UNIVERSITY OF EXETER

Economics Department Discussion Papers Series

ISSN 1473 – 3307

The Role of Heterogeneity in a Model of Strategic Experimentation

Kaustav Das

Paper number 15/07

The Role of Heterogeneity in a model of Strategic Experimentation [*]

Kaustav Das[†]

May 2, 2015

Abstract

In this paper, I examine the effect of introducing heterogeneity between players in a model of strategic experimentation. I consider a two-armed bandit problem in continuous time 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. At a time point, a player can choose only one of the arms. I show that if the degree of heterogeneity between the players is high enough, then there exists a unique Markov perfect equilibrium in simple cut-off strategies. The non-cooperative equilibrium in the heterogeneous model in terms of welfare, always gets a higher rank than any non-cooperative equilibrium of a homogeneous players model with same or more amount of experimentation in the benchmark.

JEL Classification Numbers:C73, D83, O31.

Keywords: Two-armed Bandit, Free-Riding, Learning

[*]. A very preliminary version of the model was available under the title *Strategic Experimentation with heterogeneous agents and payoff externalities*. My greatest intellectual debt is to Kalyan Chatterjee for his helpful suggestions and for providing me with continuous encouragement. I also thank Dieter Balkenborg, Martin Cripps, Rajiv Sarin and Jusso Valimaki for their helpful comments. A significant idea of this paper was developed while I was visiting the University of Bonn. A special note of thanks to Sven Rady for the helpful and valuable discussions with him. Thanks are also due to the participants of the ESEM 2014, Toulouse, NASM 2014, Minnesotta. 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 informational externalities and players are heterogeneous.

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 making 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 over to 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. Informational externalities arise in these models 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. In this paper, I study a variant of the standard exponential bandit model with two arms by introducing heterogeneity between the players. This means, along a particular arm, players differ with respect to their innate abilities. Hence, given that a reward occurs along this arm, the expected time required to get that reward differs among players. I show that with heterogeneous agents and only informational externalities, there is a unique Markov perfect equilibrium in simple cut-off strategies, provided the degree of heterogeneity between the players is sufficiently high. This non-cooperative equilibrium always involves free-riding. However, I show that welfare wise, the model with two heterogeneous players always do better than a model with homogeneous players where each player's ability to learn is equal to the average ability to learn in the heterogeneous players model and the benchmark amount of experimentation is the same or more.

The analysis starts with introducing heterogeneity in the now canonical form of Two-armed Bandit Model (*a.la* Keller, Rady and Cripps). Each player faces a common two armed exponential bandit in continuous time. 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 $g_1 > g_2 > s$. At a time point, a player can choose only one of the arms.

We first examine 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 the risky arm is good (in this setting this implies belief being close to 1), then both the players are made to access the risky arm. For interim beliefs, the weaker player (player 2) is allocated to the safe arm and the stronger player (player 1) is allocated to the risky arm. Lastly, if it is very likely that the risky arm is bad(implying belief being close to 0) then both players are made to access the safe arm.

For the analysis of the 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. I show that if the degree of heterogeneity is high enough then there exists a unique inefficient diversification equilibrium. This means that when players differ with respect to each other to a large extent, then there exists a Markov perfect equilibrium where players use simple cut-off strategies and this equilibrium is unique. The belief at which all experimentation ceases is greater than that in the optimal solution of the planner's problem. This is due to the fact that player 1 does not internalise the benefit to player 2 from his experimentation. 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 non cooperative equilibrium in a model with two heterogeneous players to that in a model with homogeneous players. In the homogeneous players model, we restrict ourselves to the class of equilibria where players switch strategies at finite number of points. We know that if the degree of heterogeneity is high enough, then in the model with two heterogeneous players, we have a unique equilibrium in simple cut off strategies. I compare this model with a homogeneous players model where each player's Poisson intensity is equal to the average Poisson intensity of the heterogeneous players model and the benchmark amount of experimentation is at least the same as in the heterogeneous players model. I show that the amount of experimentation in the unique equilibrium of the heterogeneous players model is always higher than that in any equilibrium in simple strategies of the homogeneous players model. Hence, welfare wise the heterogeneous players model does better.

In real world, there are many instances where agents have alternative potential approaches to pursue and along one of the approaches players might differ with respect to their abilities to generate payoffs. Consider a situation in the academic world. Suppose a doctoral student is jointly supervised by two faculty members. They enter into an agreement that any research output of this student during doctoral studies will have both the supervisors as co-authors. Suppose each of the supervisors is expert in their respective sub-fields and they can guarantee a steady flow of papers if research is conducted on that sub-field. In addition to this, there is an interesting and more challenging problem which requires knowledge of both these sub-fields. A priori it is not known whether this challenging problem has a solution or not. However, if it could be solved, then it would result in a more prestigious publication than the papers in each of the sub-fields. This situation can be visualised as a strategic experimentation problem in two armed bandits. Each of the sub-fields can be interpreted as a safe arm, and the challenging problem can be seen as a risky arm. In this situation, each of the supervisors has to make a choice of whether to conduct research along the safe arm or the risky arm. Notice that here each supervisor can free ride on the other. If one of them is conducting research on the challenging problem, then a success will also give a payoff to the other supervisor. In this scenario, it is interesting to analyse if it is better to have supervisors who differ with respect to their abilities in solving the challenging problem or to have supervisors with equal ability. In this paper, I show that having supervisors who differ in their ability to solve the challenging problem results in higher amount of research on it.

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 ([2]) Keller,Rady and Cripps([4]), Keller and Rady([5]), Klein and Rady ([7]) and Thomas([9]). In all of these papers except ([9]) and ([7]), players have replicas of bandits and *Free-riding* is a common feature in all the above models except ([9]). This leads to an inefficient level (too little) of experimentation. The present work contributes in two ways. First, I show the effect of heterogeneity and find that unlike in a model with homogeneous players, for certain range of parameters, there exists a unique Markov perfect equilibrium in simple cut-off strategies. The amount of experimentation in this equilibrium is more than that in any Markov perfect equilibrium (by restricting ourselves to the class of equilibria where players switch action at finite number of points) of a two armed bandit game with homogeneous players.

Thomas([9]) 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 Poisson arrival rates differs across the exclusive arm. The present paper differs from this in the way that here Poisson arrival rates along the same risky arm differs across players. Further, we do not have congestion along any of the arms.

Klein([6])) studies a model where each player has an access to a bandit with two risky arms and one safe arm. He shows that there exists efficient equilibrium if the stakes are high enough. In the present paper, I show that even in a bandit model with a safe arm and a risky arm, heterogeneity between the players can give rise to unique Markov perfect equilibrium.

The rest of the paper is organised as follows. Section 2 lays down the detail of the setting with heterogeneous players and the condition under which a unique Markov perfect equilibrium in simple cut-off strategies exists. Section 3 describes the welfare comparison to a model with homogeneous players and finally section 4 concludes the paper.

