Strategic Experimentation with Heterogeneous Agents and Payoff Externalities *

Kaustav Das†

August 28, 2014

Abstract

In this paper, I examine the effect of introducing heterogeneity between players in models of strategic experimentation. I consider a two-armed bandit problem with one safe arm and a risky arm. There are two players and each has an access to such a bandit. A player using the safe arm experiences a safe flow payoff . The risky arm can either be good or bad. A bad risky arm is worse than the safe arm and the good risky arm is better than the safe arm. Players start with a common prior about the probability of the risky arm being good. We show that without any payoff externalities, heterogeneous players in non cooperative equilibrium do more experimentation than a model with homogeneous players. When competition is introduced between the players, inefficiency in non cooperative equilibrium is in form of too much of experimentation along the risky arm.

JEL Classification Numbers:C73, D83, O31.

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

*This is based on chapter 1 of my Ph.D dissertation written at the Pennsylvania State University. Earlier versions of the paper were circulated under the titles *Competition in R&D and Sharing of Innovative Knowledge* and *Competition, Duplication and Learning in R&D*. I thank my advisor Kalyan Chatterjee for his helpful suggestions and for providing me with continuous encouragement. I also thank Vijay Krishna, Qingmin Liu, Venky Venkateswaran, David Kelsey, Dieter Balkenborg, Rajiv Sarin, James Jordan, Tymofiy Mylovanov, Neil Wallace, and Alexander Monge-Naranjo for their helpful comments. Pathikrit Basu was kind enough to go through this draft in detail and make intricate comments. The section on heterogeneous players without payoff externalities was developed while I was visiting University of Bonn. A special note of thanks to Sven Rady for the helpful discussions with him. The remaining errors are my own responsibility.

†Department of Economics, University of Exeter Business School, Email:K.Das@exeter.ac.uk

1 Introduction

In this paper, I address the problem of optimal behavior of players in a game of strategic experimentation with two-armed bandits where there are both informational and payoff externalities as well as heterogeneous players.

In the economics literature, the two-armed bandit models have been extensively used to formally address the issue of trade-offs between exploration and exploitation in dynamic decision problems with learning. In the standard continuous time exponential bandit model, an agent has to decide how long to experiment along an arm to get rewarded before switching to experimenting along another arm. As the agent experiments along a particular arm without getting rewarded, the likelihood he attributes to ever getting rewarded along that arm is revised downwards. In this paper, I study models of strategic experimentation that incorporate variants of this standard exponential bandit with two arms. Both informational and payoff externalities are present in the models and players are heterogeneous. Informational externalities arise from the fact that an agent's learning about the state of the reward process along an arm is not only influenced by his own experimentation experiences but also by the behavior of other agents. On the other hand, payoff externalities imply that the extent to which an agent can convert a reward into a meaningful payoff depends on the order in which he gets the reward with respect to other agents. Finally, heterogeneous players mean that players have different innate abilities along different arms. Given that a reward occurs along an arm, the expected time required to get that reward differs among agents. With these features, I show that in a game of strategic experimentation, the non-cooperative equilibrium (markovian) always involves inefficient experimentation. The inefficiency is in the form of too much duplication. This means that there arise instances when all agents experiment along the same arm, though the social planner would have prefered the agents to diversify their experimentation along different arms.

The analysis starts with first introducing only heterogeneity in the now cannonical form of Two-armed Bandit Model (*a.la* Keller, Rady and Cripps). That is, each player faces a common two armed exponential bandit. One of the arms is safe and a player accessing it gets a flow payoff of $s > 0$. The other arm is either good or bad. A player who accesses the good risky arm gets an arrival according to a Poisson process with known intensity. Each arrival gives a lumpsum payoff, which is drawn from a time-invariant distribution with mean $h > 0$. Players differ with respect to their innate abilities. This means the poisson intensity with which a player experiences an arrival along a good risky arm differs across

players. Player 1's intensity is λ_1 and that of player 2 is λ_2 with $\lambda_1 > \lambda_2$. Hence player 1's flow payoff along a good risky arm $g_1 = \lambda_1 h$ and that of player 2 is $g_2 = \lambda_2 h$ such that $g_1 > g_2 > s$.

We start with analysing the social planner's problem, which aims to maximise the sum of the expected surplus of the players. The planner, in a continuous time, decides on allocating players to one of the arms. The social optimal involves *specialisation* for extreme range of beliefs and *diversification* for interim range of beliefs. This means that if it is too likely that one of the arms is good (in this setting this implies belief being close to either 0 or 1), then both the players are made to access this arm. For interim beliefs, the wekaer player (player 2) is allocated to the safe arm and the stronger player (player 1) is allocated to the risky arm.

For the analysis of noncooperative solutions, we restrict ourselves to Markovian strategies with the common posterior belief as the state variable. The first main result shows that there cannot be an efficient equilibrium. We show that if the degree of heterogeneity is large enough then there exists a unique inefficient diversification equilibrium. The belief at which all experimentation c eases is greater than that in the optimal solution of the planner's problem. Also, player 2 shifts to the risky arm at a belief greater than that in the planner's solution. This is due to free riding.

Next, I compare the extent of experimentation in a model with heterogeneous players to that in model with homogeneous players. It has been shown that if the degree of heterogenity is high enough then keeping the total ability constant (which is the sum of the poisson intensities of the players) the amount of experimentation in a model with heterogeneous players is more than that in any non cooperative equilibrium with homegeneous players.

In the second setting, we introduce payoff externalities . The previous model is modified as follows. A player acessing a safe arm experiences arrivals according to a poisson process with intensity $\pi_0 > 0$. Player 1 along a good risky arm experiences arrivals according to a poisson process with intensity π_1 and player 2 experiences arrivals with intensity π_2. We have $\pi_1 > \pi_2 > \pi_0$. Payoff externalities arises from the fact that only the first arrival across any of the arms yields a payoff of one unit to the player who experiences it. The planner's solution is qualitatively similar to that in the previous setting. When it is very likely (unlikely) that the risky arm is good(bad)[1], both the players are made to access the risky (safe) arm. For interim range of beliefs, the player who is relatively better off along a risky arm is made to access it and the other player is made to access the safe arm. The

[1] that is for high (low) beliefs

noncooperative equilibrium obtained is unique in the class of markovian equilibria in cut-off strategies. It is inefficient and inefficiency is in the form of too much duplication. Thsi means that there exists a range of beliefs over which the less efficient player still accesses the risky arm when efficiency would require him to access the safe arm. As before, we find that for homogeneous players, the equilibrium is always efficient.

In real world, there are many instances where agents have alternative potential approaches to pursue in order to achieve the same goal and they compete for success. Consider a situation where competing agents who are trying to make the same discovery, have a choice between potential alternate methods and the rent accruing to the second inventor is disproportionately lower than the first. This is true in many contexts. We can think of two firms engaged in a R&D race, who have alternate research methods or hypothesis to pursue. Firms do not know which method would lead to success. However, they are aware of a likelihood by which each avenue could lead to success. In this regard, one can cite an example from the pharmaceutical industry, where firms are competing to invent a drug for the *Alzheimer's* disease. Firms know that either eliminating the *beta*-amyloid protein or the *tau*-protein would eradicate the disease. Hence, firms need to decide on which hypothesis to adopt and over time they learn about the quality of the methods in the light of their search experiences. Given the high perceived valuation of a possible drug, it is evident that whoever invents the drug first would make a disproportionately higher amount of money than the later inventor(s). One could also think of a situation where two researchers are attempting to explain a scientific phenomenon. There may be alternative forms of explanation, any of which might or might not be correct. At a time there could be only one correct explanation. For example, in the $17th$ century, the Phlogiston theory used to be put forward to explain the process of combustion. However, by the end of the eighteenth century this theory was challenged and finally became void when the new Calorific theory came in. There could be similar situations in a firm also. Consider a manager who has two or more employees under his control. The manager needs to get an assignment done and would reward the employee in form of a bonus to the one who does it first. The employees have to choose among several alternate avenues to get the assignment done, although they are not sure which avenue would finally lead to success. In this case it is possible that one of the avenues will surely lead to success, but there is an alternate avenue which can either lead to success at a faster rate or can lead to failure. Clearly here each of the employees competes with others to be the first one to do the assignment successfully. In all the above situations it could be possible that conditional on an avenue being the correct one, agents

