The Role of Heterogeneity in a model of Strategic Experimentation *

Kaustav Das†

September 15, 2016

Abstract

In this paper, I examine a situation where economic agents facing a trade-off between *exploring* a new option and *exploiting* their existing knowledge about a safe option differ with respect to their innate abilities in exploring the new option. I consider a two-armed bandit framework in continuous time with one safe arm and a risky arm. There are two players and each has access to a replica of a safe arm and a risky arm. A player using the safe arm experiences a safe flow payoff whereas the payoff from a bad risky arm is worse than the safe arm and that of 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. 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. For moderate levels of heterogeneity, equilibrium in both cut-off and non cut-off strategies exist. However, welfare wise, the cut-off equilibrium is the best. For low levels of heterogeneity, no equilibrium in cut-off strategies exists. However, at the higher end of the low range of heterogeneity, the intensity of experimentation is highest in the most heterogeneous equilibrium. As the level of heterogeneity increases, this

*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, Siddhartha Bandyopadhyay, Martin Cripps, Johannes Horner, 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 in Bonn and during his visit to Exeter. Thanks are also due to the participants of the ESEM 2016, Geneva, GAMES 2016, Maastricht and internal seminar, Penn State University, Theory workshop University of Birmingham where some subset of the results were presented. The remaining errors are my own responsibility.

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

most heterogeneous equilibrium coincides with the cut-off equilibrium and eventually is the only surviving equilibrium.

JEL Classification Numbers:C73, D83, O31.

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

1 Introduction

Economic agents are often faced with situations where they have to decide whether to exercise a new risky option or to carry on with an option of which they have complete knowledge. This leads to trade-offs between *exploration* and *exploitation*. In this paper, I address how economic agents behave in situations when they differ with respect to their abilities in exploring the new option and success in exploring the new risky option affect other agents positively. This is particularly relevant in case of teamworks where economic agents work jointly in exploring a new risky option but cannot contract upon the action of each constituent agent[1].

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 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 behaviour of other agents. In the current work, I adopt this framework of exponential bandit model with two arms to analyse how heterogeneous players behave when they face trade-offs between exploration and exploitation and there are informational externalities. The players are heterogeneous in the sense that along a particular arm, they 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 characterise the set of Markov perfect equilibria for all ranges of heterogeneity between the players. If the degree of heterogeneity between the players is sufficiently high, then there is a unique Markov perfect equilibrium which happens to be in

[1]Examples are discussed later in the introduction

simple cut-off strategies. For moderate degree of heterogeneity, there exist Markov perfect equilibria in both cut-off and non cut-off strategies. However, the intensity of experimentation in the equilibrium in cut-off strategies is always higher than that in *any* equilibrium in non cut-off strategies. When the degree of heterogeneity is low, then all equilibria are in non cut-off strategies. In this case, we can identify the *most heterogeneous* equilibrium. This is the equilibrium where the stronger player chooses the risky arm and the weaker player chooses the safe arm for largest possible range of beliefs. In this equilibrium, the weaker player uses a cut-off strategy and the stronger player uses a non cut-off strategy. As the degree of heterogeneity increases, this equilibrium eventually coincides with the equilibrium in cut-off strategies. For other equilibria, players switch arms at a lower belief. However, as in the game with homogeneous players, it is not unambiguously true that the most heterogeneous equilibrium has the lowest intensity of experimentation compared to any equilibrium where players switch arms at a lower belief[2]. In this case, we can identify two sub-levels of heterogeneity. Above the higher sub-level, the most heterogeneous equilibrium is better than any other equilibrium. Below the lower sub-level, we have the most heterogeneous equilibrium to be the worst compared to any other equilibrium. Hence, as players tend to become homogeneous, we get back the already established result in the literature. For degree of heterogeneity between these two sub-levels, the ranking between the most heterogeneous equilibrium and other equilibria is ambiguous.

The analysis starts with introducing heterogeneity in the now canonical form of the Two-armed Bandit Model (*a.la* [4]). 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 is $g_1 = \lambda_1 h$ 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 continuous time, decides on allocating players to one of the arms. The social optimum involves *specialisation* at the extremes and

[2] As in [4]

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. Then, I characterise the set of Markov perfect equilibria for all possible kinds of heterogeneity between the players. A common feature of any equilibrium is that both the beliefs where all experimentation ceases and below which only one of the players experiment are higher than the corresponding beliefs in the planner's solution. This is due to free riding by the weaker player and also due to the fact that player 1 does not internalise player 2's benefit from his own experimentation.

With respect to the nature of the equilibrium, we can characterise three regions of heterogeneity. For the highest region, there exists a unique equilibrium and this equilibrium is in cut-off strategies. For the middle range, equilibria in both cut-off and non cut-off strategies exist but compared to any equilibrium in non cut-off strategies, the intensity of experimentation is higher in the equilibrium in cut-off strategies. For the lowest range, all equilibria are in non cut-off strategies. There exists a most heterogeneous equilibrium where player 2 uses a cut-off strategy and player 1 uses a non-cutoff strategy. Within this lowest range, given λ_1, two sub-thresholds of the value of λ_2 can be identified. For λ_2 less than the lower sub-threshold, the most heterogeneous equilibrium has the highest intensity of experimentation compared to any other equilibrium. For λ_2 greater than the higher sub-threshold, the most heterogeneous equilibrium has the lowest intensity of experimentation. In between these two sub-thresholds, the ranking between the most heterogeneous equilibrium and other equilibria is ambiguous. Hence, most of the time we can characterise and identify the *best* equilibrium.

In reality, there are many issues which fit with the problem discussed in this paper. For example, consider a situation in the academic world when two heterogeneous researchers collaborate on a risky project. Thus, any breakthrough in the project will benefit both the researchers as the publication resulting out of it will have names of both. Each of these researchers can have his own area of expertise in the sense that if time is devoted towards it, then there will be a steady flow of outputs. However, if the project they collaborate on has a breakthrough, then the resulting publication will be more prestigious than the outputs from the areas of expertise. This situation can be visualised as a strategic experimentation problem in two armed bandits. The area of expertise of each researcher can be interpreted as a safe arm, and the more challenging problem can be interpreted as the risky arm. Each of the researchers has to make a choice of whether to conduct research along his safe arm or the risky arm. Here each researcher 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 researcher. Similar situation can arise in case of political coalitions. In parliamentary democracies, during elections, often political parties enter into a negotiation of seat sharing in a particular area where none of them have that much of a stronghold. The constituent parties may have their own stronghold areas. Hence, each constituent political party has to allocate its time of campaigning between its stronghold area and the area where they have formed a coalition with other parties. Campaigning in the stronghold area can be interpreted as choosing the *safe arm* and the later can be interpreted as choosing the *risky arm*. Observe that in the area where parties have formed a coalition, campaigning activities by any party gives a positive benefit to all the constituent parties. Thus, there is room for free-riding.

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 players are homogeneous. Except ([9]) and ([7]), they have a replica of bandits and *Free-riding* is a common feature in all the above models except ([9]). The paper which is closest to the current work is Keller, Rady and Cripps([5]). They find that an equilibrium in cut-off strategies never exists. The present work contributes in two ways. First, I show that with hetreogeneous players, it is not only possible to have an equilibrium in cut-off strategies, but also under certain circumstances, it can become the only surviving equilibrium. Also, except for very low level of heterogeneity, the best equilibrium is the one which is most heterogeneous[3], unlike in the model 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

[3]As stated in the introduction and will be seen later, the most heterogeneous equilibrium coincides with the equilibrium in cut-off strategies when the later exists.

by one player only. Hence, there is congestion along an arm. The Poisson arrival rates differ across the exclusive arm. The present paper differs from Thomas([9]) in three ways. First, the type of risky arm in the present paper is the same for both players unlike Thomas([9]) where types were stochastically independent. Second, conditional on the risky arm being good, the arrival rates in the present paper differ between players. Finally, there is no congestion along any arm.

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 an 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 and also for all kinds of stakes, in most cases we can identify and characterise the *best* equilibrium.

The rest of the paper is organised as follows. Section 2 lays down the details of the setting with heterogeneous players and characterise the equilibria for different ranges of heterogeneity. Section 3 briefly describes a model where heterogeneity is only with respect to the payoffs in the safe arm. 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[4]

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

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

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

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

Like ([4]), 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 exist 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^*.*

[4] We do away with the argument of v in the subsequent analysis. This is to keep the notations simple

Proof. This proposition is proved in two steps. First, from the proposed policy or solution, the planner's payoff is computed. Then, by a verification argument it is shown that this computed payoff solves the bellman equation of the planner.