2 Two armed bandit model with heterogeneous players

The Model:

There are two players (1 and 2) and each of them 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 the risky arm is good, then a player accessing 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 $h > 0$. Player 1 experiences these arrivals according to a Poisson process with intensity $\lambda_1 > 0$ and player 2 experiences these according to a Poisson process with intensity $\lambda_2 > 0$ 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 $g_2 = \lambda_2 h$. We have $g_1 > g_2 > s$. The uncertainty in this model arises from the fact that it is not known whether the risky arm is good or bad. Players start with a common prior p_0, which is the probability with which the risky arm is good. A player in continuous time has to decide whether to choose the safe arm or the risky arm. At a time point, a player can choose only one arm. Players' actions and outcomes are publicly observable and based on these, they update their beliefs. Players discount the future according to a common continuous time discount rate $r > 0$.

To describe this formally, let p_t be the common belief at time $t \geq 0$. The belief evolves according to the history of experimentation and payoffs. Since players start with a common prior and the actions and outcomes of players are publicly observable, we will always have a common belief at all times $t > 0$. Player i ($i \in \{1,2\}$) chooses a stochastic process $\{k_i(t)\}_{(t \geq 0)}$. This stochastic process is measurable with respect to the information available up to time t with $k_i(t) \in \{0,1\}$ for all t. $k_i(t) = 1(0)$ implies that the player has chosen the risky arm (safe arm). Each player's objective is to maximise his total expected discounted payoff, which is given by

$$E\{\int_{t=0}^{\infty} re^{-rt}[(1-k_i(t))s + (k_i(t)p_t)g_i]\,dt\}$$

The expectation is taken with respect to the processes $\{k_i(t)\}_{t \in R^+}$ and $\{p_t\}_{t \in R^+}$. From the objective function it can be seen that there does not exist any payoff externalities between the players. The effect of the presence of the other player is only via the effect on the belief through the informations generated by his experimentation.

Evolution of beliefs:

In the present model, only a good risky arm can yield a positive payoff in form of lump sums. This implies that the breakthroughs are completely revealing. Hence, if any player experiences a lump sum in a risky arm at time $t = \tau \geq 0$, then $p_t = 1$ for all $t > \tau$. On the other hand, suppose at the time point $t = \tau$, $p_t \in (0,1)$ and no player achieves any breakthrough till the time point $\tau + \Delta$ where $\Delta > 0$. Using Bayes' Rule, the posterior at the time point $t = \tau + \Delta$ is

$$p_{\tau+\Delta} = \frac{p_\tau e^{-\int_{\tau}^{\tau+\Delta}[\lambda_1 k_1(t) + \lambda_2 k_2(t)]\,dt}}{p_\tau e^{-\int_{\tau}^{\tau+\Delta}[\lambda_1 k_1(t) + \lambda_2 k_2(t)]\,dt} + (1-p_\tau)}$$

Since beliefs evolve in continuous time, conditional on no breakthrough, the process $\{p_t\}_{t \in R^+}$ will evolve according to the following law of motion

$$dp_t = -(\lambda_1 k_1(t) + \lambda_2 k_2(t))p_t(1-p_t)\,dt$$

In the following subsection, we consider the benchmark case when the actions of both players are controlled by a benevolent social planner.

2.1 Planner's Problem

Suppose there is a benevolent social planner, who controls the actions of both the players. Let $(k_1(p_t), k_2(p_t))$ be the action profile of the planner, such that $k_i \in \{0,1\}$. $k_i = 0$ implies that 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 $v(p)$ is the value function of the planner, then using the law of motion of the beliefs we must have

$$v = \max_{k_1,k_2 \in \{0,1\}} [r\{(1-k_1)s + (1-k_2)s + k_1 p g_1 + k_2 p g_2\} dt$$

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

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

$$v = 2s + \max_{k_1,k_2 \in \{0,1\}} \big\{k_1[b_1(p,v) - c_1(p)] + k_2[b_2(p,v) - c_2(p)]\big\}$$

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

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

We can interpret the term $b_i(p)$ as the benefit of having player i on the risky arm when the current state is p. On the other hand, the term $c_i(p)$ can be interpreted as the opportunity cost of having player i on the risky arm. Note that this bellman equation is linear in both k_1 and k_2. In the following proposition, we state the planner's solution.

Proposition 1 *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^*.*

Proof. We first assume that a solution as proposed exists and evaluate the value function. Then we show that the obtained value function satisfies optimality.

Since the Bellman equation is linear in the choice variables k_1 and k_2, we can restrict to corner solutions and can thus derive closed form solutions for the value function.

First, consider the range $p \in (0, p_1^*]$. According to the conjectured solution, $k_2 = k_1 = 0$. This would then imply that $v(p) = 2s$. Next, consider the range $p \in (p_1^*, p_2^*]$. The conjectured solution implies that $k_1 = 1$ and $k_2 = 0$. Thus, from the bellman equation we can infer that the planner's value function satisfies the following O.D.E:

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

The solution to the above differential equation is:

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

where $g = (g_1 + g_2)$; $\Lambda(p) = \frac{(1-p)}{p}$ and C is the integration constant.

Suppose p_1^* is the belief where player 1 is switched to the safe arm. From the value matching 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 at p_1^* requires that both the right hand and left hand derivative of v at p_1^* is zero. This implies

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

Substituting the value of 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 $\mu_1 = \frac{r}{\lambda_1}$.

Next, consider $p > p_2^*$. The planner finds it optimal to keep both players at the risky arm. Thus, $k_1 = k_2 = 1$. This implies that for $p \geq p_2^*$, the value function then satisfies the following O.D.E

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

The solution to the above O.D.E is

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

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

At $p=p_2^*$, player 2 is switched to the safe arm. Since the value function is continuous, at the belief p_2^*, the planner is indifferent between having player 2 at the risky arm or at the safe arm. Thus, at $p=p_2^*$, we have

$$b_2(p,v)=s-g_2p$$

Smooth pasting condition at $p=p_2^*$ implies that for $p\geq p_2^*$, we have

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

Hence $b_2(p_2^*,v)$ can be written as

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

Since, $b_2(p_2^*,v)=s-g_2p_2^*$, we have

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

This is because $\lambda_1>\lambda_2$. Let $v_{sr}(.)$ be the representation of the value function when 1 is at the risky arm and 2 is at the safe arm and v_{rr} be the same when both players are at the risky arm. Imposing the value matching condition at $p=p_2^*$ gives us

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

From this, we can infer that p_2^* should satisfy

$$[\frac{\lambda_1 g + rg_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p_2^* + [\frac{s - [\frac{\lambda_1 g + rg_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p_1^*}{(1-p_1^*)[\Lambda(p_1^*)]^{\frac{r}{\lambda_1}}}](1-p_2^*)[\Lambda(p_2^*)]^{\frac{r}{\lambda_1}} = \frac{\lambda_1}{\lambda_2}s \qquad (1)$$

We will now show that there actually exists a $p_2^* \in (p_1^*, 1)$ such that the above relation holds. At $p = p_1^*$, L.H.S of (1) is equal to $s < \frac{\lambda_1}{\lambda_2}s$. At $p = 1$, the L.H.S 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 (1) holds.