would differ in their probability to achieve success along that avenue. For instance, in the pharmaceutical industry example, it is quite possible that one firm may be relatively more efficient in eradicating the β-amyloid protein, while the other may be more proficient in eradicating the τ-protein. The models of strategic experimentation analysed in this paper capture the main features of the situations described above. There are stylized facts in reality which might be due to the phenomenon of too much duplication. Again, consider the Alzheimer's drug research case. It was widely believed that the level of β-amyloid protein is the main culprit. Consequently for the past two decades almost exclusive attention was given to developing drugs to remove amyloid plaques. However, not much success has been attained in this direction. The drugs which are presently in the market, only delay the onset of this disease.([8]) As a consequence of this, the theory that β-amyloid protein is the culprit is waning and the conjecture that *tau*-proteins are to be blamed is gaining ground. However major R&D activities still involve removal of amyloid plaques. This may be due to too much duplication.

Related Literature: This paper contributes to the strategic bandit literature. Some of the works which have studied the bandit problem in the context of economics, are Bolton and Harris ([4]) Keller,Rady and Cripps([11]), Keller and Rady([12]), Klein and Rady ([14]) and Thomas([21]). In all of these papers except ([21]) and ([14]), players have replicas of bandits and *Free-riding* is a common feature in all the above models except ([21]). This leads to an inefficient level (too little) of experimentation. The present work differs from ([11]) and ([12]) in two ways. First, we show the effect of heterogeneity and find that unlike in a model with homogeneous players, for certain range of parameters we can obtain unique non-cooperative equilibrium.

Next, we introduce payoff externalities. Due to this, the phenomenon of free riding does not arise . Secondly, agents differ with respect to their innate abilities. This gives us inefficiency in equilibrium, the nature of which is very different from the ones in ([11]) and ([12]).

Thomas([21]) analyses a set-up where each player has access to an exclusive risky arm, and both of them have access to a common safe arm. At a time the safe arm can be accessed by one player only. Hence, there is congestion along an arm. The present paper differs from this in the way that here each of the arms can be accessed by all the players. Further, we do not have congestion along any of the arms.

This paper also contributes to the relatively less explored area of the broad literature

on R&D races. It shows that in presence of heterogeneity and competition among agents, there is always a distortion in the choice of research avenue in a non-cooperative interaction. Bhattacharya and Mookerjee([3]), Dasgupta and Maskin([6]) are two of the early papers which explore this issue in a static framework. Chatterjee and Evans ([5]) analyses similar issues in a dynamic setting. The first setting of this paper has similarities with [5]. However, we consider a continuous time framework with heterogeneous players. Here we can show that we always have too much duplication in the non-cooperative interaction. Some other papers to look into similar issues are Fershtman and Rubinstein ([9]) and Akcigit and Liu([1]). ([9]) studies a two-stage model in which agents simultaneously rank a finite set of boxes. Exactly one of the boxes contains the prize. Players commit to opening the boxes according to their ranked order. Inefficiency arises due to the fact that the box which is most likely to have the prize is not opened first. Their model is basically static in nature. Hence, the present paper lays down dynamic models which show that inefficiency in R&D with respect to choice of research method, is in form of too much duplication.

The rest of the paper is organised as follows. Section 2 lays down the detail of setting with introducing only heterogeneity and section 3 introduces payoff externalities.

2 Heterogeneous Agents, No Payoff externalities

Model: There are two players (1 and 2). Each player faces a continuous time two-armed bandit. One of the arms is safe and a player who uses it gets a flow payoff of $s > 0$. The risky arm can either be good or bad. If a risky arm is good then a player choosing it experiences arrivals according to a Poisson process with a known intensity. Each arrival gives lumpsum payoffs to the player who experiences it. These lump sums are drawn from a time invariant distribution with mean h. Player 1 experiences this arrivals acording to a Poisson process of intensity $\lambda_1 > 0$ and player 2 experiences this according to a Poisson process with itensity $\lambda_2 > 0$ such that $\lambda_1 > \lambda_2$. Hence along a good risky arm player 1 experiences a flow payoff of $g_1 = \lambda_1 h$ and player 2 experiences a flow payoff of $g_2 = \lambda_2 h$. A player can either choose the safe arm or the risky arm. Player discount the future according to a common discount rate $r > 0$.

2.1 Planner's Problem

Suppose the players are controlled by a benevolent social planner. Let (k_1,k_2) be the action profile of the planner. $k_i \in \{0,1\}$. $k_i = 0$ implies player i is in the safe arm and $k_i = 1$ implies that player i is in the risky arm. The planner wants to maximise the sum of the expected discounted payoffs of the players. If v is the value function of the planner then we must have

$$v = \max_{k_1,k_2\in\{0,1\}} [r\{(1-k_1)s+(1-k_2)s+k_1pg_1+k_2pg_2\}\,dt$$

$$+(1-r\,dt)\{p(k_1\lambda_1+k_2\lambda_2)\,dt(g_1+g_2)+\big(1-p(k_1\lambda_1+k_2\lambda_2)\,dt\big)(v-v^{'}p(1-p)(\lambda_1k_1+\lambda_2k_2)\,dt)\}]$$

Simplifying above and ignoring the terms of order $o(dt)$ we have

$$v = 2s+k_1[b_1(p,v)-c_1(p)]+k_2[b_2(p,v)-c_2(p)]$$

where $c_i(p) = [s-pg_i]$ and

$$b_i(p,v) = \lambda_i p\frac{\{(g_1+g_2)-v-v^{'}(1-p)\}}{r}$$

Conjectured Solution:

There exists thresholds p_1^*, p_2^* with $0 < p_1^* < p_2^* < 1$ such that player 2 is switched to the safe arm at p_2^* and player 1 is switched to the safe arm at p_1^*.

Consider $p < p_2^*$. Thus according to the conjectured solution $k_2 = 0$. If it is optimal for the planner to have 1 in the safe arm then $k_1 = 0$. Thus $v(p) = 2s$. If it is optimal for the planner to have 1 at the risky arm then $k_1 = 1$. $v(.)$ then satisfies the following O.D.E:

$$v^{'}+v\frac{[r+\lambda_1p]}{p(1-p)\lambda_1} = \frac{rs}{p(1-p)\lambda_1}+\frac{[rg_1+\lambda(g_1+g_2)]}{(1-p)\lambda}$$

The solution to the above diferential equation is:

$$v = s+[\frac{\lambda_1g+rg_1}{\lambda_1+r}-\frac{s\lambda_1}{r+\lambda_1}]p+C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

where $g = (g_1+g_2)$; $\Lambda(p) = \frac{(1-p)}{p}$.