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 implies 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 $p > p_2^*$. According to the proposed solution, the planner keeps 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

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

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

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

Since, $b_2(p_2^*, v) = s - g_2 p_2^*$, 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. Since value matching condition is satisfied at $p = p_2^*$, we have

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

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

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

We will now show that there exists a $p_2^* \in (p_1^*, 1)$ such that the above relation holds. At $p_2^* = p_1^*$, L.H.S of (1) is equal to $s < \frac{\lambda_1}{\lambda_2} s$. At $p_2^* = 1$, 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_2^* 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

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

The obtained value function is

$$v(p) = \begin{cases} gp + \{\frac{\frac{\lambda_1+\lambda_2}{\lambda_2} s - g p_2^*}{(1-p_2^*)[\Lambda(p_2^*)]^{\frac{r}{\lambda}}}\}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}} \equiv v_{rr} & : \text{ If } p \in (p_2^*, 1], \\ & : \\ s + [\frac{\lambda_1 g + r g_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}] p + \{\frac{s - [\frac{\lambda_1 g + r g_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}] p_1^*}{(1-p_1^*)[\Lambda(p_1^*)]^{\frac{r}{\lambda_1}}}\}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}} \equiv v_{sr} & : \text{ if } p \in (p_1^*, p_2^*], \\ & : \\ 2s & : \text{ if } p \in (0, p_1^*]. \end{cases}$$

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 ■

From the above proposition one can see that the belief where player 1 is shifted to the safe arm from the risky arm is greater than the belief at which the players will be shifted if both players' poisson arrival rates are λ_1. This is because of the fact that as one of the players's innate ability declines ($\lambda_2 < \lambda_1$), the benefit of having player 1 experimenting along the rsiky arm decreases. Hence, player 1 is shifted at a higher belief.

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

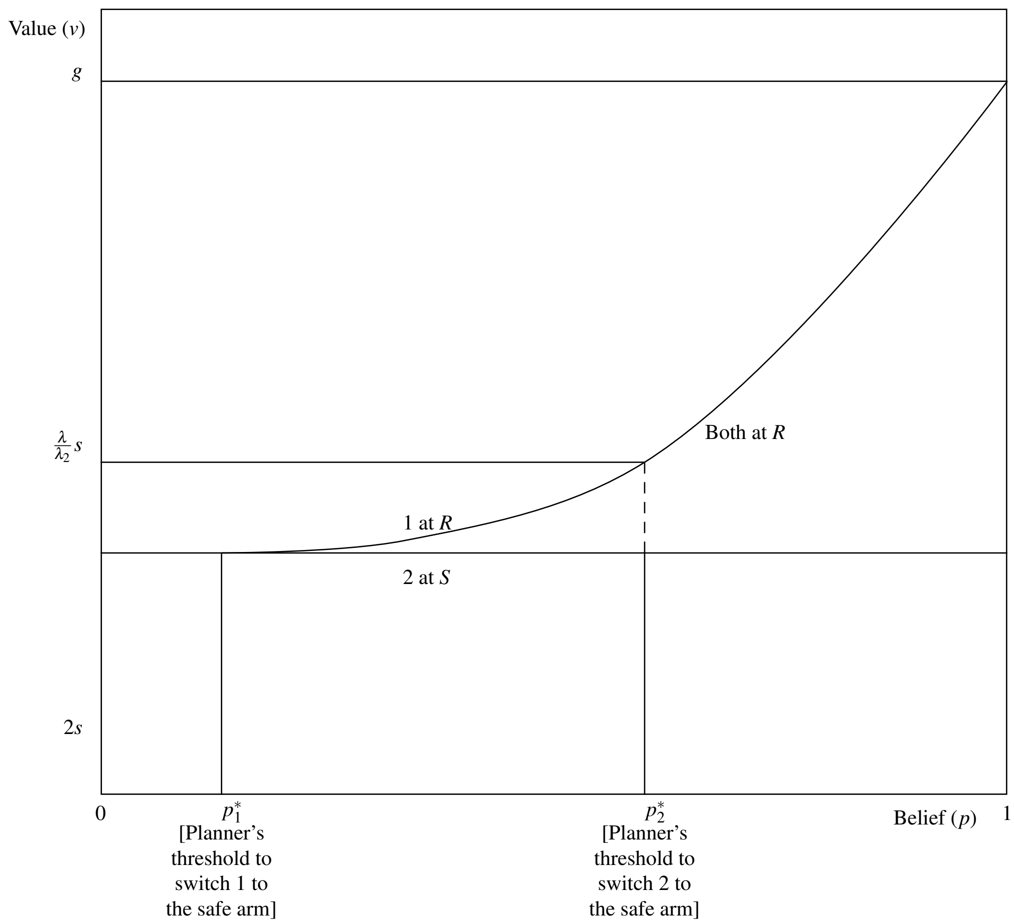

Figure 1.

The optimal value function of the planner 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.

2.2 Non-cooperative game

In this subsection, we carry out the analysis of the non-cooperative game between the players. We will 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 $k_i : [0,1] \to \{0,1\}$ ($i = 1,2$). 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 will 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

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

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

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

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

where

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

$\lambda_1 b_1^n(p, v_1)$ can be interpreted as the additional payoff accrued to player 1 due to the information generated from his own experimentation and $\lambda_2 b_1^n(p, v_1)$ 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. These interpretations are similar to ([4]).

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

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

where

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

In the same way as above, we can explain the terms $\lambda_2 b_2^n(p, v_2)$, $\lambda_1 b_2^n(p, v_2)$ 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 $\lambda_1 b_1^n(p, v_1)$) is greater than or equal to the opportunity cost of choosing the risky arm (given by $s - g_1 p$). The term $k_2[\lambda_2 b_1^n(p, v_1)]$ 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 (2) (the optimal equation of player 1), we have

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

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

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

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 there exists a region which lies above the line D_1 and below the line D_2. This gives rise to the possibility of having an equilibrium where players use cut-off strategies.

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

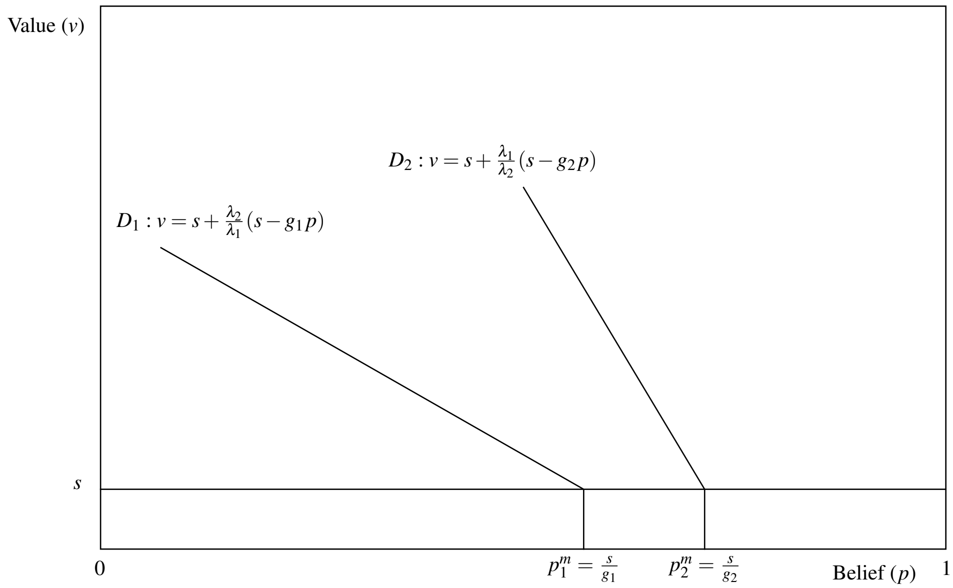

Figure 2.

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 Bellman 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 Bellman 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 Bellman 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 get

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

C in all cases represents the integration constant 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, $v_i(p) = s$ at $p = p_l$. As $p_l < \frac{s}{g_i}$, in the first case player i's payoff will 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 will now characterise the equilirbia of this game exhaustively and will examine how their nature is dependent on the degree of heterogeneity between the players. However, before doing that we define some important threshold beliefs in the following subsection.

2.3 Important threshold beliefs

In this subsection, we discuss some important threshold beliefs. These will be important for the characterisation of equilibria later.

First, consider the function $\bar{v_1}^{rs}(p)$ such that

$$\bar{v_1}^{rs}(p) = g_1 p + C_1^{rs}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

with $C_1^{rs} = \frac{s - g_1 p}{(1-\bar{p_1})[\Lambda(\bar{p_1})]^{\frac{r}{\lambda_1}}}$

We first show that there exists a unique $p_1^{'} \in (\bar{p_1}, 1)$ such that

$$\bar{v_1}^{rs}(p_1) = D_1(p)$$

where $D_1 : v = s + \frac{\lambda_2}{\lambda_1} s - g_2 p$

$\bar{v_1}^{rs}$ is strictly convex and increasing in p. D_1 is monotonically decreasing in p. At $p_1^{'} = \bar{p_1}$, $\bar{v_1}^{rs} < D_1$. At $p_1^{'} = 1$, $\bar{v_1}^{rs} > D_1$. Hence, there exists a unique $p_1^{'} \in (\bar{p_1}, 1)$ such that $\bar{v_1}^{rs} = D_1$.

Next, it can be observed that lower is the value of λ_2, lower is $p_1^{'}$. This is because as λ_2 goes down, the line D_1 shifts downwards and rotates towards left. On the other hand the function $\bar{v_1}^{rs}(p)$ is independent of the the value of λ_2. Hence, the belief at which $\bar{v_1}^{rs}$ intersects D_1 goes down.