The integration constant of v_{rr} is given by

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

The obtained value function is

$$v(p) = \begin{cases} gp + \{\frac{\frac{\lambda_1+\lambda_2}{\lambda_2}s - gp_2^*}{(1-p_2^*)[\Lambda(p_2^*)]^{\frac{r}{\lambda}}}\}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}} \equiv v_{rr} & : \quad \text{If } p \in (p_2^*, 1], \\ & : \\ s + [\frac{\lambda_1 g + rg_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p + \{\frac{s - [\frac{\lambda_1 g + rg_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p}{(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}}\}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}} \equiv v_{sr} & : \quad \text{if } p \in (p_1^*, p_2^*], \\ & : \\ 2s & : \quad \text{if } p \in (0, p_1^*]. \end{cases}$$

with $v_{rr}(p_2^*) = v_{sr}(p_2^*) = \frac{\lambda}{\lambda_2}s$ and $v_{sr}(p_1^*) = 2s$.

By standard verification arguments, it can be shown that this value function satisfies optimality. This is shown in appendix A ■

The planner's solution is depicted in the Figure 1.

The planner's value function is a smooth convex curve and it lies in the range $[2s, g)$. At the belief $p_2^*(p_1^*)$, player 2 (1) is switched to the safe arm from the risky arm.

The next subsection describes the non-cooperative game between the players.

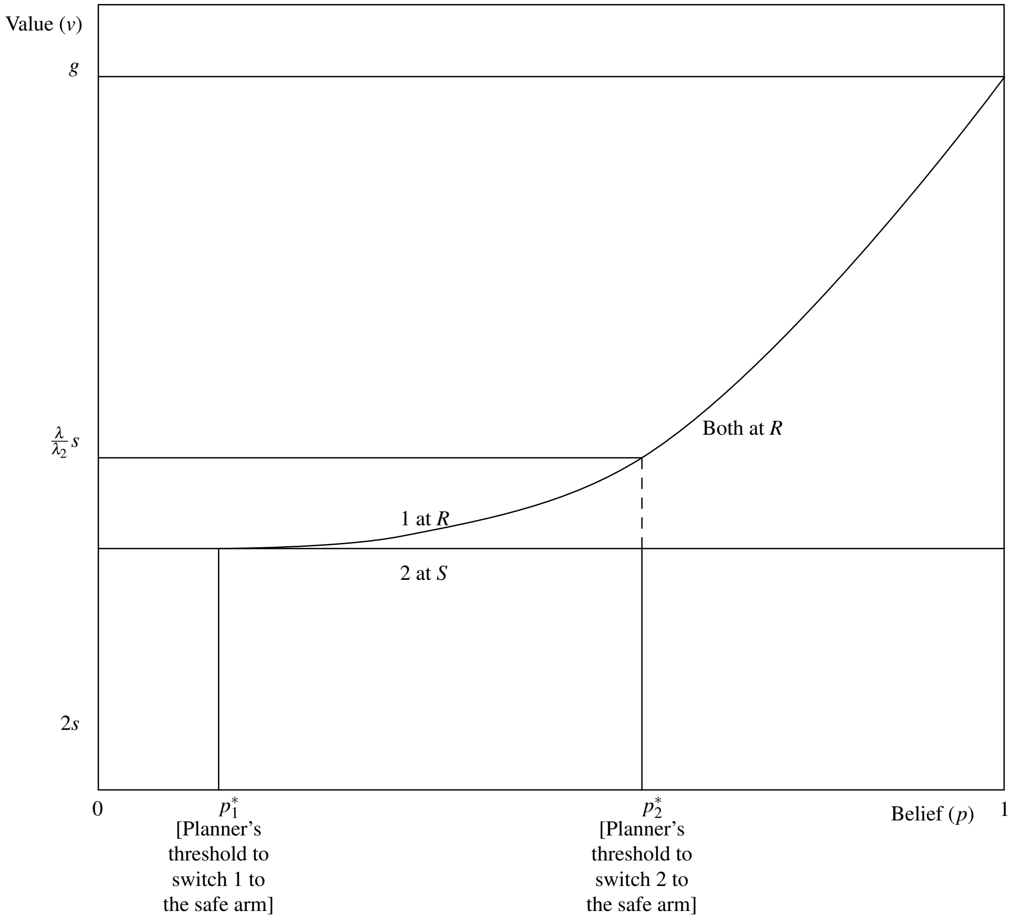

Figure 1.

2.2 Non-cooperative game

In this subsection, we carry out the analysis of the non-cooperative game between the players. We would focus on Markov perfect equilibria with the players' common posterior belief as the state variable. A Markov strategy of player i is any piecewise continuous function $k_i : [0,1] \rightarrow \{0,1\}$ ($i = 1,2$). This function is continuous at all but a finite number of points. Further, we have $k_i(0) = 0$ and $k_i(1) = 1$. This ensures that player i chooses the dominant action under subjective certainty.

We assume that the strategies of players are left continuous. Suppose at a time point $t \geq 0$, the common prior is p_t. Then, given a strategy pair $(k_1(p_t), k_2(p_t))$ and conditional on there being no breakthrough, from our previous arguments we know that the common posterior beliefs evolve in continuous time according to the following law of motion

$$dp_t = -(\lambda_1 k_1(p_t) + \lambda_2 k_2(p_t))p_t(1-p_t)\,dt$$

Given these, we first discuss the best responses of the players.

Best Responses:

Let v_1 be the optimal value function of player 1. Then given player 2's strategy, and by the principle of optimality, v_1 should satisfy

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

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

After ignoring the terms of the order $(o(dt))$ and rearranging the remaining terms, we have

$$v_1(p) = s + k_2[\lambda_2 b_1(p,v_1)] + \max_{k_1 \in \{0,1\}} k_1[\lambda_1 b_1(p,v_1) - (s - g_1 p)] \qquad (2)$$

where

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

$\lambda_1 b_1(p,v_1)$ can be interpreted as the additional payoff accrued to player 1 due to the information generated from his own experimentation. $b_1(p,v_1)$ is the expected discounted benefit to player 1 from the risky arm if information is generated according to a Poisson process with intensity 1. This $b_1(p,v_1)$ has two parts. The first part, $p[g_1 - v_1]$, represents the expected value of the jump in the event a breakthrough occurs. On the other hand, if no breakthrough occurs, there is a revision of belief in the downward direction which results in a negative effect on the overall payoff. This aspect is captured by the part $-p(1-p)v_1^{'}$. Since player 1's experimentation along the risky arm generates information according to a Poisson process with intensity λ_1 , $\lambda_1 b_1(p,v_1)$ is the additional payoff obtained by player 1 by choosing the risky arm. Similarly, $\lambda_2 b_1(p,v_1)$ is the additional payoff to player 1 from player 2's experimentation along the risky arm. $s - g_1(p)$ is player 1's opportunity cost of choosing the risky arm.

