. Hence we must have $B_s(p^{*N}) = B_r(p^{*N})$. This implies that p^{*N} is unique and is equal to $\frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2 - \pi_0)\frac{r}{r+\pi_0}}$. This concludes the proof of the lemma.

■

The above two lemmas have described that if a symmetric markovian equilibrium as described above exists then it must be unique. We now establish the existence of such an equilibrium in the following proposition.

Proposition 1 *An equilibrium as described above always exists.*

Proof.

We begin the proof of this proposition by first proving the following claim.

Claim. If for $p > p^{*N} = \frac{\pi_0}{\pi_2 + \frac{\pi_1}{r+2\pi_2}(\pi_2 - \pi_0)\frac{r}{r+\pi_0}}$, $v_i^{'}(p) > 0$ $(i = 1, 2)$ and both players are experimenting along the risky arm, then given that one player is experimenting along the risky arm, the other player will have no incentive to switch to experiment along the safe arm. ■

Proof of the claim.

Suppose the claim is not true. That is, let $v_1^{'}(p)$ be strictly greater than 0 for all $p > p^{*N}$, if both players are experimenting along the risky arm. Let player 2's strategy be to keep experimenting along the risky arm and conditional on no observation remain there until the belief is higher than p^{*N}. Then suppose there exists some $\tilde{p} \in (p^{*N}, 1)$ such that player 1 finds it beneficial to switch to the safe arm at that belief. Now we can say that if player 1 finds it optimal to switch to the safe arm from the risky arm at the belief $\tilde{p}$, then he would still find it optimal to keep experimenting along the safe arm for any belief $p < \tilde{p}$. This is because for $p < \tilde{p}$, the prospects from the risky arm (given player 2's strategy)is lower than that it would have been at $p = \tilde{p}$. Hence $\tilde{p}$ must be an interior solution of the optimal stopping problem of player 1, given that player 2 is experimenting along the risky arm. This implies that v_1 should satisfy the smooth pasting condition at $p = \tilde{p}$. If player 2 is experimenting along the risky arm and player 1 is experimenting along the safe arm then player 1's value function's derivative with respect to p would be negative. This is because higher is p, higher is the probability of player 2 being the first one to experience a publicly observable arrival, and hence the more adverse it is for player 1. Since v_1 is continuously differentiable (smooth pasting) at $\tilde{p}$, both the right and left derivative of v_1 should be negative at $\tilde{p}$. However this is a contradiction since the right derivative is positive(by hypothesis) and the left derivative is negative. This proves the claim.

■

Next, we prove that if both players are experimenting along the risky arm if their common prior exceeds a certain threshold and conditional on no observation they switch at a common threshold belief, then the right derivative of the value function at the switching point is strictly positive if the switching point is p^{*N}.

Suppose both players are experimenting along the risky arm when $p > \bar{p}$. Hence v_1 will be given by (7) and $v_1^{'}$ by (8). Since v_1 would satisfy the value matching condition at $\bar{p}$, from (7) we obtain

$$C = \frac{\frac{\pi_0}{r+2\pi_0} - \frac{\pi_2}{r+2\pi_2}\left[\frac{r+2\pi_1+2\pi_2}{r+\pi_1+2\pi_2}\right]\bar{p}}{(1-\bar{p})[\Lambda(\bar{p})]^{\frac{r}{\pi_1+2\pi_2}}}$$

Then from (8), 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

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

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

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

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

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

clearly $p^{*N} > p^{'}$.

Hence if both players experiment along the risky arm for $p > p^{*N}$, the derivative of the value function of each player with respect to p will be strictly positive for all $p > p^{*N}$. From the claim proved at the beginning of the proof of this proposition, we can posit that no player will have any incentive to switch to the safe arm at any $p > p^{*N}$.

Next, we argue that if a player has actually deviated by not switching to the safe arm at $p = p^{*N}$, then the behavior described above constitute optimal behaviour on the player's part. Suppose the player has deviated. Then the opponent would instantaneously switch back to the risky arm. Now suppose at $p = p^{*N}$ the deviating player has observed nothing. Thus belief would be falling below p^{*N}. The deviating player knows that as soon as it would switch back to the safe arm, the opponent would follow him. Hence the players would be switching back to the safe arm at the same belief. Since $p < p^{*N}$, from our above analysis know that for the deviating player, switching to the safe arm constitutes optimal behaviour Hence the conjectured equilibrium exists and as proved in the previous lemma, it is unique. This concludes the proof of this proposition. ■

Having described the unique equilibrium in the class of symmetric markov equilibria, we would now like to compare the outcome of the equilibrium with that of the benchmark case. First of all observe that $p* < p^{*N}$. Hence there might be some distortion in the non-cooperative equilibrium.