Next, consider the function $\bar{F_2}(p)$ such that

$$\bar{F_2}(p) = s + \frac{\lambda_1}{\lambda_1 + r}[g_2 - s]p + C_2^{sr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

with $C_2^{rs} = \frac{-\lambda_1}{\lambda_1 + r}[g_2 - s]\bar{p_1}$

The function $\bar{F_2}(p)$ is strictly concave and strictly increasing in the range $p \in (\bar{p_1}, 1)$. In similar way as above it can be shown that there exists a unique $p_2^{*n} \in (\bar{p_1}, 1)$ such that the curve of the function $\bar{F_2}(p)$ intersects the line D_2 where $D_2 : v = s + \frac{\lambda_1}{\lambda_2} s - g_1 p$.

It can be argued that p_2^{*n} goes up as λ_2 goes down. This follows from the fact that as λ_2 goes down, D_2 shifts upwards and rotates towards right. On the other hand as λ_2 goes down, $\bar{F_2}(p)$ becomes flatter at every belief. This implies that the belief at which the curve of the function $\bar{F_2}(p)$ intersects D_2 goes up.

We can also argue that whenever $\lambda_2 \in (\lambda_1, \frac{s}{h})$, $\bar{p_2} < p_2^{*n}$. As $\bar{p_2}$ is monotonically decreasing in λ_2, we can infer that there exists $\lambda_2^{'}$ and λ_2^{*} satisfying $\frac{s}{h} < \lambda_2^{*} < \lambda_2^{'} < \lambda_1$ such that whenever $\lambda_2 < \lambda_2^{'}$, $p_1^{'} < p_2^{*n}$ and whenever $\lambda_2 < \lambda_2^{*}$, $p_1^{'} < \bar{p_2}$. Thus, for $\lambda_2 \in (\lambda_2^{*}, \lambda_2^{'})$, $\bar{p_2} < p_1^{'} < p_2^{*n}$

2.4 Existense of equilibrium in cut-off strategies and condition under which it is unique

In this sub-section we will demnostrate the conditions under which equilibrium in cut-off strategies exists and when it is the only surviving equilibrium. We will show that if the degree of heterogeneity between the players is high enough, then an equilibrium exists where both players use cut-off strategies and that equilibrium is the unique MPE of the game. This is illustrated in the following proposition.

Proposition 3 *If λ_2 is such that $\lambda_2 < \lambda_2^{'} < \lambda_1$, then there exists a Markov perfect equilibrium in simple cut-off strategies. The equilibrium is characterised by two 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.*

Further, if $\lambda_2 < \lambda_2^*$, then this equilibrium is the unique Markov perfect equilibrium.

Proof.

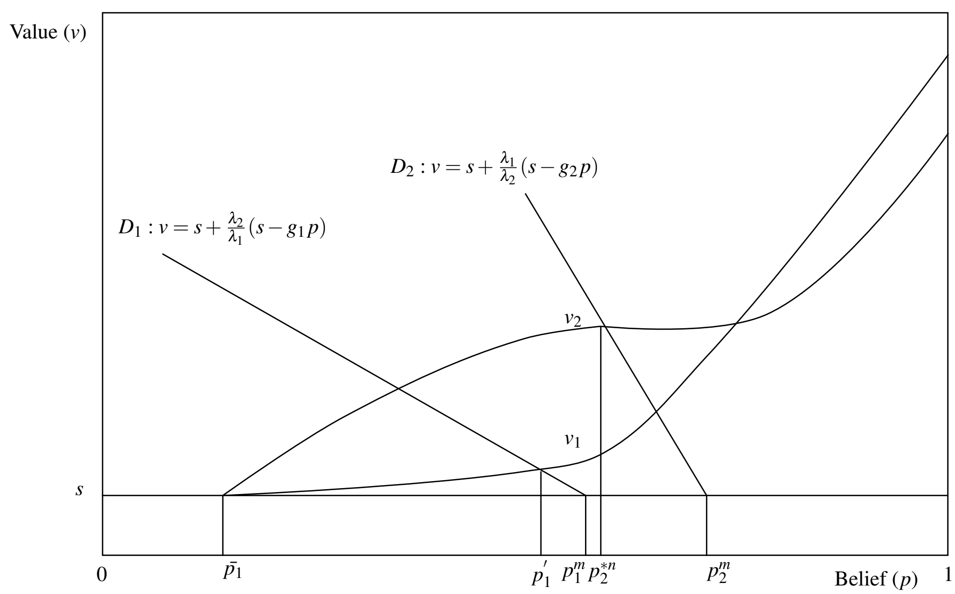

Figure 3.

As argued earlier, 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, at the right ε-neighborhood of $\bar{p_1}$, only one player can experiment along the risky arm. 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 constitutes 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[5] 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 if and 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. Appendix (B) shows that given a λ_1, we can find a $\lambda_2^{'} \in (\frac{s}{h}, \lambda_1)$ such that if $\lambda_2 < \lambda_2^{'}$, then $p_2^{*n} > p_1^{'}$ and hence, equilibrium in cut-off strategies exists.

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 will exist a range of beliefs where player 1 choosing the safe arm and player 2 choosing the risky arm will 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 and the line D_2 tends to coincide with the line D_1. 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 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 $\bar{p_2}$.

[5] 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 $p \in (\bar{p_1}, 1)$

Lastly, we show that $p_2^{*n} > \bar{p_2}$ for all $\lambda_2 > \frac{s}{h}$. 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 the unique MPE of the game.

For beliefs greater than p_2^{*n}, payoffs 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. ■

The equilibrium described above is depicted in 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 wants player 2 to experiment along the risky arm.

We will now discuss the obtained result intuitively . In an equilibrium in cut-off strategies, 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 an equilibrium in cut-off strategies 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 further and hence, player 1 can never free-ride on player 2 in any non-cooperative equilibrium.

This unique MPE in cutoff strategies is inefficient. The inefficiency arises from two channels. First, no experimentation takes place for beliefs below $\bar{p_1}$, whereas the planner wants experimentation to go on 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[6]. 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.

In the following sub-sections, we will characterise the equilbria of the game when the degree of heterogeneity between the players is not high enough. First, we focus on that

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

range of heterogeneity between the players when equilibria in both cut-off and non cut-off strategies exist and later we will focus on that range of heterogeneity when equilbria are only in non cut-off strategies.

2.5 Simultaneous existence of equilibria in both cutoff and non-cutoff strategies

In the previous subsection, we have demonstrated that an equilibrium in cut-off strategies exists if and only if $p_1^{'} < p_2^{*n}$. Further, this equilibrium is the only surviving equilibrium when $\bar{p_2} > p_1^{'}$. When $\bar{p_2} < p_1^{'}$, there exists a range of beliefs where player 1 choosing the safe arm and player 2 choosing the risky arm constitutes mutual best responses. We know that for any $\lambda_2 > \frac{s}{h}$, $p_2^{*n} > \bar{p_2}$, and given λ_1, both p_2^{*n} and $\bar{p_2}$ are monotonically decreasing in λ_2. Thus, given λ_1, there exists a range of values of λ_2 such that $p_2^{*n} > p_1^{'}$ and $\bar{p_2} < p_1^{'}$. For these values of λ_2, equilibria in both cut-off and non cut-off strategies exist.

The equilibrium in cut-off strategies is characterised as before. For beliefs below and equal to $\bar{p_1}$, both players choose the safe arm. Player 1 chooses the risky arm for any belief greater than $\bar{p_1}$. Player 2 continue choosing the safe arm for beliefs less than or equal to p_2^{*n} and chooses the risky arm for any belief greater than p_2^{*n}. The following proposition describes the equilibria in non-cutoff strategies.

To understand any non-cutoff equilibrium, we first discuss following two lemmas. The proofs are relegated to the appendix.

Lemma 1 *Consider an equilibrium in non-cutoff strategies. Suppose $p_s^1 > \bar{p_1}$ is the belief at which the payoff of player 1 meets the line D_1. If over the range of beliefs $p \in (\bar{p_1}, p_s^1)$ player 1 has switched arms at least once, then $p_s^1 < p_1^{'}$*

Lemma 2 *Consider an equilibrium in non-cutoff strategies. Let $p_2^s > \bar{p_1}$ be the belief at which the payoff of player 2 meets the line D_2. If player 2 has switched arms at least once over the range of beliefs $p \in (\bar{p_1}, p_s^2)$, then $p_s^2 > p_2^{*n}$*

The above two lemmas allow us to conclude that while comparing the equilibrium in cut-off strategies with any equilibrium in non-cutoff strategies, we can without loss of generality focus on those equilibria in non-cutoff strategies where players switch arms only once before their payoff intersect the respective best response lines.

Proposition 4 *Given λ_1, if λ_2 is such that $\lambda_2^* < \lambda_2 < \lambda_2^{'}$, then equilibria in both cut-off and non cut-off strategies exist. Any equilibrium in non-cutoff strategies is characterised by two switching points p_s and p_s^2 such that $\bar{p_2} < p_s < p_s^2 < p_1^{'}$ and a cutoff point $\hat{p_2}$ with $\hat{p_2} > p_2^{*n}$. For beliefs $(\bar{p_1}, p_s]$ and $(p_s^2, 1]$, player 1 chooses the risky arm and the safe arm for $p \in (0, \bar{p_1}]$ and $p \in (p_s, p_s^2]$. Player 2 chooses the risky arm for beliefs $p \in (p_s, p_s^2]$ and $p \in (\hat{p_2}, 1]$. For beliefs $p \in (0, p_s]$ and $p \in (p_s^2, \hat{p_2}]$, player 2 chooses the safe arm.*

Proof.