Imposing the value matching and the smooth pasting condition at p_1^* we have

$$s+[\frac{\lambda_1 g+rg_1}{\lambda_1+r}-\frac{s\lambda_1}{r+\lambda_1}]p+C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}=2s$$

$$\Rightarrow C=\frac{s-[\frac{\lambda_1 g+rg_1}{\lambda_1+r}-\frac{s\lambda_1}{r+\lambda_1}]p}{(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}}$$

Smooth pasting condition requires that both the right hand and left hand derivative of v at p_1^* is zero. Thus

$$[\frac{\lambda_1 g+rg_1}{\lambda_1+r}-\frac{s\lambda_1}{r+\lambda_1}]-C[\Lambda(p)]^{\frac{r}{\lambda_1}}(1+\frac{r}{\lambda_1 p})=0$$

Substituting C we have

$$[\frac{\lambda_1 g+rg_1}{\lambda_1+r}-\frac{s\lambda_1}{r+\lambda_1}]-\frac{s-[\frac{\lambda_1 g+rg_1}{\lambda_1+r}-\frac{s\lambda_1}{r+\lambda_1}]p}{(1-p_1^*)}(1+\frac{r}{\lambda_1 p})=0$$

$$\Rightarrow p_1^*=\frac{s\mu_1}{(\mu_1+1)g_1+g_2-2s}$$

where $\mu_1=\frac{r}{\lambda_1}$.

Now consider $p>p_2^*$. If the planner finds it optimal to keep both players at the risky arm then $k_1=k_2=1$. The value function then satisfies:

$$v^{'}p(1-p)(\lambda_1+\lambda_2)+v[r+(\lambda_1+\lambda_2)p]=pg(\lambda_1+\lambda_2+r)$$

$$\Rightarrow v(p)=gp+C(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

where $g=g_1+g_2$ and $\lambda=\lambda_1+\lambda_2$.

It is clear that at the belief p_2^*, when the planner finds it optimal to switch player 2 to the safe arm, it must be the case that

$$b_2(p,v)=s-g_2 p$$

For $p\geq p_2^*$ we have

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

Hence $b_2(p,v)$ can be written as

$$\frac{\lambda_2}{\lambda}(1-p)C[\Lambda(p)]^{\frac{r}{\lambda}} = \frac{\lambda_2}{\lambda}[v-gp]$$

Thus $b_2(p_2^*,v) = s - g_2 p_2^*$ gives us

$$v(p_2^*) = \frac{\lambda_1+\lambda_2}{\lambda_2}s > 2s$$

since $\lambda_1 > \lambda_2$. Let $v_{sr}(.)$ be the value function when 1 uses risky arm and 2 uses the safe arm and v_{rr} be the value function when boh players use the risky arm. We have

$$v_{rr}(p_2^*) = v_{sr}(p_2^*) = \frac{\lambda_1+\lambda_2}{\lambda_2}s$$

Thus p_2^* satisfies

$$[\frac{\lambda_1 g + r g_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p_2^* + [\frac{s - [\frac{\lambda_1 g + r g_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p_1^*}{(1-p_1^*)[\Lambda(p_1^*)]^{\frac{r}{\lambda_1}}}](1-p_2^*)[\Lambda(p_2^*)]^{\frac{r}{\lambda_1}} = \frac{\lambda_1}{\lambda_2}s$$

One can show that there actually exists a $p_{2*} \in (p_1^*, 1)$ such that the above relation holds. This is because the L.H.S is equal to $s < \frac{\lambda_1}{\lambda_2}s$ when $p = p_1^*$. At $p = 1$, it is equal to

$$g_1 + \frac{\lambda_1}{r+\lambda}(g_2 - s) > g_1 = \frac{\lambda_1}{\lambda_2}g_2 > \frac{\lambda_1}{\lambda_2}s$$

Since L.H.S is continuous in p and monotonically increasing, there exists a unique $p_2^* \in (p_1^*, 1)$ such that the above equality holds. This confirms that our conjectured solution is correct.

The integration constant of v_{rr} is given by

$$C = \frac{\frac{\lambda_1+\lambda_2}{\lambda_2}s - gp_2^*}{(1-p_2^*)[\Lambda(p_2^*)]^{\frac{r}{\lambda}}}$$

2.2 Non-cooperative solution

The above two players now act noncooperatively. We consider stationary Markovian strategies with the common belief as the state variable.

Best Responses: k_i denotes player i's action given the current belief. $k_i \in \{0,1\}$. $k_i = 0$ means the player is activating the safe arm and $k_i = 1$ means the player is activating the risky arm.

Let v_1 be the value function of player 1. Then it should satisfy the following Bellman equation

$$v_1(p) = \max_{k_1 \in \{0,1\}} \Big\{ r[(1-k_1)s + k_1 p g_1]\,dt + (1-r\,dt)[(k_1\lambda_1 + k_2\lambda_2)p\,dt g_1$$

$$+ (1 - k_1\lambda_1 p\,dt - k_2\lambda_2 p\,dt)(v_1 - v_1' p(1-p)(k_1\lambda_1 + k_2\lambda_2)\,dt) \Big\}$$

Player 1's expected discounted payoff is (after ignoring the terms of order $o(dt)$)

$$(1-r\,dt)(k_1\lambda_1 p\,dt g_1 + k_2\lambda_2 p\,dt g_1 + v_1 - k_1\lambda_1 p\,dt v_1 - k_2\lambda_2 p\,dt v_1 - (k_1\lambda_1 + k_2\lambda_2)p(1-p)v_1')\,dt$$

$$= v_1 + r[\frac{k_1\lambda_1 p\,dt g_1 - k_1\lambda_1 p\,dt v_1 - k_1\lambda_1 p(1-p)\,dt v_1'}{r} + \frac{k_2\lambda_2 p\,dt g_1 - k_2\lambda_2 p\,dt v_1 - k_2\lambda_2 p(1-p)\,dt v_1'}{r} - v_1]$$

After rearranging we have

$$v_1(p) = s + k_2[\lambda_2 b_1(p, v_1)] + \max_{k_1 \in \{0,1\}} k_1[\lambda_1 b_1(p, v_1) - (s - g_1 p)]$$

where

$$b_1(p, v_1) = p\frac{\{g_1 - v_1 - (1-p)v_1'\}}{r}$$

Similarly we have

$$v_2(p) = s + k_1[\lambda_1 b_2(p, v_2)] + \max_{k_2 \in \{0,1\}} k_1[\lambda_2 b_2(p, v_2) - (s - g_2 p)]$$

where

$$b_2(p, v_2) = p\frac{\{g_2 - v_2 - (1-p)v_2'\}}{r}$$

Optimal response of player 1

Consider any $k_2 \in \{0,1\}$.

Player 1's best response is:

$$k_1 = \begin{cases} 1 & : \text{ if } \lambda_1 b_1(p,v_1) > s - g_1 p, \\ \in \{0,1\} & : \text{ if } \lambda_1 b_1(p,v_1) = s - g_1 p, \\ 0 & : \text{ if } \lambda_1 b_1(p,v_1) < s - g_1 p. \end{cases}$$

Putting this in the Bellman equation satisfied by $v_1(p)$ we have

$$k_1 = \begin{cases} 1 & : \text{ if } v_1 > s + k_2 \frac{\lambda_2}{\lambda_1}[s - g_1 p], \\ \in \{0,1\} & : \text{ if } v_1 = s + k_2 \frac{\lambda_2}{\lambda_1}[s - g_1 p], \\ 0 & : \text{ if } v_1 < s + k_2 \frac{\lambda_2}{\lambda_1}[s - g_1 p]. \end{cases}$$

when $k_2 = 1$, then player 1 chooses risky,safe or is indifferent between them according as his value in the (p,v) plane lying above, below or on the line

$$D_1 : v = s + \frac{\lambda_2}{\lambda_1}[s - g_1 p]$$

If $k_2 = 0$, player 1 chooses the risky arm as long as his value is greater than s. He smoothly switches from R to S at $\bar{p_1}$. Since player 1 switches to S at $\bar{p_1}$ smoothly, we will have $v_1^{'}(\bar{p_1}) = 0$.Also $v_1(\bar{p_1}) = s$. From the bellman equation we would then have

$$\lambda_1 p(g_1 - s) = rs - rg_1 p$$

$$\Rightarrow \bar{p_1} = \frac{rs}{\lambda_1(\frac{r}{\lambda_1} g_1 + g_1 - s)}$$

$$\Rightarrow \bar{p_1} = \frac{\mu_1 s}{(\mu_1 + 1)g_1 - s)}$$

where $\mu_1 = \frac{r}{\lambda_1}$.