If v_2 is the optimal value function of player 2, then given k_1, we have

$$v_2(p) = s + k_1[\lambda_1 b_2(p,v_2)] + \max_{k_2 \in \{0,1\}} k_1[\lambda_2 b_2(p,v_2) - (s - g_2 p)] \tag{3}$$

where

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

In the same as above, we can explain the terms $\lambda_2 b_2(p,v_2)$, $\lambda_1 b_2(p,v_2)$ and $s - g_2 p$.

For a given $k_2 \in \{0,1\}$, from (2) we know that player 1's best response is

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

Thus, player 1 chooses the risky arm as long as his private additional benefit from using it (given by $\lambda_1 b_1(p,v_1)$) is greater than or equal to the opportunity cost of choosing the risky arm (given by $s - g_1 p$). The term $k_2[\lambda_2 b_1(p,v_1)]$ reflects the free-riding opportunities for player 1.

By rearranging we can infer that

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

This implies that when $k_2 = 1$, player 1 chooses the risky arm, safe arm 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 optimal 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 since player1's value function is continuous, we have $v_1(\bar{p_1}) = s$. Putting these in the optimal equation of player 1 (2), we 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, for player 2, from (3) we have

$$k_2 = \begin{cases} 1 & : \text{ if } v_2 > s + k_1 \frac{\lambda_1}{\lambda_2}[s - g_2 p], \\ \in \{0, 1\} & : \text{ if } v_2 = s + k_1 \frac{\lambda_1}{\lambda_2}[s - g_2 p], \\ 0 & : \text{ if } v_2 < s + k_1 \frac{\lambda_1}{\lambda_2}[s - g_2 p]. \end{cases}$$

This implies that 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

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

When the other player uses the risky arm, the best response of the players are depicted in figure 2.

The region lying below the line D_1 represents the free-riding opportunities for player 1 while that lying below the line D_2 represents the free-riding opportunities for player 2. Line D_2 is steeper than the line D_1. From the picture, we can see that the range of beliefs over which player 2 can free-ride is larger than the range over which player 1 can free-ride. There exists a region which lies above the line D_1 and below the line D_2. This gives rise to the possibility of equilibria with players choosing different arms over a range of beliefs.

Payoffs: Before we discuss equilibrium formally, we obtain explicit solutions for the payoffs obtained by the players under different possibilities.

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

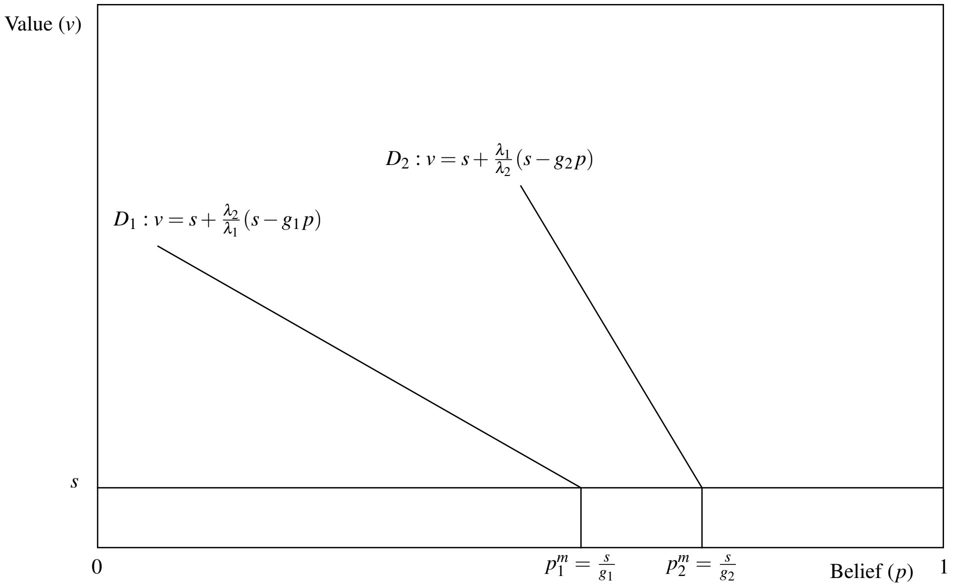

Figure 2.

chooses the risky arm . v_i^{rr} satisfies the ODE

$$v_i^{'} + v_i \frac{[r + (\lambda_1 + \lambda_2)p]}{(\lambda_1 + \lambda_2)p(1-p)} = \frac{(\lambda_1 + \lambda_2) + r}{(\lambda_1 + \lambda_2)(1-p)} g_i \tag{4}$$

This is obtained by putting $k_1 = k_2 = 1$ in the optamility equation of player i. Since v_i^{rr} is a solution to the above ODE, it can be expressed as

$$v_i^{rr} = g_i p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda}} \tag{5}$$

v_i^{rs} : payoff to player i when he chooses the risky arm and the other player chooses the safe arm. Putting $k_i = 1$ and $k_j = 0 (j \neq i)$ in the optimality equation of player i, we get the ODE which v_i^{rs} should satisfy as

$$v_i^{'} + v_i \frac{[r + \lambda_i p]}{\lambda_i p(1-p)} = \frac{\lambda_i + r}{\lambda_i(1-p)} g_i \tag{6}$$

Thus, v_i^{rs} can be expressed as

$$v_i^{rs}(p) = g_i p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_i}} \tag{7}$$

Finally, let the payoff to player i when the other player chooses the risky arm and he free rides by choosing the safe arm be denoted by F_i. Putting $k_i = 0$ and $k_j = 1$ ($j \neq i$) in the optimality equation of player i, we get the ODE satisfied by F_i. This is given by

$$v_i^{'} + \frac{r+\lambda_j p}{\lambda_j p(1-p)} = \frac{rs}{\lambda_j p(1-p)} + \frac{g_i}{(1-p)} \tag{8}$$

Solving the above ODE, we can posit that F_i can be expressed as

$$F_i(p) = s + \frac{\lambda_j}{\lambda_j + r}[g_i - s]p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_j}} \tag{9}$$

C in all cases is the integration constants and $\Lambda(p) = \frac{1-p}{p}$.

We will now show that no efficient equilibrium exists. The following proposition describes this.

Proposition 2 *The planner's solution can never be implemented in a markov perfect equilibrium*

Proof. First, we argue that in any non-cooperative equilibrium, no experimentation along the risky arm will occur for beliefs strictly less than $\bar{p_1}$. Suppose it does. Then let $p_l < \bar{p_1}$ be the lowest belief where experimentation along the risky arm ceases. Then, consider player i who is experimenting at this belief. There can be two possibilities. Either the other player ($j \neq i$) is also experimenting along the risky arm at this belief or player i is the only one experimenting. Since no experimentation occurs for beliefs strictly less than p_l and value functions of players are continuous, at p_l, $v_i = s$. As $p_l < \frac{s}{g_i}$, in the first case player i's payoff would lie below the line D_i and hence, he is not playing his best response. In the later case, since $p_l < \bar{p_i}$, player i is again not playing his best response.