Since $\bar{p_2} < p_1^{'}$, there is a range of beliefs where player 1 choosing the safe arm is a best response to player 2 choosing the risky arm. In any equilibrium, for beliefs in the range $(\bar{p_1}, \bar{p_2}]$, only mutual best responses are player 1 choosing the risky arm and player 2 choosing the safe arm. In any equilibrium, it is not possible to have player 1 choosing the safe arm and player 2 choosing the risky arm for all $p \in (p_1^{'}, 1]$. This is because once player 1's value function intersects D_1, he cannot choose safe arm anymore in equilibrium. However, for an equilibrium in non cut-off strategies, it is a necessity that player 1 chooses the safe arm and player 2 chooses the risky arm for some range of beliefs greater than $\bar{p_1}$. Hence, in any equilibrium in non cut-off strategies, the belief where players switch arms for the first time, should always lie in $(\bar{p_2}, p_1^{'})$. Let p_s be the belief where players switch arms for the first time. $p_s \in (\bar{p_2}, p_1^{'})$.

For beliefs $p \in (\bar{p_1}, p_s]$, player 1 chooses the risky arm and player 2 chooses the safe arm. Hence, player 1's payoff in this range is given by

$$v_1^{rs1}(p) \equiv v_1^{rs}(p) = g_1 p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

with $C = \frac{s - g_1 p}{(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}}$. This is obtained from the value matching condition $v_1^{rs1}(\bar{p_1}) = s$.

Player 2's payoff in this range is given by

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

with $C = -\frac{\frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p}{(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}}$. This is obtained from the value matching condition $F_2^1(\bar{p_1}) = s$.

At the right ε- neighborhood ($\varepsilon \to 0$) of p_s, the payoff to player 1 will be given by :

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

with $C = \frac{v_1^{rs1}(p_s) - \{s + \frac{\lambda_2}{\lambda_2 + r}(g_1 - s)p_s\}}{(1-p_s)[\Lambda(p_s)]^{\frac{r}{\lambda_2}}}$

The payoff to player 2 at the right ε- neighborhood of p_s is given by

$$v_2^{rs1}(p) \equiv v_2^{rs}(p) = g_2 p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

with $C = \frac{F_2^1(p_s) - g_2 p_s}{(1-p_s)[\Lambda(p_s)]^{\frac{r}{\lambda_2}}}$

Let p_s^2 be the belief where $F_1^1(p)$ intersects the line D_1 and let $\tilde{p}_2$ be the belief where $v_2^{rs1}(p)$ intersects D_2.

As long as $F_1^1(p)$ is below the line D_1, it is always above v_1^{rs1} (refer to appendix (D) for a detailed proof) and for all beliefs $p < \frac{s}{g_2}$, v_2^{rs1} will be below $F_2^1(p)$ (refer to appendix (E) for a detailed proof). These imply that F_1^1 will intersect the line D_1 at a belief p_s^2 such that $p_s^2 < p_1^{'}$ and v_2^{rs1} will intersect D_2 at a belief which is strictly greater than p_2^{*n}. Since $p_2^{*n} > p_1^{'}$, F_1^1 will intersect D_1 before v_2^{rs1} intersects D_2. Hence, when player 1's value intersects the line D_1, player 2's value is still below D_2. Thus, at the right ε-neighborhood of p_s^2, the profile which constitutes mututal best responses is player 1 choosing the risky arm and player 2 choosing the safe arm. Hence, at the right ε-neighborhood of p_s^2, payoff to player 1 is given by

$$v_1^{rs2}(p) \equiv v_1^{rs}(p)$$

with $v_1^{rs}(p_s^2) = F_1^1(p_s^2)$

and payoff to player 2 is given by

$$F_2^2(p) \equiv F_2(p)$$

with $F_2(p_s^2) = v_2^{rs1}(p_s^2)$

Since $v_2^{rs1}(p_s^2) < F_2^1(p_s^2)$, $F_2^2(p)$ will always be below F_2^1. Hence, the belief $\hat{p}_2$ at which F_2^2 intersects D_2 is strictly greater than the belief at which $F_2^1(p)$ intersects D_2. This shows that $\hat{p}_2 > p_2^{*n}$. At $p = \hat{p}_2$, player 2 switches to the risky arm. For $p > \hat{p}_2$, payoff to player 1 is given by

$$v_1^{rr} = g_1 p + C(1-p)[\Lambda(p)]^{\frac{r}{(\lambda_1 + \lambda_2)}}$$

with the integration constant being determined from $v_1^{rr}(\hat{p}_2) = v_1^{rs2}(\hat{p}_2)$.

Payoff to player 2 for $p > \hat{p}_2$ is given by

$$v_2^{rr} = g_2 p + C(1-p)[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}}$$

with the integration constant being determined from $v_2^{rr}(\hat{p}_2) = F_2^2(\hat{p}_2)$.

■

We are now in a position to compare the equilibrium in cut-off strategies with any equilibrium in non cut-off strategies. It turns out that the intensity of experimentation in the equilibrium in cut-off strategies is higher than that in any equilibrium in non cut-off strategies. The following proposition describes this.

Proposition 5 *When equilibria in both cut-off and non cut-off strategies exist, then compared to any equilibrium in non-cutoff strategies, the intensity of experimentation is always higher in the equilibrium in cut-off strategies.*

Proof.

For the equilibrium in cut-off strategies and for any equilibrium in non cut-off strategies, no experimentation takes place for beliefs less than or equal to $\bar{p}_1$. In the equilibrium in cut-off strategies, for $p \in (\bar{p}_1, p_2^{*n}]$, the weaker player (player 2) free rides and the stronger player (player 1) experiments. For $p > p_2^{*n}$, both players experiment. On the other hand, in any equilibrium in non cut-off strategies, both players experiment for beliefs $p > \hat{p}_2$, where $\hat{p}_2 > p_2^{*n}$. Hence, compared to the equilibrium in cut-off strategies, the range of beliefs over which only one player experiments is higher in any equilibrium in non cut-off strategies. However, the total range of beliefs over which any experimentation takes place is same across two kinds of equilibria. This range is $(\bar{p}_1, 1]$. Further, in the equilibrium in cut-off strategies, whenever only one player experiments, it is the stronger player who does it. However, in any equilibrium in non cut-off strategies, we have seen that there exists a range of beliefs $(p_s, p_s^2]$, when the weaker player experiments and the stronger player free rides. Hence, the intensity of experimentation in the equilibrium in cut-off strategies is always higher than that in any equilibrium in non cut-off strategies. This concludes the proof.

■

The comparison between the equilibrium in cut-off strategies and any equilibrium in non cut-off strategies is depicted in figure 4.

The black curves v_1 and v_2 depict the payoffs to player 1 and 2 respectively in the equilibrium in cut-off strategies. In the equilibrium in non cut-off strategies, payoffs are

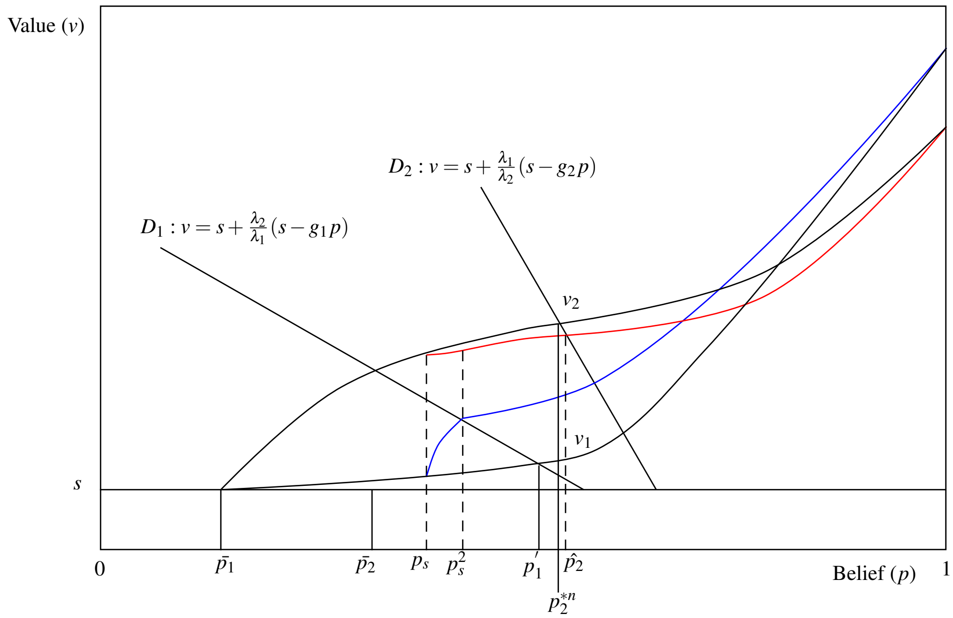

Figure 4.

same as before for beliefs less than or equal to p_s. At p_s, players switch arms. Blue curve depicts the payoff to player 1 and the red curve depicts the payoff to player 2 for $p > p_s$, in the equilibrium in non cut-off strategies . As argued, the blue curve meets the line D_1 at a belief p_s^2, which is strictly less than $p_1^{'}$. In the region $(p_s, p_s^2]$, player 2 experiments and player 1 free rides. At p_s^2, player 1 shifts to the risky arm and player 2 shifts to the safe arm. When the red curve meets the line D_2 at $\hat{p_2}$, player 2 shifts to the risky arm again. As argued, $\hat{p_2} > p_2^{*n}$.