Similarly if $k_1 = 1$, player 2 chooses risky, safe or is indifferent between them according as his value in the (p,v) plane lying above, below or on the line

$$D_2 : v = s + \frac{\lambda_1}{\lambda_2}[s - g_2 p]$$

If $k_1 = 0$, player 2 switches to the safe arm from the risky arm smoothly at $\bar{p_2}$ where

$$\bar{p_2} = \frac{\mu_2 s}{(\mu_2 + 1)g_2 - s}$$

Payoffs: Before we discuss equilibrium formally, we obtain explicit solutions for the payoffs obtained by the players. We adopt following notations:

v_i^{rr} : payoff to player i when he chooses risky and the other player also chooses risky

v_i^{rs} : payoff to player i when he chooses risky and the other player chooses safe

F_i : Payoff to player i when the other player chooses the risky arm and he free rides by choosing the safe arm

We have

$$v_i^{rr}(p) = g_i p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

$$v_i^{rs}(p) = g_i p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_i}}$$

$$F_1(p) = s + \frac{\lambda_2}{\lambda_2 + r}[g_1 - s]p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

$$F_2(p) = s + \frac{\lambda_1}{\lambda_1 + r}[g_2 - s]p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

Diversification Equilibrium:

Suppose there exists a Markov perfect equilibrium as follows:

Player 2 chooses risky arm for $p > p_2$ and safe arm for $p \leq p_2$. Player 1 chooses risky arm for $p > p_1$ and safe for $p \leq p_1$.

$$0 < p_1 < p_2 < 1$$

We would construct an equilibrium as described above and in the course of our construction we would determine the condition under which such an equilibrium actually exists.

Proposition 1 *If λ_2 is sufficiently low with respect to λ_1, then we have an equilibrium as above and it is the unique Markov perfect equilibrium. This equilibrium is inefficient and*

involves free riding.

Proof. In any noncooperative equilibrium no experimentation will occur for beliefs less than or equal to $\bar{p_1}$. Also in any equilibrium at the right $\varepsilon-$ neighborhood of $\bar{p_1}$ only player 1 will be experimenting and player 2 will be free riding. This is because if player 2 were the only player choosing the risky arm, he would do so for beliefs greater than $\bar{p_2}$. Hence player 2 will never choose the risky arm for beliefs less than or equal to $\bar{p_2}$.

We will now work backwards from $\bar{p_1}$. In any equilibrium, at the right $\varepsilon-$ neighborhood of $\bar{p_1}$ player 1 chooses the risky arm and player 2 free rides. This is because if player is the only one to choose the risky arm then he smoothly switches to the safe arm at the belief $\bar{p_1}$. Thus payoff of 1 will be given by v_1^{rs} and that of 2 will be given by $F_2()$. Since the value functions are continuous we will have

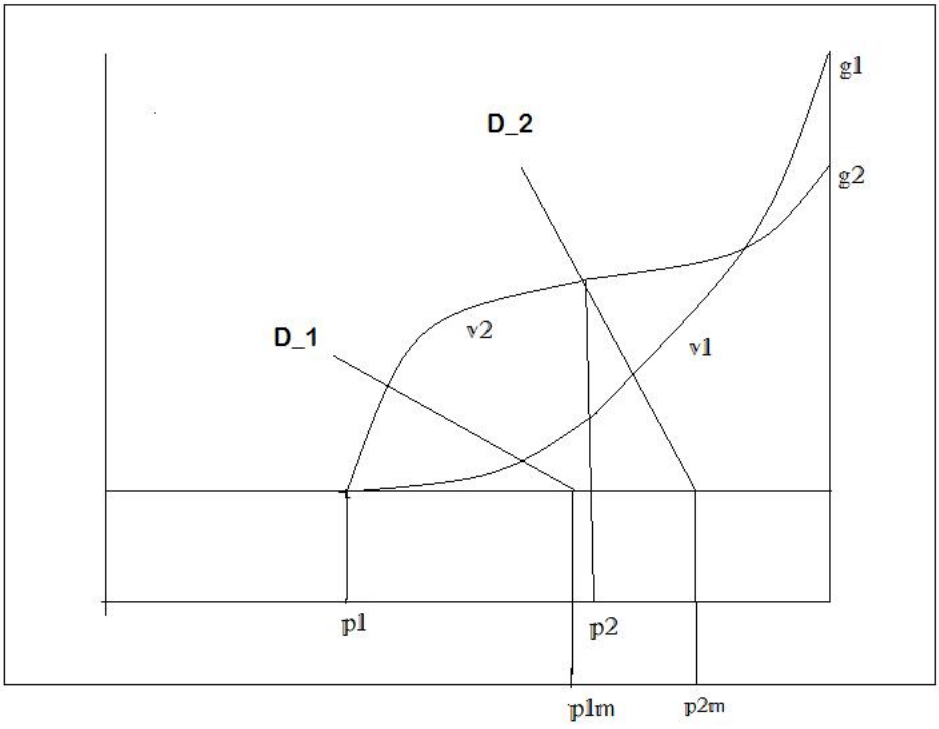
$$v_1^{rs}(\bar{p_1}) = g_1\bar{p_1} + C(1-\bar{p_1})[\Lambda(\bar{p_1})]^{\frac{r}{\lambda_1}} = s \Rightarrow C = \frac{s - g_1 p}{(1-\bar{p_1})[\Lambda(\bar{p_1})]^{\frac{r}{\lambda_1}}}$$

and

$$F_2(\bar{p_1}) = s + \frac{\lambda_1}{\lambda_1 + r}[g_2 - s]\bar{p_1} + C(1-\bar{p_1})[\Lambda(\bar{p_1})]^{\frac{r}{\lambda_1}} = s \Rightarrow C = -\frac{\frac{\lambda_1}{\lambda_1+r}[g_2 - s]\bar{p_1}}{(1-\bar{p_1})[\Lambda(\bar{p_1})]^{\frac{r}{\lambda_1}}}$$

The integration constant for v_1^{rs} is positive and thus it is strictly convex. The slope of v_1 at $\bar{p_1}$ is 0. Hence v_1^{rs} is strictly increasing for $p > \bar{p_1}$. On the other hand, the integration constant of F_2 is negative and thus it is strictly concave. At $\bar{p_1}$, the slope of F_2 is strictly positive. Hence at the right $\varepsilon-$ neighborhood of $\bar{p_1}$, F_2 will lie above v_1^{rs}.

Player 2 will stop free riding as soon as in the (p, v) plane, the payoff function of 2 crosses the line D_2 . This is demonstrated in the diagram below.



Since $g_1 = \lambda_1 h$ and $g_2 = \lambda_2 h$ we have

$$D_1 : v = s + \frac{\lambda_2}{\lambda_1}(s - g_1 p) = s + \frac{\lambda_2}{\lambda_1} s - g_2 p; D_2 : v = s + \frac{\lambda_1}{\lambda_2}(s - g_2 p) = s + \frac{\lambda_1}{\lambda_2} s - g_1 p$$

Hence D_1 has a negative slope of magnitude g_2 and D_2 has a negative slope of magnitude g_1. Since $g_1 > g_2$, D_1 is flatter than D_2. D_1 intersects the horizontal line $v = s$ at $p1m = \frac{s}{g_1}$ and D_2 intersects at $p2m = \frac{s}{g_2}$.