At this juncture, it must be stated that by just comparing the threshold probabilities of switching (p^* in the planner's case and p^{*N} in the non-cooperative case) we cannot infer whether there is too much or too-little experimentation along the risky arm in the non-cooperative equilibrium. This is because 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} then same action profile would give rise to different system of beliefs. In the non-cooperative equilibrium the beliefs are private (although same across individuals) and in the benchmark case it is public. In the present work, we determine the nature of inefficiency in the following manner.

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

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

Proposition 2 *The non-cooperative equilibrium involves inefficiency. There exists a $p_{0*} \in (p^{*N}, 1)$ such that if the prior $p_0 > p_{0*}$, then conditional on no arrival we have excessive experimentation and for $p^0 < p^{*0}$ we have too little experimentation. By* *excessive experimentation we mean that starting from a prior the duration for which players experiment along the risky arm is more than that a planner would have liked to.*

Proof. Let $t^n_{p_0}$ be the duration of experimentation along the risky line by the firms in the non-cooperative equilibrium described above when they start from the prior p_0. From the non-cooperative equilibrium described above we know that of the firms start out from the prior p_0 then they would carry on experimentation along the risky line until the posterior reaches p^{*N}. From the dynamics of the posterior we know that

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

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

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

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

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

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

We have excessive experimentation when $t^n_{p_0} > t^p_{p_0}$. This is the case 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 $\pi_1\log[\Lambda(p_0)] \equiv \tau(p)$. Since logarithm is a monotonically increasing function and $\Lambda(p)$ is monotonically decreasing in p. Hence $\tau(p)$ is monotonically decreasing in p.

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

The R.H.S can be written as

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

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

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

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

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

These imply that

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

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^*, then there is excessive experimentation along the risky arm and if it is below the threshold there is too little experimentation along the risky arm in the non-cooperative equilibrium.

This concludes the proof of this proposition ■

In the non-cooperative equilibrium, distortion arises from two sources. One, is what we call the *implicit* free-riding effect. This comes from the fact that if a player experiences a private arrival of information, then the benefit from that is also reaped by the other competing player. This is possible here because of instantaneous costless switching back to the risky arm. In fact, if information arrival to firms would have been public, then the non-cooperative equilibrium would always involve free-riding. This follows directly from ([8]). Thus this implicit free riding effect tends to reduce the duration of experimentation along the risky arm.

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

4 Conclusion

This paper has analysed a tractable model to explore the situation when there can be private arrival of information. We show that there can be a non-cooperative equilibrium where depending on the prior we can have both too much and too little experimentation along the risky line. This result has been obtained under the assumption that players can switch between arms without incurring any cost (revocable switching). It will be interesting to see how the results change if a player after switching to the safe arm is unable to revert back to the risky arm immediately. Hence switching back to the risky arm is costly. In addition to it, once we introduce payoff from revealing informational arrival, then there might be situations where a player would have incentive to reveal a private observation. These issues will be addressed in my near future research.

References

[1] Akcigit, U., Liu, Q., 2013: "The Role of Information in Competitive Experimentation. ", *mimeo, Columbia University and University of Pennsylvania.*

[2] Bolton, P., Harris, C., 1999 "Strategic Experimentation. ", *Econometrica* 67, 349 – 374.

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

[4] Das, K. 2015: "The Role of Heterogeneity in a Model of Strategic Experimentation ", *Working paper, University of Exeter*

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

[6] Graham, M.B.W., 1986 "The Business of research ", *New York:Cambridge University Press.*

[7] Heidhues, P., Rady, S., Strack, P., 2015 "Strategic Experimentation with Private Payoffs ", *Forthcoming in Journal of Economic Theory*

[8] Keller, G., Rady, S., Cripps, M., 2005: “Strategic Experimentation with Exponential Bandits ”, *Econometrica* 73, 39 − 68.

[9] Keller, G., Rady, S., 2010:“Strategic Experimentation with Poisson Bandits ”, *Theoretical Economics* 5, 275 − 311.

[10] Klein, N., 2013: “Strategic Learning in Teams ”, *forthcoming in Games and Economic Behavior*

[11] Klein, N., Rady, S., 2011: “Negatively Correlated Bandits ”, *The Review of Economic Studies* 78 693 − 792.

[12] Presman, E.L., 1990: “Poisson Version of the Two-Armed Bandit Problem with Discounting, *Theory of Probability and its Applications*

[13] Stokey,N.L., 2009: “The Economics of Inaction ”, *Princeton University Press.*

[14] Thomas, C., 2011: “Experimentation with Congestion ”, *mimeo, University College of London and University of Texas Austin*