Thus no experimentation will ever occur for beliefs less than $\bar{p_1}$. However, in the planner's solution experimentation occurs till the belief reaches the point p_1^* and $p_1^* < \bar{p_1}$. This proves the proposition. ∎

Having proved that all markov perfect equilibria are inefficient, we now investigate whether we are able to obtain equilibria which is qualitatively similar to the planner's solution.

Diversification Equilibrium:

It is worthwhile to explore if there exists a non-cooperative equilibrium which is qualitatively similar to the planner's solution. We would show that if the degree of heterogeneity between the players is high enough, then such an equilibrium exists and is unique. This is illustrated in the following proposition.

Proposition 3 *If λ_2 is sufficiently low with respect to λ_1, then there exists a unique Markov perfect equilibrium in simple cutoff strategies with thresholds $\bar{p}_1$ and p_2^{*n}. $\bar{p}_1$ is as defined above and p_2^{*n} is such that $p_2^{*n} \in (\bar{p}_1, 1)$ and $p_2^{*n} > p_2^*$. For $p \in (p_2^{*n}, 1)$, both players choose the risky arm, for $p \in (\bar{p}_1, p_2^{*n}]$, player 1 chooses the risky and 2 chooses the safe arm and for $p \leq \bar{p}_1$, both players choose the safe arm.*

Proof. Before formally proving the existence and uniqueness of this equilibrium, let us try to understand it intuitively from figure 3.

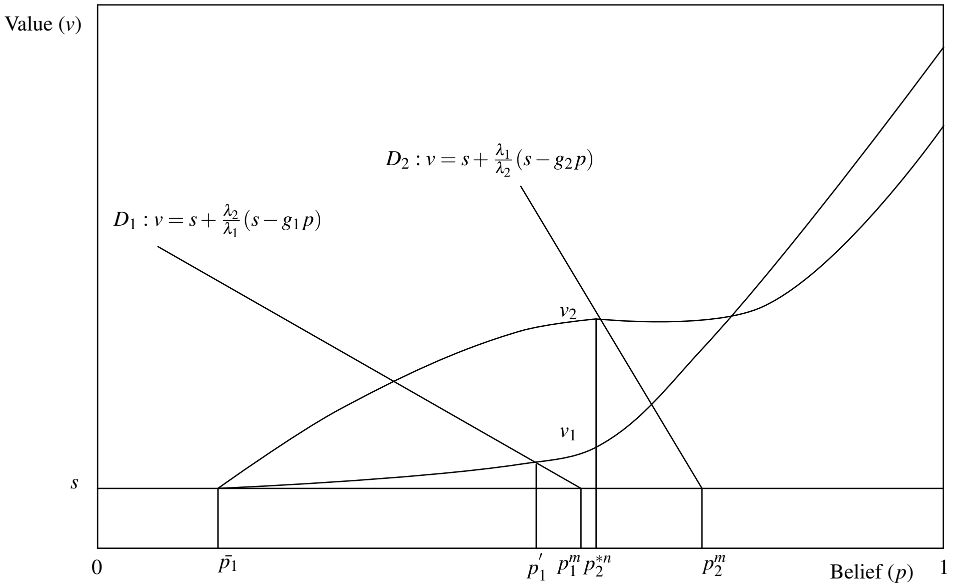

Figure 3.

As before, line D_i $(i = 1,2)$, describes the free-riding opportunities for player i. 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 $p = p_1^m = \frac{s}{g_1}$ and D_2 intersects at $p = p_2^m = \frac{s}{g_2}$.

The upper curve v_2 depicts the payoff of player 2 and the lower curve v_1 depicts the payoff of player 1. For all beliefs less than or equal to $\bar{p}_1$, both players choose the safe arm. At the right neighborhood of $\bar{p}_1$, only player 1 experiments along the risky arm and player 2 free rides. Hence, the payoff curve of player 1 is strictly convex and that of player 2 is strictly concave. At the belief $\bar{p_1}$, the derivative of the payoff of player 1 is zero and that of player 2 is strictly positive. Thus, at $p = \bar{p_1}$, v_2 lies strictly above v_1. v_2 intersects the line D_2 at $p = p_2^{*n}$. At this point, player 2 stops free-riding and starts choosing the risky arm as well. Hence, the curve v_2 now becomes convex and there is a kink in v_1 at this point. p_2^{*n} is strictly greater than p_2^*, the belief upto which the planner would have wanted player 2 to experiment along the risky arm.

We now formally establish this equilibrium. As argued above, in any noncooperative equilibrium, no experimentation along the risky arm will occur for beliefs less than or equal to $\bar{p_1}$. We can now work backwards from $\bar{p_1}$.

First, I show that in any equilibrium, at the right $\varepsilon-$ neighborhood ($\varepsilon \to 0$) of $\bar{p_1}$, only player 1 will be experimenting along the risky arm and player 2 will be free riding.

Suppose, at the right ε- neighborhood of $\bar{p_1}$, both players experiment along the risky arm. Since the value functions are continuous, both will have their values close to s. In the (v, p) plane, $(s, \bar{p_1})$ lies below both the lines D_1 and D_2. Hence, none of the players are playing their best responses. This shows that in any non-cooperative equilibrium, only one player can experiment along the risky arm at the right ε-neighborhood of $\bar{p_1}$. It is not possible to have player 2 experimenting along the risky arm and player 1 choosing the safe arm. This is because if player 1 chooses the safe arm, choosing the risky arm constitutes a best response for player 2 only if $p \geq \bar{p_2} > \bar{p_1}$. Hence, the only possibility is to have player 1 experimenting along the risky arm and player 2 choosing the safe arm. This constitute playing best responses by both the players. Thus, in any non-cooperative equilibrium, for beliefs at the right ε- neighborhood of $\bar{p_1}$, Player 1 chooses the risky arm and 2 chooses the safe arm. At $p = \bar{p_1}$, player 1 smoothly switches to the safe arm. Hence, payoffs for player 1 and 2 for this range of beliefs will be given by v_1^{rs} and F_2 respectively. Since the value functions are continuous, we will have

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

and

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

This is the manifestation of the value matching conditions at $p = \bar{p_1}$. 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}.

With player 1 choosing the risky arm, choosing safe arm will be a best response of player 2, as long as F_2 lies left of D_2.

We will now show, that there exists a unique $p_2^{*n} \in (\bar{p_1}, 1)$ such that $F_2(p_2^{*n}) = s + \frac{\lambda_1}{\lambda_2}(s - g_2 p_2^{*n}) \equiv D_2(p_2^{*n})$. That is, there exists a unique belief in the range $(\bar{p_1}, 1)$ where F_2 meets the line D_2.