The upper curve $v2$ depicts the payoff of player 2 and the lower curve $v1$ depicts the payoff of player 1. Suppose the payoff curve of player 2 intersects the line D_2 at $p = p_2$. This is determined from

$$F_2(p_2) = s + \frac{\lambda_1}{\lambda_2}(s - g_2 p_2)$$

We have $F_2(\bar{p}_1) = s < s + \frac{\lambda_1}{\lambda_2}(s - g_2\bar{p}_1)$ and $F_2(1) = s + \frac{\lambda_1}{\lambda_1+r}[g_2 - s] > s + \frac{\lambda_1}{\lambda_2}(s - g_2)$. Since F_2 is monotonically increasing and D_2 is monotonically decreasing, there exists a unique $p_2 \in (\bar{p}_1, 1)$ such that $F_2(p_2) = s + \frac{\lambda_1}{\lambda_2}(s - g_2 p_2)$.

Next, observe that given λ_1, a decrease in λ_2 results in an increase in p_2. To see this consider a p_2 for particular given values of λ_1 and λ_2. As λ_2 decreases, the slope of F_2 at $\bar{p}_1$ goes down and since F_2 is strictly concave, at the present p_2, F_2 becomes lower. Thus L.H.S decreases. On the other hand, R.H.S increases. Given the nature of F_2 and D_2, to restore equality the new p_2 should be higher.

Now a diversification equilibrium as conjectured will exist if at p_2 the payoff curve of player 1 lies above the line D_1. Referring to the diagram above this implies that the lower curve should intersect D_1 before the upper curve intersects D_2.

The payoff curve of player 1 lies above the line D_1 when

$$v_1^{rs}(p_2) > s + \frac{\lambda_2}{\lambda_1}s - g_2 p_2 \Rightarrow \frac{\lambda_2}{\lambda_1} < \frac{g_2 p_2}{s} + \frac{v_1^{rs}}{s} - 1$$

Since $p_2 < \frac{s}{g_2}$ and $v_1^{rs}(p_2) > s$, $\frac{g_2 p_2}{s} + \frac{v_1^{rs}}{s} - 1 < 1$. Given a λ_1, we can always choose λ_2 so that the above condition is satisfied. This is because p_2 increases with the decrease of λ_2. Thus a diversification equilibrium exists only when the degree of heterogenity is high enough.

The other integration constants are determined as follows:

$$C \text{ for } v_1^{rr} \text{ from } v_1^{rr}(p_2) = v_1^{rs}(p_2)$$

$$C \text{ for } v_2^{rr} \text{ from } v_2^{rr}(p_2) = F_2(p_2) = s + \frac{\lambda_1}{\lambda_2}[s - g_2 p_2]$$

This concludes the proof.

■

The above result can be intuitively understood from the diagram. In a diversification equilibrium player 1 should never free ride. Given λ_1, if λ_2 decreases then the line D_1 becomes flatter. This reduces the range of beliefs over which player 1 would have wanted to free ride if 2 was choosing the risky arm. The area under the line D_1 represents the opportunities to free ride for player 1.

The diversification equilibrium is inefficient. The inefficiency arises from two channels. First, no experimentation takes place for beliefs below $\bar{p}_1$, whereas the planner would have wanted experimentation up to $p = p_1^* < \bar{p}_1$. Clearly player 1 does not internalise the benefit

to player 2 from his experimentation. Secondly, player 2 inefficiently free rides for some range of beliefs. At p_2, player 2's private return is equal to the private cost $s - g_2 p_2$. However the social benefit is higher, since player 2 does not internalise the benefit to player 1 from his experimentation. Thus $p_2^* < p_2$ and there is inefficient free riding for $p \in (p_2^*, p_2)$. We call this inefficient free riding as the planner makes player 2 to free ride over some range of beliefs as part of the efficient solution.

Welfare Comparison: Homogeneity and Heterogeneity

One natural question to ask is whether with heterogeneous players we would have relatively more experimentation in the non-cooperative equilibrium than that in a model with homogeneous players. To make a meaningful comparison of this sort we first define an index to compare the extent of experimentation in two models of strategic experimentation. This index is defined below

Index of Experimentation: Suppose there are two models of strategic experimentation. In both the models suppose players start with a common prior p_0. Let p_1^c be the belief where conditional on no breakthrough all experimentation ceases in model 1 and let p_2^c be the belief where conditional on no breakthrough all experimentation ceases in model 2. Without loss of generality let $p_1^c > p_2^c$ .Then the index η is defined as:

$$\eta = [\frac{p_0 - p_1^c}{\tau_1}] - [\frac{p_0 - p_1^c}{\tau_2^{p_1^c}} + \frac{p_1^c - p_2^c}{\tau_2^{p_2^c}}]$$

τ_1 is the time required to reach to the belief p_1^c starting from p_0 in model 1. $\tau_2^{p_1^c}$ is the time required to reach p_1^c, starting from p_0 and $\tau_2^{p_2^c}$ is the time required to reach p_2^c from p_1^c in model 2.

η positive(negative) indicates that the extent of experimentation is more in model 1 (2) is more. Note that this index of comparison captures both the range of beliefs over which experimentation takes place and also the total intensity with which the experimentation has taken place(reflected in τ.). In particular if in two models experimentation has taken place for the same range of beliefs but slower in one of them then the extent of experimentation should be lower for that model. This is because slower experimentation means lower intensity of experimentation. Similarly suppose in a model we have experimentation over large range of beliefs but very slowly and in another model we have experimentation over a smaller range of beliefs but at a much faster pace. Thus it is possible to have the extent

of experimentation to be lower in the first case.

The following proposition now states that there exist situations where with heterogeous players we can acheive more experimentation in noncooperative equilibrium than in a model with homogeneous players, keeping the sum of the Poisson intensities of the players constant.

Proposition 2 *Suppose there are two players with Poisson intensities λ_1 and λ_2 respectively ($\lambda_1 > \lambda_2$). Let $\lambda = \lambda_1 + \lambda_2$. Then there exists a λ_2^* such that for all $\lambda_2 \in (0, \lambda_2^*)$, keeping the summation of λ_1 and λ_2 constant, the extent of experimentation in the noncooperative equilibrium is more than that in any noncooperative equilibrium in a model of strategic experimentation with each player having a Poisson intensity of $\frac{\lambda}{2}$.*

Proof.

First, from the previous proposition we can infer that keeping the sumation of λ_1 and λ_2 constant, we can find a $\lambda_2^{'}$ such that for all $\lambda_2 \in (0, \lambda_2^{'})$, there is a unique diversification equilibrium in Markovian Strategies.

Next, from Keller et.al (2005) we know that in any equilibrium of the noncooperative game of the model with each player having a Poisson intensity of $\frac{\lambda}{2}$, all experimentation ceases at the belief

$$p_c^1 = \frac{\mu_{hom} s}{(\mu_{hom} + 1)(\frac{g}{2} - s) + \mu_{hom} s}$$

where $\mu_{hom} = \frac{r}{\frac{\lambda}{2}}$, $g = \lambda h$.

With heterogeneous players all experimentation ceases at the belief $\bar{p_1}$, the monopoly cutoff belief for the player with Poisson intensity λ_1. Since $\lambda_1 > \frac{\lambda}{2}$, we have $p_c^1 > \bar{p_1}$.

Consider a $\lambda_2 \in (0, \lambda_2^{'})$. We know that there exists a unique equilibrium with thresholds p_2 and $\bar{p_1}$ ($p_2 > \bar{p_1}$) such that player 2 switches to the safe arm at p_2 and player 1 at $\bar{p_1}$.

In any non cooperative equilibrium of the model with homogeneous players, there exists ranges of beliefs when only one of the players is experimenting. Hence we must have

$$\tau_1 > \frac{1}{\lambda}[\log[\Lambda(p_c^1)] - \log[\Lambda(p_0)]]$$

where $\frac{1}{\lambda}[\log[\Lambda(p_c^1)] - \log[\Lambda(p_0)]]$ is the time required to reach p_c^1 from the prior p_0 had all players been experimenting. Hence this is the time required to reach the belief p_c^1 from the prior p_0 if the players were controlled by a social planner. τ_1 is the time required by the players in an arbitrary noncooperative equilibrium of a model with homogeneous players.