In appendix (H) we prove that the aggregate payoffs of players in equilibrium in cut-off strategies is strictly higher than that in the equilibrium in non-cutoff strategies.

In the next sub-section, we discuss about equilibria when the degree of heterogeneity is such that no equilibrium in cut-off strategies exists.

2.6 Low degree of heterogeneity: All equilibria are in non cut-off strategies

In this subsection, we consider the situation when the degree of heterogeneity between the players is such that there does not exist any equilibrium in cut-off strategies. Given λ_1, this happens when the value of λ_2 is such that $\lambda_2^{'} < \lambda_2 < \lambda_2$

Proposition 6 *When no equilibrium in cut-off strategies exists, then the most heterogeneous equilibrium is characterised as follows. In this equilibrium, only the weaker player*

(player 2) uses a cut-off strategy. There is a switching point $\hat{p}_1 \in (p_2^{*n}, p_1')$, such that for $p \in (\bar{p}_1, p_2^{*n}]$ and $p \in (\hat{p}_1, 1]$, player 1 chooses the risky arm and for $p \in (0, \bar{p}_1]$ and $p \in (p_2^{*n}, \hat{p}_1]$, he chooses the safe arm. Player 2 chooses the safe arm for $p \leq p_2^{*n}$ and the risky arm for $p > p_2^{*n}$. As λ_2 decreases below the threshold λ_2', this equilibrium becomes the equilibrium in cut-off strategies.

Proof. By most heterogenoeus equilibrium we mean the equilibrium where actions of players are different to the maximum possible extent. As $p_1^{'} > p_2^{*n}$, no equilibrium in cut-off strategies exists. As before, at the right ε neighborhood of $\bar{p}_1$, player 1 choosing the risky arm and player 2 choosing the safe arm constitutes mutual best responses. Since $p_2^{*n} < p_1^{'}$, for $\bar{p}_1 < p < p_2^{*n}$, player 1 choosing the risky arm and player 2 choosing the safe arm constitutes mutual best responses. Thus for $\bar{p}_1 < p \leq p_2^{*n}$, payoff to player 1 is given by $v_1^{rs1}(p)$ with $v_1^{rs1}(\bar{p}_1) = s$. Player 2's payoff is given by $F_2^1(p)$ with $F_2^1(\bar{p}_1) = s$. At the belief $p = p_2^{*n}$, player 2 shifts to the risky arm and player 1 shifts to the safe arm. Payoff to player 1 is given by $F_1^1(p) \equiv F_1(p)$ with $F_1(p_2^{*n}) = v_1^{rs1}(p_2^{*n})$. $F_1^1(p)$ is strictly concave and from our previous arguments we know that it will lie above $v_1^{rs1}(p)$ as long as $F_1^1(p) < D_1(p)$. Hence, the belief at which $F_1^1(p)$ will intersect D_1 is strictly less than p_1'. Let this belief be $\hat{p}_1$. Once $F_1^1(p)$ intersects D_1 at $\hat{p}_1$, player 1 switches to the risky arm.

Payoffs to player 2 for $p_2^{*n} < p \leq \hat{p}_1$ is given by $v_2^{rs1}(p) \equiv v_2^{rs}(p)$ with $v_2^{rs}(p_2^{*n}) = F_2^1(p_2^{*n})$. For beliefs higher than $\hat{p}_1$, payoffs to player 1 is given by $v_1^{rr}(p)$ with $v_1^{rr}(\hat{p}_1) = F_1^1(\hat{p}_1)$. Payoff to player 2 is given by $v_2^{rr}(p)$ with $v_2^{rr}(\hat{p}_1) = v_2^{rs}(\hat{p}_1)$.

The range of beliefs over which player 1 free-rides is $(p_2^{*n}, \hat{p}_1] \in (p_2^{*n}, p_1^{'}]$. As λ_2 decreases, p_2^{*n} gets closer to $p_1^{'}$ and hence the range of beliefs over which player 1 free rides shrinks. When $\lambda_2 = \lambda_2'$, $p_2^{*n} = p_1^{'}$. In that case, both $v_1^{rs1}(p)$ and $F_2^1(p)$ intersect D_1 and D_2 respectively at the same belief. It is from this point that the equilibrium in cut-off strategies begins to exist. Hence, for all $\lambda_2 < \lambda_2'$, if in an equilirbium player 2 uses a cut-off strategy then player 1 also have to use a cut-off strategy. Hence, at $p_2^{*n} = p_1^{'}$, the most heterogeneous equilibrium becomes the equilibrium in cut-off strategies. It continues to be so for all $\lambda_2 < \lambda_2'$.

■

In the above equilibrium, player 2 uses a cut-off strategy and player 1 uses a non cut-off strategy. The equilibrium is depicted in figure 5. v_1 and v_2 denote the payoffs to player 1 and 2 respectively. The range of beliefs over which player 1 free rides is $(p_2^{*n}, p_s]$. It can be seen from the figure that as λ_2 decreases, p_2^{*n} comes closer to $p_1^{'}$ and hence the range of beliefs over which player 1 free rides shrinks and eventually as p_2^{*n} crosses $p_1^{'}$, player 1 stops free-riding and we have equilibrium in cut-off strategies.

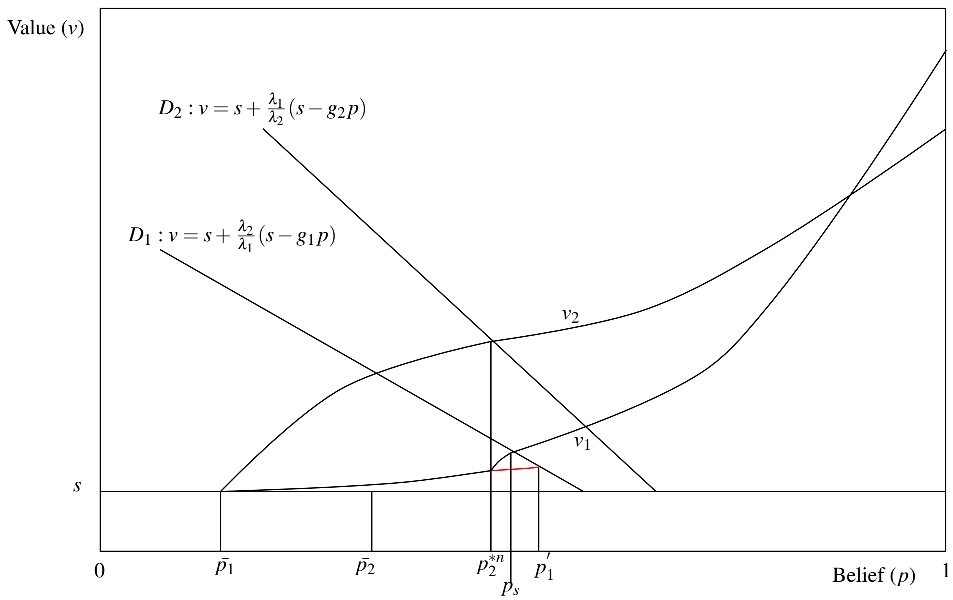

Figure 5.

We will now characterise other equilibria in non cut-off strategies. The following proposition demonstrates this.

Proposition 7 *If $p_2^{*n} < p_1^{'}$, then apart from the most heterogeneous equilibrium, other equilibria in non cut-off strategies also exist. They can be of the following two kinds*[7]

1 . *There exists a switching point p_s and two thresholds $\tilde{p}_1$ and $\tilde{p}_2$ such that $p_s \in (\bar{p}_2, p_2^{*n})$ and $p_2^{*n} < \tilde{p}_2 < \tilde{p}_1 < p_1^{'}$. Player 1 chooses the safe arm for $p \in (0, \bar{p}_1]$, and $p \in (p_s, \tilde{p}_1]$. He plays risky arm for $p \in (\bar{p}_1, p_s]$ and $p \in (\tilde{p}_1, 1]$. Player 2 chooses the safe arm for $p \in (0, p_s]$ and the risky arm for $p \in (p_s, 1]$. Hence, player 1 uses a non cut-off strategy and player 2 uses a cut-off strategy.*

2. *There exists a switching point p_s and two thresholds $\tilde{p}_1$ and $\tilde{p}_2$ such that $p_s \in (\bar{p}_2, p_2^{*n})$ and $\tilde{p}_1 < \tilde{p}_2$. Player 1 chooses the safe arm for $p \in (0, \bar{p}_1]$ and $p \in (p_s, \tilde{p}_1]$. He chooses the risky arm for $p \in (\bar{p}_1, p_s]$ and $p \in (\tilde{p}_1, 1]$. Player 2 chooses the safe arm for $p \in (0, p_s]$ and $p \in (\tilde{p}_1, \tilde{p}_2]$. He chooses the risky arm for $p \in (p_s, \tilde{p}_1]$ and $p \in (\tilde{p}_2, 1]$. Hence, both players use non-cutoff strategies.*

[7]Again, without loss of generality, we consider only those equilibria where players switch arms only once over the range of beliefs when both players' payoff functions are below their respective best response lines.