We have $F_2(\bar{p_1}) = s < s + \frac{\lambda_1}{\lambda_2}(s - g_2 \bar{p_1}) \equiv D_2(\bar{p_1})$, since $\bar{p_1} < p_1^m$. On the other hand, $F_2(1) = s + \frac{\lambda_1}{\lambda_1 + r}[g_2 - s] > s + \frac{\lambda_1}{\lambda_2}(s - g_2)$ as $g_2 > s$. Since F_2 is monotonically increasing[1] and D_2 is monotonically decreasing in p, there exists a unique $p_2^{*n} \in (\bar{p_1}, 1)$, such that $F_2(p_2^{*n}) = s + \frac{\lambda_1}{\lambda_2}(s - g_2 p_2^{*n}) \equiv D_2(p_2^{*n})$.

Let the belief at which v_1^{rs} meets D_1 be denoted as $p_1^{'}$. For all $p \in (\bar{p_1}, p_2^{*n})$, player 1 choosing the risky arm and player 2 choosing the safe arm are best responses to each other. The conjectured equilibrium exists if both players choosing the risky arm for $p > p_2^{*n}$ are best responses to each other. This happens only if $p_1^{'} < p_2^{*n}$. Thus, v_1^{rs} should meet D_1 at a belief which is strictly lower than the belief at which $F_2(.)$ meets D_2. This is because, if $p_1^{'} > p_2^{*n}$, then for $p \in (p_2^{*n}, p_1^{'})$, choosing the risky arm is not a best response of player 1 when the other player is choosing the risky arm.

However, for this equilibrium to be unique, we need to ensure that there does not exist any range of beliefs such that player 1 choosing the safe arm and player 2 choosing the risky arm are best responses to each other. This requires the belief at which the curve v_1^{rs} meets the line D_1 to be lower than $\bar{p_2}$. Thus, we require $p_1^{'} < \bar{p_2}$. If the belief at which v_1^{rs} meets D_1 is higher than $\bar{p_2}$, then there would exist a range of beliefs where player 1 choosing the safe arm and player 2 choosing the risky arm would be best responses to each other. It will be established below, that $p_1^{'} < \bar{p_2}$, only if the degree of heterogeneity is high enough. Further, $p_1^{'} < \bar{p_2}$ also guarantees existence of the equilibrium.

Consider λ_2 very close to λ_1. That is, λ_2 is such that $\lambda_1 - \lambda_2 > 0$ and $(\lambda_1 - \lambda_2) \to 0$. In this case, $\bar{p_2} \to \bar{p_1}$ from above. Since, v_1^{rs} is independent of λ_2, the belief at which it will meet D_1 will be strictly higher than $\bar{p_2}$.

Next, keeping λ_1 fixed, consider λ_2 close to $\frac{s}{h}$. That is $\lambda_2 - \frac{s}{h} > 0$ and $\lambda_2 \to \frac{s}{h}$ from above. In this case, $\bar{p_2} \to 1$. Thus, the belief at which v_1^{rs} meets the line D_1 is strictly less

[1] This is because $F_2^{'} = \frac{\lambda_1}{\lambda_1 + r}[g_2 - s] - C[\Lambda(p)]^{\frac{r}{\lambda_1}}$. For $p = 1$, $F_2^{'} = \frac{\lambda_1}{\lambda_1 + r}[g_2 - s] > 0$. Since $F_2^{'}(\bar{p_1}) > 0$ and is strictly concave, $F_2^{'} > 0$ for $p \in (\bar{p_1}, 1)$

than $\bar{p_2}$.

Keeping λ_1 constant, as λ_2 goes down, the line D_1 becomes flatter and pivots downward along the point $(\frac{s}{g_1}, s)$. Thus, the belief at which v_1^{rs} meets D_1 goes down. Hence, the belief at which v_1^{rs} meets D_1 is monotonically increasing in λ_2. On the other hand, $\bar{p_2}$ is monotonically decreasing in λ_2. This implies that there exists a $\lambda_2^* \in (\frac{s}{h}, \lambda_1)$, such that if $\lambda_2 < \lambda_2^*$, then v_1^{rs} always meets D_1 at a belief strictly less than $\bar{p_2}$.

Lastly, we show that $p_2^{*n} > \bar{p_2}$. Since $F_2()$ is strictly increasing in p, to establish this formally, we need to show that

$$D_2(\bar{p_2}) > F_2(\bar{p_2})$$

As the integration constant of F_2 is strictly negative, we have $F_2(\bar{p_2}) < s + \frac{\lambda_1}{\lambda_1+r}[g_2 - s]\bar{p_2}$. From the expression of $\bar{p_2}$, we then have

$$F_2(\bar{p_2}) < s + \frac{\lambda_1}{\lambda_1 + r}[g_2 - s]\bar{p_2} = s + \frac{\lambda_1}{\lambda_1 + r}[g_2 - s]\frac{\mu_2 s}{(\mu_2 + 1)g_2 - s} \equiv f$$

On the other hand, $D_2(\bar{p_2}) = s + \frac{\lambda_1}{\lambda_2} s[g_2 - s]\frac{1}{(\mu_2+1)g_2 - s}$. This implies

$$D_2(\bar{p_2}) - f = \frac{\lambda_1 s[g_2 - s]}{\{(\mu_2 + 1)g_2 - s\}\lambda_2(\lambda_1 + r)}\lambda_1 > 0$$

Hence, $D_2(\bar{p_2}) > F_2(\bar{p_2})$. This establishes the fact that $p_2^{*n} > \bar{p_2}$. Thus, whenever $p_1^{'} < \bar{p_2}$, $p_1^{'} < p_2^{*n}$.

This proves that if the difference $\lambda_1 - \lambda_2$ exceeds a threshold, then the conjectured equilibrium exists and is unique.

For beliefs greater than p_2^{*n}, payoff of player 1 and 2 are given by v_1^{rr} and v_2^{rr} respectively. The integration constants are determined as follows:

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

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

This concludes the proof. ■

Let us intuitively explain the above result. In a diversification equilibrium, player 1 should never free ride and there should be a range of beliefs over which 1's best response should be choosing R and 2's best response should be free-riding on 1's experimentation. Given λ_1, if λ_2 decreases then the line D_1 becomes flatter. This reduces the free-riding opportunities of player 1. Hence, the area between the two lines D_1 and D_2 increases. This explains why the degree of heterogeneity should be high enough for a diversification equilibrium to exist. For this equilibrium to be unique, there should not exist any range of beliefs where player 1 choosing the safe arm constitutes a best response to player 2 choosing the risky arm. This is ensured by having the degree of heterogeneity even higher. Keeping λ_1 fixed, as λ_2 goes down, the line D_1 becomes flatter and this reduces the free riding opportunities of player 1 and hence player 1 can never free-ride on player 2 in any non-cooperative equilibrium.

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^{*n}, 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[2]. Thus, $p_2^* < p_2^{*n}$ and there is inefficient free riding for $p \in (p_2^*, p_2^{*n})$. We call this *inefficient free riding* because the planner in his efficient solution, makes player 2 to free ride over some range of beliefs.