Next, suppose $p_2 > p_c^1$. Then the time required to reach p_c^1 from p_0 in the noncooperative game with heterogeneous players is

$$\tau_2^{p_c^1} = \frac{1}{\lambda_1}[\log[\Lambda(p_c^1)] - \log[\Lambda(p_2)]] + \frac{1}{\lambda}[\log[\Lambda(p_2)] - \log[\Lambda(p_0)]]$$

This is becuase for the range $(p_2, p_0]$ both players are experimenting and for the range (p_c^1, p_2) only player 1 is experimenting.

Now as $\lambda_2 \to 0$ (keeping λ constant), $\tau_2^{p_c^1} \to \frac{1}{\lambda}[\log[\Lambda(p_c^1)] - \log[\Lambda(p_0)]]$ from below. Thus there exists $\lambda_2^* \leq \lambda_2^{'}$ such that for $\lambda_2 \in (0, \lambda_2^*)$ we have

$$\tau_1 > \tau_2^{p_c^1} > \frac{1}{\lambda}[\log[\Lambda(p_c^1)] - \log[\Lambda(p_0)]]$$

Since $\frac{p_c^1 - \bar{p_1}}{\tau_2^{\bar{p_1}}} > 0$, we have

$$\eta = \frac{p_0 - p_c^1}{\tau_1} - [\frac{p_0 - p_c^1}{\tau_2^{p_c^1}} + \frac{p_c^1 - \bar{p_1}}{\tau_2^{\bar{p_1}}}] < 0$$

If $p_2 < p_c^1$, then for any λ_2 we have $\tau_2^{p_c^1} = \frac{1}{\lambda}[\log[\Lambda(p_c^1)] - \log[\Lambda(p_0)]$. Thus η will be negative.

Hence the extent of experimentation with heterogeneous players is more. This concludes the proof. ■

The economic intuition behind the above result is as follows. With homogeneous players, both players free ride on each other. However when we introduce heterogeneity, player 1 never free rides on the other player in the noncooperative equilibrium. Player 2 however has an increased opportunities for free riding. In the figure the area under the line D_2 represents the opportunities to free ride for player 2. As λ_2 goes down, λ_1 increases since the summation remains constant. This makes the line D_2 steeper and in turn increases the opportunities for player 2 to free ride. However since the intensity of player 2's experimentation is low and in the noncooperative equilibrium only player 2 free rides we have more experimentation in the game with heterogeneous players.

3 Introducing Payoff Externalities

Two players face a common continuous time two-armed bandit. Each of the arms is accessible by both the players. The bandit is of exponential type. One of the arms is *safe*(*S*) and the other arm is *risky*(*R*). A player who activates the safe arm, gets arrival according to a Poisson process with intensity $\pi_0 > 0$. A risky arm can either be good or bad. If player i activates a *good* risky arm, then he experiences arrival according to a Poisson process with intensity π_i, such that

$$\pi_1 \geq \pi_2 > \pi_0 > 0$$

No arrivals are experienced along a bad risky arm.

Players start with a common prior p^0, which is the probability with which the risky arm is good. Players observe each other's actions and the arrivals experienced by them. Hence at each time point t, players share a common posterior belief p_t. Only the first player to experience an arrival gets a payoff of 1 unit. We start with the case when the players are homogeneous, i.e $\pi_1 = \pi_2$.

3.1 Symmetric Players

In this subsection, we lay out the analysis with homogeneous players. Thus players' ability to learn along the risky arm is the same. They both experience arrivals at the good risky arm according to a Poisson process with intensity $\pi_1 > \pi_0$.

We start our analysis with the benchmark case, i.e the social planner's problem.

3.1.1 Social Planner's problem: The efficient benchmark

Consider the problem of a benevolent social planner who wants to maximise the sum of expected discounted payoff of the players. Hence at each instant, based on p, he allocates each of the players to activate one of the arms. k_t denotes the action profile chosen by the planner at the instant t. $k_t \in \{0,1,2\}$. k_t denotes the number of players made to activate the risky arm at the instant t. $k_t(t \geq 0)$ is such that it is measurable with respect to the information available at time t

It is assumed that if the planner is indifferent between making a player to activate the risky arm or the safe arm, then he makes him to activate the safe arm. Thus the planner's action is left continuous.

From now on we will do away with the time subscript. Let $v(p)$ be the value function of the planner. Since actions are left continuous and beliefs can move only in the left direction, left continuity of $v(p)$ can always be assumed.

Then $v(p)$ should satisfy,

$$v(p) = \max_{k\in\{0,1,2\}} \{(2-k)\pi_0\,dt + kp\pi_1\,dt +$$

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

since $(v(p+dp) = v(p) + v^{'}(p)\,dp)$ and $dp = -k\pi_1 p(1-p)\,dt$.

After expanding and rearranging the above and ignoring the terms of order $o(dt)$ we have

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

Proposition 3 *The planner's optimality involves making both the players to activate the risky arm as long as $p > p^*$, where $p^* = \frac{\pi_0}{\pi_1}$. For $p \leq p^*$, both are made to activate the safe arm.*

Proof. Since (1) is linear in k, we know that at the optimum, k will either be 2 or 0. When both players are optimally made to activate the risky arm, the value function satisfies:

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

where $\Lambda(p) = \frac{1-p}{p}$ and C is the integration constant. This is derived by solving the O.D.E obtained by putting $k = 2$ in (1).

When both players are optimally made to activate the safe arm, then $v = \frac{2\pi_0}{r+2\pi_0}$. Since $v(p)$ satisfies the *value matching* and *smooth pasting* conditions at $p = p^*$, we get

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

This concludes the proof. ■

3.1.2 The non-cooperative game

Player i chooses actions $\{k_{it} \in \{0,1\}\}$, such that k_{it} is measurable with respect to the information available at time t. We restrict our attention to Markovian strategies, such that

strategy of player i is defined by the mapping $k_i : [0,1] \to \{0,1\}$. We allow only those k_i functions which satisfy the property that $k_i^{-1}(1)$ and $k_i^{-1}(0)$ are disjoint unions of a finite number of non-degenerate sub-intervals in $[0,1]$, such that $k_i(0) = 0$ and $k_i(1) = 1$. This ensures that the game is well-defined in the continuous time framework.

Players simultaneously update their belief about the risky arm to be good as long as there is at least one player activating the risky arm and there is no arrival(at any of the arms). Both k_1 and k_2 are left continuous, which guarantee the existence of a well defined law of motion of the posterior.

Let v_i be the value function (equilibrium payoff) of player i ($i = 1,2$)in the non-cooperative game. If (k_1, k_2) is an equilibrium strategy profile then given k_j ($j = 1,2$), k_i ($i = 1,2; i \neq j$) and v_i should satisfy

$$v_i = \max_{k_i \in \{0,1\}} \{(1-k_i)\pi_0\, dt + k_i \pi_1 p\, dt +$$

$$(1 - r\, dt)(1 - \pi_0\, dt(2 - k_i - k_j) - p\pi_1(k_i + k_j)\, dt)(v_i - v_i' p(1-p)\pi_1(k_i + k_j)\, dt)\}$$

Simplifying the above, we obtain

$$rv_i = \max_{k_i \in \{0,1\}} \{(1-k_i)\pi_0(1 - v_i) + k_i(\pi_1 p[1 - v_i - v_i' p(1-p)])$$

$$- (1-k_j)\pi_0 v_i - k_j \pi_1 p(v_i + (1-p)v_i')\} \tag{2}$$

Proposition 4 *There exists an efficient equilibrium.*

Proof. Consider the following strategy profile: *Each player activates R for $p > p^*$ and S for $p \leq p^*$* (Hence p^* is the switching point). This is a symmetric strategy profile and the outcome implied by this profile is the efficient outcome. We need to show that this profile constitutes an equilibrium.