Proof.

Consider a point $p_s \in (\bar{p}_2, p_2^{*n})$. For $p \in (\bar{p}_1, p_s]$, player 1 chooses the risky arm and player 2 chooses the safe arm. Since $p_2^{*n} < p_1^{'}$, from our previous analysis we know that at $p = p_s$, player i's ($i = 1,2$) payoff function lies below the line D_i. Hence, at the right ε neighborhood of p_s, one player choosing the risky arm and the other choosing the safe arm constitute mutual best responses. Hence, we can have an equilibrium where both players switch arms at a belief $p_s \in (\bar{p}_2, p_2^{*n})$. In that case, at the right ε neighborhood of p_s , payoff to player 1 is given by

$$F_1^1(p) \equiv F_1(p) \text{ with } F_1(p_s) = v_1^{rs1}(p_s)$$

This function is concave and as argued earlier, as long as it is below the line D_1, it lies above $v_1^{rs1}(p)$. Hence, the belief at which $F_1^1(p)$ intersects the line D_1 is strictly less than $p_1^{'}$. Let the belief at which $F_1^1(p)$ intersects the line D_1 be $\tilde{p}_1$.

The payoff to player 2 is given by

$$v_2^{rs1}(p) \equiv v_2(p) \text{with } v_2(p_s) = F_2^1(p_s)$$

This function is strictly convex and always lies below $F_2^1(p)$ as long as $p < \frac{s}{g_2}$. Let $\hat{p}_2$ be the belief at which $v_2^{rs1}(p)$ intersects the line D_2. We have $\hat{p}_2 > p_2^{*n}$. There are two possibilities:

1. $\hat{p}_2 < \tilde{p}_1$: In this case, we have the equilibrium of the first kind and the threshold $\tilde{p}_2 = \hat{p}_2$.

2. $\hat{p}_2 > \tilde{p}_1$: In this case player 1' value function intersects D_1 at a belief which is less than the belief at which player 2's value function intersects the line D_2. We will have the equilibrium of the second kind. Hence, at $\tilde{p}_1$, player 1 will switch to the risky arm and player 2 will switch to the safe arm. At the right ε neighborhood of $\tilde{p}_1$, player 1's payoff is given by

$$v_1^{rs2}(p) \equiv v_1^{rs}(p) \text{ with } v_1^{rs}(\tilde{p}_1) = F_1^1(\tilde{p}_1)$$

and player 2's payoff is given by

$$F_2^2(p) \equiv F_2(p) \text{ with } F_2(\tilde{p}_1) = v_2^{rs}(\tilde{p}_1)$$

Let $\tilde{p}_2$ be the belief at which $F_2^2(p)$ intersects the line D_2. Beyond this belief, both players choose the risky arm.

This concludes the proof of the proposition. ■

Comparing different equilibria when all equilibria are in non cut-off strategies:

Consider an arbitrary equilibrium in non cut-off strategies. A particular feature of this class of equilibria is that player 1 shifts to the safe arm from the risky arm at a belief which is in the interval $(\bar{p}_2, p_1^{'})$. Player 1 after switching to the safe arm continue to choose the safe arm as long as his payoff does not meet D_1. Thereafter, player 1 chooses the risky arm. The following lemma establishes the fact that sooner the players initially switch, lower is the belief at which player 1's payoff meets the best response line D_1.

Lemma 3 *Consider $p_2^{*n} < p_1^{'}$ and two equilibria in non cut-off strategies such that for one the initial switching point is p_{s_1} and for the other it is p_{s_2} with $p_{s_1} < p_{s_2}$. The belief at which 1's payoff meets D_1 is strictly lower for the equilibrium with the initial switching point p_{s_1}.*

Proof.

Consider the equilibrium with the initial switching point p_{s_1}. The payoff to player 1 at the right ε neighborhood of p_{s_1} is given by $F_1^{11}(p) \equiv F_1(p)$ with $F_1(p_{s_1}) = v_1^{rs1}(p_{s_1})$. This payoff function will be concave and will be given by

$$F_1^{11}(p) = s + \frac{\lambda_2}{r+\lambda_2}(g_1 - s)p + C_1^{sr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

Let $\tilde{p_1}$ be the belief at which $F_1^{11}(p)$ meets D_1. The proof is trivial if the other equilibrium has the initial switching point $p_{s_2} > \tilde{p_1}$. Hence, consider $p_{s_2} < \tilde{p_1}$. From our previous arguments we know that for $p \in (p_{s_1}, \tilde{p_1}]$, $F_1^{11}(p) > v_1^{rs1}(p)$. For the equilibrium with switching point p_{s2}, the payoff to player 1 at the right ε neighborhood of p_{s_2} is given by $F_1^{12}(p) \equiv F_1(p)$ with $F_1(p_{s_2}) = v_1^{rs1}(p_{s_2})$. This payoff function will be concave and will be given by

$$F_1^{12}(p) = s + \frac{\lambda_2}{r+\lambda_2}(g_1 - s)p + C_2^{sr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

We have $F_1^{11}(p_{s_2}) > v_1^{rs1}(p_{s_2}) = F_1^{12}(p_{s_2})$. Since $F_1^{11}(p)$ and $F_1^{12}(p)$ cannot intersect for $p < 1$, the belief at which $F_1^{12}(p)$ meets D_1 is higher. This concludes the proof of the lemma.

■

The above lemma implies that for any equilibrium where both players use non cut-off strategies, the belief at which player 1's payoff function meets D_1 is strictly less than that in the most heterogeneous equilibrium. This is because, in the later case the belief at which player 1 switches to the safe arm is p_2^{*n} and in the former case it is at a belief less than p_2^{*n}. The following proposition shows that unlike in the case with homogeneous players, the most heterogeneous equilibrium is not always the one with lowest intensity of experimentation.

Proposition 8 *Given λ_1, if λ_2 is such that $p_2^{*n} < p_1^{'}$, then there exist sub-thresholds λ_2^{lh} and λ_2^{ll} such that $\lambda_2^{ll} < \lambda_2^{lh}$. If $\lambda_2 < \lambda_2^{ll}$ then the intensity of experimentation is higher in the most heterogeneous equilibrium than all other equilibria. If $\lambda_2 > \lambda_2^{lh}$, then the intensity of experimentation is lowest for the most heterogeneous equilibrium.*

Proof.

Consider $\lambda_2 = \lambda_2^{'}$. Thus, $p_2^{*n} = p_1^{'}$. At this point, the most heterogenoeus equilibrium coincides with the equilibrium in cut-off strategies. From our previous discussions we know that in any other equilibrium, the intensity of experimentation goes down through two channels. First, the range of beliefs over which both players experiment shrinks. Next, there emerges a range of beliefs where player 1 free rides. Hence, any other equilibrium will have lower intensity of experimentation than the most heterogeneous equilibrium. In any equilibrium which is not the most heterogeneous, the belief at which the payoff of player 2 meets D_2 is a continuous function of λ_2. For values of λ_2 at the right ε neighborhood of $\lambda_2^{'}$, the belief above which both players experiment in the most heterogeneous equilibrium is $\hat{p_1} \approx p_1^{\prime}$. From the previous proposition we can infer that for λ_2 at the right ε neighborhood of $\lambda_2^{\prime}$, any other equilibrium will be of the second kind and $\tilde{p_2} > \hat{p_1}$. Thus, the range of beliefs over which both players experiment will be the highest for the most heterogeneous equilibrium.

The effect through the other channel is ambiguous when $\lambda_2 > \lambda_2^{'}$. In this case, even in the most heterogeneous equilibrium, there is a range of beliefs when player 1 free rides. Hence, if the players switch arms earlier then it can have two effects. The range of beliefs over which player 1 free rides can increase. At the same time, there is some higher range of beliefs over which in the most heterogeneous equilibrium player 1 free rides while in the other equilibrium player 2 free-rides. Thus, when we move from the most heterogeneous equilibrium to any other equilibrium, we give up some good experiments at the lower belief in return of having some good experiments at the higher belief. Thus the effect through this channel is ambiguous. However, the range of beliefs over which player 1 free rides in the most heterogeneous equilibrium goes to zero as $\lambda_2 \to \lambda_2^{'}$. Thus, at the right ε neighborhood of $\lambda_2^{'}$, there is a negative effect through this channel and this effect is a continuous function of λ_2. As already argued, the range of beliefs over which both players experiment is higher for the most heterogeneous equilibrium when λ_2 is in the right ε neighborhood of $\lambda_2^{'}$. Since, the beliefs at which players' value functions meet the best response lines are continuous fucntions of λ_2, there exists a $\lambda_2^{ll} > \lambda_2^{'}$ such that if $\lambda_2 \in (\lambda_2^{'}, \lambda_2^{ll})$, the intensity of experimentation will be highest for the most heterogeneous equilibrium.

Next, given λ_1, consider λ_2 such that $\lambda_2 \to \lambda_1$. Then, from the results of the model with homogeneous players, we know that the most heterogeneous equilibrium is always the one with lowest intensity of experimentation. Since the beliefs at which the payoff functions intersect the best response lines are all continuous functions of λ_2, there exists a λ_2^{lh} with $\lambda_2^{ll} < \lambda_2^{lh} < \lambda_1$, such that for all $\lambda_2 \in (\lambda_2^{lh}, \lambda_1)$, the intensity of experimentation will be lowest for the most heterogeneous equilibrium.