We conclude this section by making a note. When the extent of heterogeneity between the players is not large enough then there are two possibilities. First, an equilibrium in simple cut off strategies may not exist. In that case, like in Keller et.al (2005), we will have equilibria in simple strategies where players switch actions at finitely many points and players free ride on each other in turn. Secondly, if an equilibrium in simple cutoff strategies exist but $p_1^{'} > \bar{p_2}$, then there will be additional equilibria where players switch actions at finite number of points and free-rid eon each other in turn.

3 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, we first define what we mean by the amount of experimentation in a two-armed bandit model with heterogeneous players. In fact, the way we define the measure of experimentation below will turn out to be appli-

[2] Please refer to appendix (B) for a formal proof to show that this is true

cable for models of strategic experimentation with both homogeneous and heterogeneous players.

Suppose, we have a two armed bandit model with N players. Let λ_i be the Poisson intensity with which player i ($i = 1,2,...,N$) obtains breakthroughs along the good risky arm. Let $\bar{\lambda}$ be the average Poisson intensity of the players. Hence $\bar{\lambda} = \frac{\sum_{i=1}^{N} \lambda_i}{N}$. Then, we define the strength of each player i as $s_i = \frac{\lambda_i}{\bar{\lambda}}$. Intuitively, s_i is the relative strength of a player with respect to the average strength. Note that the total strength of all players combined is equal to N ($\sum_{i=1}^{N} s_i = N$). At a time point t, the total resource allocated to the risky arm is defined as

$$K_t = \sum_{i=1}^{N} I_{it} s_i$$

where I_{it} is the indicator variable denoting whether player i at time t is in the risky arm or not. Hence, K_t is basically the sum of the relative strengths of the players who are using the risky arm at time t. The total amount of experimentation performed up to time T is then given by

$$E = \int_{t=0}^{T} K_t \, dt$$

Interestingly, the total amount of experimentation depends only on the prior, the belief at which all experimentation stops and the average Poisson intensity of the players. The following lemma describes this

Lemma 1 *Let the common prior be p_0. Suppose there is no breakthrough along the risky arm, and at the belief p_c, all experimentation stops. Then, the amount of experimentation performed is given by $\frac{[\log(\Lambda(p_c)) - \log(\Lambda p_0)]}{\bar{\lambda}}$, where $\Lambda(p) = \frac{(1-p)}{p}$*

Proof. At time t, we have

$$dpt = -[\sum_{i=1}^{N} I_{it} \lambda_i] p_t (1-p_t) \, dt = -[K_t \bar{\lambda}] p_t (1-p_t) \, dt$$

as $\lambda_i = s_i \bar{\lambda}$. Hence the total amount of experimentation performed is given by

$$E = \int_{t=0}^{\infty} K_t \, dt = \int_{p=p_0}^{p=p_c} -\frac{1}{\bar{\lambda}} \frac{1}{p_t(1-p_t)} \, dp_t$$

$$= \frac{[\log(\Lambda(p_c)) - \log(\Lambda p_0)]}{\bar{\lambda}}$$

This concludes the proof of the lemma. ■

From Keller, Rady and Cripps (2005), we know that this is the amount of experimentation performed in a model with homogeneous players with each player having an intensity of $\bar{\lambda}$. In fact, the measure suggested here is also applicable to a model with homogeneous players. In that case, $s_i = 1\ \forall i$ and K_t is then the total number of players experimenting along the risky arm.

We now compare the heterogeneous players model with two players to a model with homogeneous players such that in the later the intensity of each player is equal to the average intensity of each player in the heterogenous players model and the benchmark amount of experimentation is at least as much as in the heterogeneous players model. Also, in the heterogeneous players model, the degree of heterogeneity is high enough to ensure that a diversification equilibrium exists and is unique. The following proposition summarizes the result.

Proposition 4 *Consider a two armed bandit model with two heterogeneous players with Poisson intensities λ_1 and λ_2. $\lambda_1 > \lambda_2 > \frac{s}{h}$. $(\lambda_1 - \lambda_2)$ is high enough so that there is a unique Markov perfect equilibrium. Also, consider a model with homogeneous players such that the benchmark amount of experimentation is at least as high as in the model with heterogeneous players and also the intensity of each player is equal to the average intensity of the players in the first model (i.e intensity of each player is $\lambda = \frac{\lambda_1+\lambda_2}{2}$). Then, the amount of experimentation in the Markov perfect equilibrium of the heterogeneous players model is always higher than any Markov perfect equilibrium in simple strategies of the game with homogeneous players. Thus, the heterogeneous players model does better than the homogeneous players model in terms of efficiency*

Proof. To prove this, first we show that if the amount of experimentation in a model with homogeneous players as described above is at least as large as that in the model with 2 heterogeneous players, then it must be the case that in the homogeneous players model there are at least three players. The following lemma describes this.

Lemma 2 *Consider a homogeneous players model with n players with each player's Poisson intensity being $\lambda = \frac{\lambda_1+\lambda_2}{2}$. If the benchmark amount of experimentation is the same as in the heterogeneous players model, then it must be the case that $n \geq 3$.*

Proof of Lemma. From Keller et.al we know that in the homogeneous players model, the planner would cease all experimentation at the belief p^{*hom} such that

$$p^{*hom} = \frac{1}{\lambda}[\frac{rs}{rh+ng-ns}]$$

where $g = \lambda h$.

From the analysis carried out in the present paper, we know that in the heterogeneous players model, the planner would cease all experimentation at the belief p_1^* such that

$$p_1^* = \frac{1}{\lambda_1}[\frac{rs}{rh+g_1+g_2-2s}]$$

Thus the benchmark amount of experimentation performed in the first model is

$$E^n_{hom} = \frac{[\log(\Lambda(p^{*hom})) - \log(\Lambda(p_0))]}{\lambda}$$

In the heterogeneous players model, the benchmark amount of experimentation is

$$E_{het} = \frac{[\log(\Lambda(p_1^*)) - \log(\Lambda(p_0))]}{\lambda}$$

To have $E^n_{hom} \geq E_{het}$, we must have $p^{*hom} \leq p_1^*$. This is because $\log(\Lambda(p))$ is strictly decreasing in p.

For $n = 2$, we have

$$p^{*hom} = \frac{1}{\lambda}[\frac{rs}{rh+2g-2s}] = \frac{1}{\lambda}[\frac{rs}{rh+g_1+g_2-2s}] > \frac{1}{\lambda_1}[\frac{rs}{rh+2g-2s}] = p_1^*$$

as $\lambda < \lambda_1$ and $g_1 + g_2 = 2g$. Since p^{*hom} is strictly decreasing in n, if $p^{*hom} \leq p_1$, then we must have $n \geq 3$.

This proves the lemma. ■

We will now show that the amount of experimentation performed in any Markov perfect equilibrium in simple strategies(i.e when player change actions at finite number of points) of the homogeneous players model is always strictly less than that in the unique non-cooperative equilibrium of the model with heterogeneous players.