Suppose player 2 follows the above strategy. We will determine the best response of player 1. It is clear that for $p = 1$, player 1 will choose R. Thus the optimal switching point of player 1 is to be determined. It is shown in the supplemental appendix that the unique optimal switching point for player 1 is p^*. Similarly, this can be shown for player 2.

This concludes the proof. ■

This is an interesting point to note. From [11] we know that with homogeneous players, efficient equilibrium in threshold type strategies never exists. Here we observe that just by introducing payoff externalities, we can obtain efficient equilibrium in threshold

type strategies. Hence we see that competition among players brings in efficiency which intuitively makes sense.

Next, we move on to our analysis with heterogeneous players. We find that the nature of distortion in the non-cooperative game with respect to the benchmark case (social planner's problem) is exactly the same as obtained in the previous environment.

3.2 Heterogeneous Players

Consider the setting where players are heterogeneous, i.e their ability to learn across the risky arm is different. Hence we have $\pi_1 > \pi_2 > \pi_0$.

To start with, as before, we first analyse the social planner's problem which is intended to be the efficient benchmark.

3.2.1 The Social Planner's problem

The planner's objective is the same as before. Let (k_1, k_2) be his action profile. $k_i \in \{0,1\}$, for $i = 1,2$. $k_i = 1(0)$ implies that the planner has made the ith player to activate risky(safe) arm. Let $v(p)$ be the value function of the planner. Then it should satisfy

$$v(p) = \max_{k_i \in \{0,1\}} \{(2-k_1-k_2)\pi_o\, dt + k_1 p\pi_1\, dt + k_2 p\pi_2\, dt +$$

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

$$\Rightarrow rv = \max_{k_i \in \{0,1\}} \{(2-k_1-k_2)\pi_0[1-v] + k_1(p\pi_1[1-v-v^{'}(1-p)]) + k_2(p\pi_2[1-v-v^{'}(1-p)])\} \tag{3}$$

This is because $v(p+dp) = v(p) + v^{'}(p)\,dp$ and $dp = -(k_1\pi_1 + k_2\pi_2)p(1-p)\,dt$.

The following lemma establishes a property for an interior solution of the planner's problem.

Lemma 1 *If there exists an interior solution (i.e there exists $p_i^* \in (0,1)$ such that for $p > p_i^*$ player i is made to activate R and for $p \leq p_i^*$, player i is made to activate S) then optimality requires diversification over a range of beliefs. That is, there exists a range of beliefs over which the planner will make one player to activate the risky arm and the other player to activate the safe arm.*

Proof of Lemma. Suppose not. This implies that the planner's optimality requires him to switch both the player from the risky arm to the safe arm at the same p, say $p^{'}$. At the

optimum the smooth pasting condition must hold which implies that $v^{'}(p^{'}) = 0$. From (3), we know that optimality requires,

$$p^{'}\pi_2[1-v] = p^{'}\pi_1[1-v(p^{'})] = \pi_0[1-v(p^{'})]$$

However since $\pi_1 > \pi_2$, $p^{'}\pi_2[1-v(p)] < p^{'}\pi_1[1-v(p^{'})]$. This is a contradiction.

This proves the lemma. ■

The next lemma shows that if the planner's solution involves diversification, then player 2 is to be switched to the safe arm at a higher belief than the one at which player 1 is switched.

Lemma 2 *Player* 2 *is to be switched to the safe arm from the risky arm at a higher p than player* 1.

Proof of Lemma. Suppose not. From lemma (1) we know that this implies player 1 is switched to the safe arm at a higher p than player 2. Let this switching point be p_1^*. From (3), we know that at p_1^* we must have, $\pi_0[1-v(p_1^*)] = p_1^*\pi_1[1-v(p_1^*)-v^{'}(p_1^*)(1-p_1^*)]$. Since $\pi_2 < \pi_1$, we have $\pi_0[1-v(p_1^*)] = p_1^*\pi_1[1-v(p_1^*)-v^{'}(p_1^*)(1-p_1^*)] > p_1^*\pi_2[1-v(p_1^*)-v^{'}(p_1^*)(1-p_1^*)]$. This is a contradiction to the claim that it is optimal to keep player 2 at the risky arm at $p = p_1^*$. This proves the lemma. ■

With the help of the above two lemmas we are now in a position to describe the planner's solution. The following proposition does this.

Proposition 5 *There exists a solution to the planner's problem, where both the players are made to activate the risky arm for $p > p_2^*$, player* 2 *is made to activate the safe arm and* 1 *to activate the risky arm for $p \in (p_1^*, p_2^*]$, and both players are made to activate the safe arm for $p \leq p_1^*$ where $p_1^* = \frac{\pi_0}{\pi_1}$.*

Proof. First, assume that there exists some $\frac{\pi_0}{\pi_1} < p_2^* < 1$, such that it is optimal to switch player 2 to the safe arm at p_2^*. $v(p)$ in the range of beliefs over which 2 is made to activate the safe arm and 1 is made to activate the risky arm, should satisfy

$$v = \frac{\pi_0}{r+\pi_0} + \frac{r\pi_1 p}{(r+\pi_0)(r+\pi_0+\pi_1)} + C_2(1-p)[\Lambda(p)]^{\frac{r+\pi_0}{\pi_1}} \equiv v_{SR}$$

This is derived through solving the O.D.E obtained by putting $k_2 = 0$ and $k_1 = 1$ in (3). Suppose p_1^* is the belief where 1 is to be switched to the safe arm. Since at p_1^*, both players

are activating S, optimality would require to have $v^{'}(p_1^*) = 0$(smooth pasting condition). According to lemma (2), player 2 is switched from R to S at a higher p. Then from the value matching condition, we know that we should have $v_{SR}(p_1^*) = v(p_1^*)$. This gives us $C_2 = \frac{\frac{r\pi_0}{(r+\pi_0)(r+2\pi_0)} - \frac{r\pi_1 p_1^*}{(r+\pi_0)(r+\pi_0+\pi_1)}}{(1-p_1^*)[\Lambda(p_1^*)]^{\frac{r+\pi_0}{\pi_1}}}$. Observe that $C_2 > 0$. Also, the smooth pasting condition at p_1^* implies $v_{SR}^{'}(p_1^*) = 0$. This gives us

$$\frac{r\pi_1}{(r+\pi_0)(r+\pi_0+\pi_1)} - C_2[\Lambda(p_1^*)]^{\frac{r+\pi_0}{\pi_1}}[1+\frac{(r+\pi_0)}{\pi_1 p_1^*}] = 0 \Rightarrow p_1^* = \frac{\pi_0}{\pi_1}$$

We now need to prove the existence of a $p_2^* \in (p_1^*, 1)$, such that at p_2^*, the planner finds it optimal to switch player 2 from R to S. The existence of such a p_2^* is proved in the supplemental appendix.

This concludes the proof of the proposition. ■

Corollary 1 *$p_2^* > \frac{\pi_0}{\pi_2}$, the threshold p where the planner would have switched player 2 from R to S had he been dealing with this player only.*

Proof. Suppose not. Then $p_2^* \leq \frac{\pi_0}{\pi_2}$. At p_2^*, $v^{'}(p_2^*) = v_{SR}^{'}(p_2^*) > 0$. Since v is strictly convex for $p > \frac{\pi_0}{\pi_1}$, $v^{'}(\frac{\pi_0}{\pi_2}) > 0$. Therefore at $p = \frac{\pi_0}{\pi_2}$, $\pi_0[1-v] > \pi_2 p[1-v-v^{'}(1-p)]$. From (3), we can see that this contradicts the claim that $p_2^* \leq \frac{\pi_0}{\pi_2}$. This proves the corollary. ■

3.2.2 The non-cooperative game

This is similar to the non-cooperative game with homogeneous players. Thus $k_1(.)$ and $k_2(.)$ are the Markovian strategies of the players.