For $\lambda_2 \in (\lambda_2^{ll}, \lambda_2^{lh})$, the ranking between the most heterogeneous equilibrium and the other equilibria is ambiguous. This happens because of the effect from the second channel as explained above.

■

In the next section, we will briefly discuss that if we consider a model where the players are identical in getting breakthrough along the good risky arm (i.e $\lambda_1 = \lambda_2$) but differ with respect to the flow payoff obtained at the safe arm, the qualitative results obtained in the paper this far are not altered.

3 Heterogeneity in the safe arm payoffs

Consider a variant of the model considered in the paper. Suppose players are identical in their innate abilities in exploring the risky arm. That is, $\lambda_1 = \lambda_2 = \lambda$. However, they differ with respect to the payoffs obtained by choosing the safe arm. Let s_i be the flow payoff obtained by player i by choosing the safe arm such that

$$s_1 < s_2 < g$$

where $g = \lambda h$.

We first show that the Planner's solution will be of the same nature as obtained in the model where players only differ with respect to their abilities to get a breakthrough along the good risky arm.

Planner's Solution: The planner's objective is to maximise the sum of the expected payoffs of the players. Planner's action is denoted by the pair (k_1, k_2) $(k_i \in \{0,1\})$. $k_i = 0(1)$ denotes that the planner has allocated player i at the safe(risky) arm. If $v(p)$ is the optimal value function of the planner, then the bellman equation of the planner is given by

$$v(p) = s_1 + s_2 + \max_{k_1 \in \{0,1\}} k_1[b(p,v) - c_1(p)] + \max_{k_2 \in \{0,1\}} k_2[b(p,v) - c_2(p)]$$

where $b(p,v) = \frac{\lambda p\{2g - v - v'(1-p)\}}{r}$ and $c_i(p) = s_i - gp$

Analogous to the planner's solution obtained in the previouis section, we can show that there exist two thresholds p^*_{12} and p^*_{22} such that

$$\frac{s_1\mu}{(2+\mu)g - (s_1+s_2)} = p^*_{12} < p^*_{22} < 1$$

where $\mu = \frac{r}{\lambda}$. For all beliefs greater than p^*_{22}, both players are made to choose the risky arm. For beliefs greater than p^*_{12} and less than equal to p^*_{22}, player 1 is made to choose the risky arm and player 2 is made to choose the safe arm. For all beliefs less than or equal to p^*_{12}, both players are made to choose the safe arm.

Non-cooperative game:

We restrict ourselves to markovian strategies with the comon posterior as the state variable. Let the strategy of player i be denoted as k_i. It is defined by the mapping $k_i : [0,1] \to \{0,1\}$. $k_i = 0(1)$ denotes that player i is choosing the safe(risky) arm. Let $v_i(p)$ $(i = 1,2)$ be the optimal value function of the players. Then, analogous to the previous section, the individual bellman equations are given as

$$v_1 = s_1 + k_2[b_n(p,v_1)] + \max_{k_1 \in \{0,1\}} k_1[b_n(p,v_1) - (s_1 - gp)]$$

and

$$v_2 = s_2 + k_1[b_n(p,v_2)] + \max_{k_2 \in \{0,1\}} k_2[b_n(p,v_2) - (s_2 - gp)]$$

where $b_n(p,v_i) = \frac{\lambda p\{g - v_i - v_i'(1-p)\}}{r}$

We will now determine the best responses of the players. Consider player 1. Given that player 2 is choosing the risky arm (i.e $k_2 = 1$), player 1's best response is to choose the risky arm as long as $b_n(p, v_1) > s_1 - gp$. This implies that when player 1 is optimally choosing the risky arm, we will have

$$v_1 \geq s_1 + s_1 - gp$$

Hence, given that the other player is choosing the risky arm, choosing the risky arm constitutes a best response for player 1 as long as in the $v - p$ plane, player 1's value lies above the line

$$D_1 : v_1 = s_1 + [s_1 - gp]$$

Similarly, for player 2, given that player 1 is choosing the risky arm, choosing the risky arm constitutes a best response for player 2 as long as in the $v - p$ plane, player 2's value lies above the line

$$D_2 : v_2 = s_2 + [s_2 - gp]$$

If player 2 is choosing the safe arm, then player 1 chooses the risky arm as long as the belief is greater than $\bar{p_{1s}}$ where

$$\bar{p_{1s}} = \frac{\mu s_1}{(1+\mu)g - s_1}$$

Similarly, if player 1 is choosing the safe arm, then player 2 chooses the risky arm as long as the belief is greater than $\bar{p_{2s}}$ where

$$\bar{p_{2s}} = \frac{\mu s_2}{(1+\mu)g - s_2}$$

Since, $s_1 < s_2$, we have $\bar{p_{1s}} < \bar{p_{2s}}$.

The best responses of the players are depicted in figure 6.

From figure 6, it can be observed that heterogeneity in safe arm payoffs makes the best response lines of the players to lie apart from each other. When $s_1 = s_2$, then the lines coincide with each other and when $s_2 = g$, the lines are farthest apart from each other. Hence, we see that the qualitative effect on the best response lines due to heterogeneity in safe arm payoffs is the same as it is in the model where players differ only with respect to their innate abilities in getting a breakthrough along the good risky arm. Thus, we can posit that same kind of equilibrium analysis for different ranges of heterogeneity can be done. The analytical characterisations will be different but they can be obtained in the identical

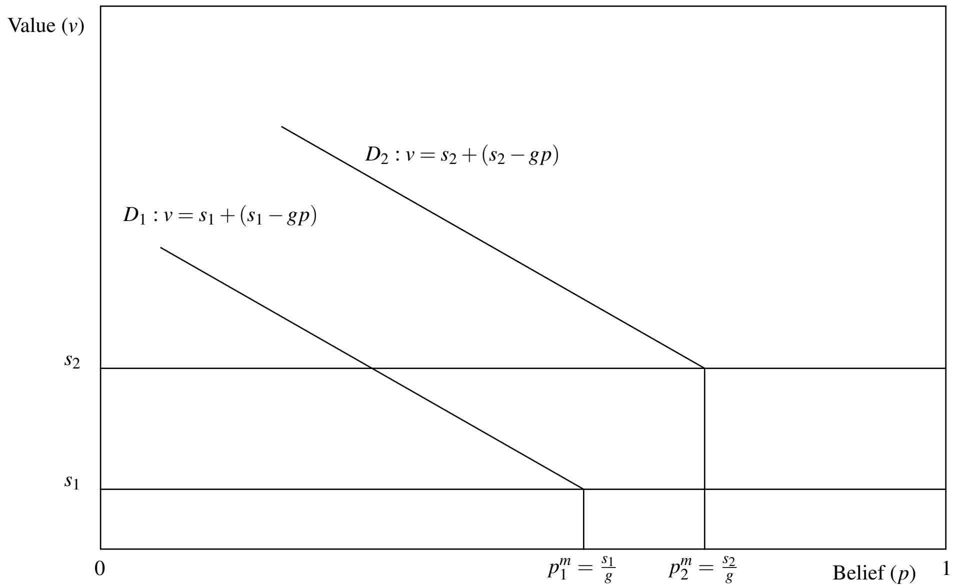

Figure 6.

way as done in the previous section.

4 Conclusion

This paper has exhaustively characterised the equilbria in a two armed bandit game when players are heterogeneous. As agents become heterogeneous, the most heterogeneous equilibrium tends to become the equilibrium with maximum intensity of experimentation. As heterogeneity between the agents increases, equilibrium in cut-off strategies exists. When equilibria in both cut-off and non cut-off strategies exist, the intensity of experimentation is always highest in the former. As heterogeneity increases further, the equilibrium in cut-off strategies becomes the only surviving Markov perfect equilibrium. Thus, one of the crucial take away of the paper is that, except for very low level of heterogeneity, we can always characterise and identify the best equilibrium outcome of the game. Also, heterogeneity in any for has the same qualitative effects on the non-cooperative solutions of the game.

References

[1] Akcigit, U., Liu, Q., 2015 : "The Role of Information in Competitive Experimentation. ", *Journal of the European Economic Association*.

[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., 2013: “Experimentation with Congestion ”, *mimeo, University College of London and University of Texas Austin*

[10] Tonjes, Katharina., 2009: “Strategisches Experimentieren mit asymmetrischen Spielern ”, *mimeo, Unpublished Ph.D thesis, LMU, Munich*

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

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

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

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

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

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

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

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 Condition for existence of equilibrium in cutoff strategies

We have argued that for an equilibrium in cutoff strategies to exist, we must have $p_2^{*n} > p_1^{'}$.

Given a λ_1, consider $\lambda_2 \to \lambda_1$. In this case, the line D_2 almost coincides with D_1 and from ([4]), we know that $p_2^{*n} < p_1^{'}$.

As $\lambda_2 \to \frac{s}{h}$, the belief at which D_2 intersects the horizontal line s approaches 1. Also, $F_2 \approx s$ as $\lambda_2 \to \frac{s}{h}$. Hence, $p_2^{*n} \to 1 > p_1^{'}$.