From Keller et.al(2005), we know that in the model with homogeneous players, in any markov perfect equilibrium where players change actions at finite number of points, all experimentation stops at the belief p_c^1 such that

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

where $\mu_{hom} = \frac{r}{\lambda}$ and $g = \lambda h$. To recall, this is the cutoff for each person, i.e the belief at which each individual would have switched to the safe arm, had he been the only one experimenting along the risky arm.

We know from our above analysis that in the unique equilibrium of the model with heterogeneous players, the belief at which all experimentation stops is given by $\bar{p_1}$, such that

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

where $\mu_1 = \frac{r}{\lambda_1}$. Since $\lambda_1 > \lambda$, $g_1 > g$. This implies that $p_c^1 > \bar{p_1}$.

Hence, in the homogeneous players model, the amount of experimentation in any Markov perfect equilibrium in simple strategies is given by

$$E_1 = \frac{2[\log(\Lambda[p_c^1]) - \log(\Lambda(p_0))]}{\lambda}$$

The amount of experimentation in the noncooperative equilibrium of the model with heterogeneous players is given by

$$E_2 = \frac{2[\log(\Lambda[\bar{p_1}]) - \log(\Lambda(p_0))]}{\lambda}$$

Since $\bar{p_1} < p_c^1$ and $\Lambda(.)$ is decreasing in p, $E_2 > E_1$.

This shows that welfare wise, the Markov perfect equilibrium of the model with heterogeneous players will always get a higher rank than any equilibrium of the considered class of equilbria of the model with homogeneous players where each player's Poisson intensity is equal to the average intensity of the players in the heterogeneous players model and the benchmark amount of experimentation is at least as large as in the model with heterogeneous players.

This concludes the proof of the proposition. ■

We conclude this section by discussing the intuition of the above established result. In a homogeneous players model, all players have equal free riding opportunities and this reduces the amount of experimentation. However, in the heterogeneous players model, the free riding opportunities for players are different. The amount of experimentation goes up through two channels. First, the free riding opportunities of player 1 is lower. Hence, this increases the amount of experimentation. Further, although player 2 free rides in equilibrium, since his intensity of experimentation is lower, the negative effect on the total amount of experimentation is relatively less than that in the model with homogeneous players.

4 Conclusion

This paper has shown that when the players are heterogeneous with respect to their ability to learn along the risky arm, then efficiency requires diversification, i.e each player to experiment along an exclusive arm. Keeping the average Poisson intensity of the players constant, if the degree of heterogeneity is high enough then we have a unique Markov perfect equilibrium in simple cut-off strategies. The amount of experimentation in this equilibrium is more than that in any Markov perfect equilibrium in simple strategies of the equilibrium is more than that in any non-cooperative equilibrium with homogeneous players model. We consider only those homogeneous players model where the amount of experimentation in the benchmark is at least as high as in the heterogeneous players model.

References

[1] Akcigit, U., Liu, Q., 2011: “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] Fershtman, C., Rubinstein, A., 1997 “A Simple Model of Equilibrium in Search Procedures. ”,*Journal of Economic Theory* 72, 432 – 441.

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

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

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

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

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

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

APPENDIX

A Verification arguments for the planner's solution

First consider the range of beliefs $p \in (p_2^*, 1)$. From the planner's value function we know that $v(p)$ is this range satisfies

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

where $g = \lambda h$ and $\lambda = \lambda_1 + \lambda_2$. We have to show that $b_i(p,v) \geq s - g_i p$ for $i = 1,2$.

From the expression of the value function we have

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

This gives us

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

Thus

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

Hence,

$$b_i(p,v) \geq s - g_i p \text{ requires } v \geq \frac{\lambda}{\lambda_i} s$$

Since for $p \geq p_2^*$, we have $v \geq \frac{\lambda}{\lambda_2} s$, $v > \frac{\lambda}{\lambda_1} s$ as $\lambda_1 > \lambda_2$. This implies that the value function satisfies optimality on this range of beliefs. Further, since $v(p_2^*) = \frac{\lambda}{\lambda_2} s$, we can see that at $p = p_2^*$, the planner is just indifferent between having player 2 at the risky arm or at the safe arm.

Next, consider the range $p \in [p_1^*, p_2^*]$. $v(p)$ in this range satisfies

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

This gives us

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

Hence,

$$b_1(p,v) = \lambda_1 p \frac{[g - v - v^{'}(1-p)]}{r} = v - s - g_1 p$$

Thus,

$$b_1(p,v) \geq s - g_1 p \text{ requires } v - s - g_1 p \geq s - g_1 p \Rightarrow v \geq 2s$$

Since this is satisfied for the range of beliefs considered, it is indeed optimal to keep player 1 at the risky arm.

On the other hand we have

$$b_2(p,v) = \lambda_2 p \frac{[g - v - v^{'}(1-p)]}{r} = \frac{\lambda_2}{\lambda_1}[v - s - g_1 p]$$

It is optimal to keep player 2 at the safe arm if

$$b_2(p,v) \leq s - g_2 p \Rightarrow v \leq \frac{\lambda}{\lambda_2} s$$

For the range of beliefs considered, this condition is satisfied. Hence, we can infer that it is indeed optimal to keep player 2 at the safe arm.

Further, since $v(p_1^*) = 2s$ we can infer that the planner is indifferent between having player 1 at the safe arm or at the risky arm at the belief $p = p_1^*$.

Finally, we check for the region $p < p_1^*$. $v = 2s$ for this region of beliefs. Thus we have

$$b_i(p,v) = \frac{\lambda_i}{r}[g - 2s]$$

$$b_i(p,v) \leq s - g_i p \Rightarrow p \leq \frac{s\mu_i}{(\mu_i + 1)g_i + g_{j,j\neq i} - 2s}$$

where $\mu_i = \frac{r}{\lambda_i}$. From the expression of p_1^* we can infer that it is optimal to keep both players at the safe arm for $p < p_1^*$.

B To show that $p_2^* < p_2^{*n}$

At p_2^{*n} we have

$$\lambda_2 p_2^{*n} \frac{\{g_2 - v_2 - (1-p)v_2^{'}\}}{r} = s - g_2 p_2^{*n}$$

The social planner's payoff is given by $v = v_1 + v_2$. In this range of beliefs $v_i = v_i^{rr}$ for $i = 1,2$. From the planner's problem we know that social planner's benefit from making 2 use the risky arm is

$$\lambda_2 p \frac{(g_1+g_2) - v - v^{'}(1-p)}{r} = \frac{\lambda_2}{\lambda} C_1 [\Lambda(p)]^{\frac{r}{\lambda}} + \lambda_2 p_2^{*n} \frac{\{g_2 - v_2 - (1-p)v_2^{'}\}}{r} > s - g_2 p_2^{*n}$$

C_1 is the integration constant of v_1. This implies that at $p = p_2^{*n}$ the social benefit from having 2 choosing the risky arm is higher than the cost of doing that. Hence, this proves that $p_2^* < p_2^{*n}$