Let $v_1(p)$ and $v_2(p)$ be the payoff functions of players 1 and 2 respectively in a Markovian equilibrium. v_i along with k_i should then satisfy

$$rv_i = \max_{k_i \in \{0,1\}} \{(1-k_i)[\pi_0(1-v_i)] + k_i[\pi_i p(1-v_i-v_i^{'}(1-p))] - [(1-k_j)\pi_0 v_i + k_j p(v_i + v^{'}(1-p))]\} \tag{4}$$

This implies that given k_j, at any p optimality on player i's part requires choosing $k_i(p) = 0(1)$ if $[\pi_0(1-v_i)] \geq (<)[\pi_i p(1-v_i-v_i^{'}(1-p))]$.

We determine the non-cooperative equilibrium in following steps.

Lemma 3 *Suppose player 2 follows the strategy of activating R for $p > p_2^{*N}$ and S for $p \leq p_2^{*N}$ such that $\frac{\pi_0}{\pi_1} < p_2^{*N} < 1$. Then player 1's best response is to activate R for $p > p_1^*$ and S for $p \leq p_1^{*N}$ where $p_1^* = \frac{\pi_0}{\pi_1}$.*

Proof of Lemma. First, consider the range $p \leq p_2^{*N}$. If $k_1 = 1$ ($k_2 = 0$ by hypothesis), then by putting $i = 1$ in (4) we know that v_1 should solve

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

This is a first order O.D.E. Solving this we have,

$$v_1 = \frac{\pi_1}{r+\pi_0+\pi_1} p + C(1-p)[\Lambda(p)]^{\frac{r+\pi_0}{\pi_1}} \equiv v_1^{RS}(p) \tag{5}$$

where C is an integration constant. If he choose $k_1 = 0$ then $v_1(p)$ should satisfy,

$$v_1 = \frac{\pi_0}{r+2\pi_0} \tag{6}$$

Initially, we assume that player 1 indeed behaves in the way as claimed, for $p \leq p_2^{*N}$. Later, we will show that the value function thus obtained for the specified range will satisfy the Bellman equation for this range. This is shown in the supplemental appendix.

Next, consider the range $p > p_2^{*N}$. As before we conjecture that it is optimal for 1 to choose $k_1 = 1$ and derive the value function. Then we show that the obtained value function indeed satisfy the bellman equation. Again, this is shown in the supplemental appendix.

This concludes the proof. ■

Lemma 4 *Suppose player* 1 *plays the following strategy: Activate R for* $p > p_1^{*N} = \frac{\pi_0}{\pi_1}$ *and Activate S for* $p \leq p_1^{*N}$. *Then there exists a* $p_2^{*N} \in (p_1^{*N}, \frac{\pi_0}{\pi_2})$, *such that player* 2*'s best response is to activate R for* $p > p_2^{*N}$ *and activate S for* $p \leq p_2^{*N}$.

Proof of Lemma. Consider $p \leq p_1^{*N}$. First, as before we conjecture that it is optimal for player 2 to be at S. Then $v_2 = \frac{\pi_0}{r+2\pi_0}$ for $p \leq p_1^{*N}$. From (4) one can conclude that $\pi_0(1-v_2) > \pi_2 p[1-v_2-v_2^{'}(1-p)]$ for $p \leq p_1^{*N}$. This supports our conjecture.

Now consider the optimal stopping problem of player 2 in the range $[p_1^{*N}, 1]$, given player 1's strategy. This is done in the supplemental appendix, which shows the existence of a unique $p_2^{*N} \in (p_1^{*N}, 1)$.

From (4), we know that at the optimal we shall have $[\pi_2 p_2^{*N}(1-v_2(p_2^{*N}) - v_2^{'}(p_2^{*N})(1-p_2^{*N}))] = \pi_0(1-v_2(p_2^{*N}))$. Since $[1-v_2(p_2^{*N})] < [1-v_2(p_2^{*N}) - v_2^{'}(p_2^{*N})(1-p_2^{*N})]$, we have $p_2^{*N} < \frac{\pi_0}{\pi_2}$. ■

The above two lemmas now allow us to formally state the non-cooperative equilibrium. The following proposition describes this.

Proposition 6 *Player* 1 *activating R (S) for* $p > (\leq) p_1^{*N}$ *and player* 2 *activating R (S)for* $p > (\leq) p_2^{*N}$ *constitutes a unique Markovian equilibrium in threshold type strategies.*

Proof. The proof of this proposition follows directly from lemma (3) and (4). ■

The above proposition describes the unique equilibrium in threshold type Markovian strategies. Since $p_2^{*N} < \frac{\pi_0}{\pi_2} < p_2^*$, there exists a range of beliefs (p_2^{*N}, p_2^*) when efficiency requires player 2 to switch to the safe arm, but it does not. This shows, that the non-cooperative equilibrium outcome involves the phenomenon of too-much duplication.

4 Conclusion

We have demonstrated that when the players are heterogeneous with respect to learn across the risky arm(s), then efficiency requires diversification, i.e each player to experiment along an exclusive arm. Keeping the total ability constant, if the degree of heterogeneity is high enough then the extent of experimentation in the non-cooperative equilibrium is more than that in any non-cooperative equilibrium with homogeneous players. When payoff externalities is introduced then with heterogeneous players we have inefficiency in equilibrium in form of too much of duplication.

References

[1] Akcigit, U., Liu, Q., 2011: "The Role of Information in Competitive Experimentation. ", *mimeo, Columbia University and University of Pennsylvania*.

[2] d'Aspremeont, C., Bhattacharya, S., Gerard-Varet, L., 2000 "Bargaining and Sharing Innovative knowledge. ", *The Review of Economic Studies* 67, 255 – 271.

[3] Bhattacharya, S., Mookerjee D., 1986 "Portfolio choice in research and development. ", *Rand Journal of Economics* 17, 594 – 605.

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

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

[6] Dasgupta, P., Maskin, E., 1987: "The Simple Economics of Research Portfolios ", *The Economic Journal* 581 – 595

[7] Dasgupta, P., Stiglitz, J., 1980 "Uncertainty, Industrial Structure and the Speed of R&D ", *Bell Journal of Economics* 111 – 28

[8] "No end to Dementia ", *The Economist, June* 2010

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

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

[11] Keller, G., Rady, S., Cripps, M., 2005: "Strategic Experimentation with Exponential Bandits ", *Econometrica* 73, 39 – 68.

[12] Keller, G., Rady, S., 2010:"Strategic Experimentation with Poisson Bandits ", *Theoretical Economics* 5, 275 – 311.

[13] Klein, N., 2011: "Strategic Learning in Teams ", *mimeo University of Bonn*

[14] Klein, N., Rady, S., 2011: "Negatively Correlated Bandits ", *The Review of Economic Studies* 78 693 – 792.

[15] Lee, T., Wilde, L., 1980: "Market Structure and Innovation: A Reformulation", *Quarterly Journal of Economics* 94 429 – 436

[16] Loury,G.C., 1979 "Market Structure and Innovation ", *Quarterly Journal of Economics* 93395 – 410.

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

[18] Reinganum, J. 1982 "A dynamic Game of R&D Patent Protection and Competitive Behavior ", *Econometrica* 50 671 – 688.

[19] Scherer, F.M., "International High-Technology Competition ", *Cambridge, Mass:, Harvard University Press*

[20] Stokey,N.L., 2009: "The Economics of Inaction ", *Princeton University Press*.

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