As λ_2 decreases from λ_1 to $\frac{s}{h}$, D_2 shifts right with slope remaining the same and F_2 becomes flatter. Hence, the belief at which F_2 intersects D_2 is monotonically decreasing in λ_2. On the other hand, as λ_2 decreases, D_1 becomes flatter and pivots downward along the point $(\frac{s}{g}, s)$. Hence, $p_1^{'}$ is monotonically increasing in λ_2.

Thus there exists a $\lambda_2^{'}$ such that for all $\lambda_2 < \lambda_2^{'}$, $p_2^{*n} > p_1^{'}$

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

At p_2^{*n} we have

$$\lambda_2 b_2^n(p, v_2) = \lambda_2 p_2^{*n} \frac{\{g_2 - v_2 - (1-p)v_2^{'}\}}{r} = s - g_2 p_2^{*n}$$

This is because the private benefit to player 2 by staying at the risky arm is equal to the cost to player 2 by moving to the safe arm.

However, due to player 2's experimentation, player 1's benefit is $\lambda_2 b_1^n(p, v_1)$. Thus, if player 2 continues to experiment along the risky arm at and at the left ε neighborhood of p_2^{*n}, the sum of payoffs will be higher. Since the sum of payoffs is highest for the planner's solution, we must have $p_2^{*n} > p_2^*$.

D $v_2^{rs1}(p)$ lies below $F_2^1(p)$ for all $p < \frac{s}{g_2}$

We first show that v_2^{rs1} is strictly convex.

Consider the function

$$v_2^{rs}(p) = g_2 p + C(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

with C being determined from $v_2^{rs}(p_s) = s$. Since $p_s < \frac{s}{g_2}$, the integration constant will be strictly positive and hence the function will be strictly convex. By definition,

$$v_2^{rs1} \equiv v_2^{rs}(p)$$

with $v_2^{rs}(p_s) = F_2^1(p_s)$. Since $F_2^1(p_s) > s$, we must have the integration constant of v_2^{rs1} to be strictly greater than that of v_2^{rs} with $v_2^{rs}(p_s) = s$. Hence, v_2^{rs1} is strictly convex.

We have

$$v_2^{rs1}(p) = g_2 p + C_2^{rs}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

and

$$F_2^1(p) = s + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p + C_2^s(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

This gives us

$$F_2^{1'}(p_s) = \frac{\lambda_1}{\lambda_1 + r}(g_2 - s) - C_2^{sr}[\Lambda(p_s)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{\lambda_1 p_s}]$$

and

$$v_2^{rs1'}(p_s) = g_2 - C_2^{rs}[\Lambda(p_s)]^{\frac{r}{\lambda_2}}[1 + \frac{r}{\lambda_2 p_s}]$$

From the value matching condition at $p = p_s$, we have

$$C_2^{rs}[\Lambda(p_s)]^{\frac{r}{\lambda_2}} - C_2^{sr}(1 - p_s)[\Lambda(p_s)]^{\frac{r}{\lambda_1}} = \frac{s + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p - g_2 p}{(1 - p)}$$

Next, since $[1 + \frac{r}{\lambda_2 p_s}] > [1 + \frac{r}{\lambda_1 p s}]$ and $C_2^{sr} > 0$, we have

$$v_2^{rs'}(p_s) = g_2 - C_2^{rs}[\Lambda(p_s)]^{\frac{r}{\lambda_2}}[1 + \frac{r}{\lambda_2 p_s}] < g_2 - C_2^{rs}[\Lambda(p_s)]^{\frac{r}{\lambda_2}}[1 + \frac{r}{\lambda_1 p_s}] = \hat{v_2}$$

Hence,

$$F_2^{1'}(p_s) - \hat{v_2} = \{\frac{\lambda_1}{\lambda_1 + r}(g_2 - s) - g_2\} + \frac{[1 + \frac{r}{\lambda_1}]\{-g_2 p + s + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p_s\}}{(1 - p_s)}$$

$$= \frac{r}{\lambda_1}[\frac{s}{p_s} - g_2] > 0$$

since $p_s < \frac{s}{g_2}$.

Since $F_2^1(p)$ is strictly concave and v_2^{rs1} is strictly convex, for all $p < \frac{s}{g_2}$, $F_2^1(p)$ will be strictly above $v_2^{rs1}(p)$.

E $F_1^1(p)$ lies above $v_1^{rs1}(p)$ as long as $F_1^1(p) < D_1(p)$

At $p = p_s$, player 2 changes to risky arm. Hence, at the right ε neighborhood of p_s, the best response of player 1 is to choose the safe arm. Given that player 2 is choosing the risky arm, player 1's best response will be to choose the safe arm as long as $F_1^1(p) < D_1(p)$. Since player 1 always had the option of choosing the risky arm, it must be the case that as long as $F_1^1(p) < D_1(p)$, we must have

$$F_1^2(p) \geq v_1^{rr}(p) = g_1 p + C(1-p)[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}}$$

with $v_1^{rr}(p_s) = v_1^{rs1}(ps)$.

Next, we prove that whenever $v_1^{rr}(p)$ and $v_1^{rs1}(p)$ intersect at a belief $p < 1$, it must be the case that $v_1^{rr'}(p) > v_1^{rs'}(p)$.

Since at the belief p, $v_1^{rr}(p) = v_1^{rs}(p)$, we have

$$g_1(p) + C_1^{rr}(1-p)[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}} = g_1 p + C_1^{rs}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

$$\Rightarrow C_1^{rr}[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}} = C_1^{rs}[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

$C_1^{rr} > 0$ and $C_1^{rs} > 0$.

This gives us

$$v_1^{rr'}(p) = g_1 - C_1^{rr}[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}}[1 + \frac{r}{(\lambda_1+\lambda_2)p}] = g_1 - C_1^{rs}[\Lambda(p)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{(\lambda_1+\lambda_2)p}]$$

$$> g_1 - C_1^{rs}[\Lambda(p)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{\lambda_1 p}] = v_1^{rs1'}(p)$$

Consider the function $v_1^{rr}(p)$ such that $v_1^{rr}(p_s) = v_1^{rs1}(p_s)$. From our above conclusion we can infer that at the right ε neighborhood of p_s, $v_1^{rr}(p) > v_1^{rs1}(p)$. Further, they cannot cross again since if they have to cross then $v_1^{rs1'}()$ would have to be strictly greater than $v_1^{rr'}()$. However, as argued, this is not possible. Thus, for all $p_s < p < 1$, $v_1^{rr}(p) > v_1^{rs1}(p)$. This implies that as long as $F_1^1(p) < D_1(p)$,

$$F_1^1(p) \geq v_1^{rr}(p) > v_1^{rs}(p)$$

F Proof of lemma 1

Proof.

Consider an equilibrium in non-cutoff strategies where player 1 switches arms at least once over the range of beliefs $p \in (\bar{p}_1, p_s^1)$.

In equilibrium, before player 1's payoff meets D_1 from below, if player 1 is choosing the risky (safe) arm then player 2 must be choosing the safe (risky) arm. Hence, if player 1 is choosing the risky arm, his payoff is given by

$$v_1^{rs}(p) = g_1 p + C_1^{rs}(1-p)[\lambda(p)]^{\frac{r}{\lambda_1}}$$

The integration constant C_1^{rs} is determined from the value matching condition at the belief where player 1 switches to the risky arm.

On the other hand, when player 1 chooses the safe arm, then his payoff is given by

$$F_1(p) = s + \frac{\lambda_2}{\lambda_2 + r}(g_1 - s)p + C_2^{sr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

The integration constant C_2^{sr} is determined from the value-matching condition at the belief where player 1 switches to the safe arm.

Suppose at a belief $\tilde{p}_1$, player 1's equilibrium payoff is below the line D_1 and player 1 switches to the safe arm from the risky arm. Then the payoff to 1 at and to the left neighborhood of $\tilde{p}_1$ is given by $v_1^{rs}(p)$ and at the right neighborhood of $\tilde{p}_1$ is given by $F_1(p)$ with $F_1(\tilde{p}_1) = v_1^{rs}(\tilde{p}_1)$. We will now argue that as long as $F_1(p)$ is below the line D_1, $F_1(p)$ will be strictly above $v_1^{rs}(p)$. Given that player 2 is choosing the risky arm, player 1's best response will be to choose the safe arm as long as $F_1^1(p) < D_1(p)$. Since player 1 always had the option of choosing the risky arm, it must be the case that as long as $F_1^1(p) < D_1(p)$, we must have

$$F_1^2(p) \geq v_1^{rr}(p) = g_1 p + C(1-p)[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}}$$

with $v_1^{rr}(\tilde{p}_1) = v_1^{rs}(\tilde{p}_1)$.

Next, we prove that whenever $v_1^{rr}(p)$ and $v_1^{rs1}(p)$ intersect at a belief $p < 1$, it must be the case that $v_1^{rr'}(p) > v_1^{rs'}(p)$.

Since at the belief p, $v_1^{rr}(p) = v_1^{rs}(p)$, we have

$$g_1(p) + C_1^{rr}(1-p)[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}} = g_1 p + C_1^{rs}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

$$\Rightarrow C_1^{rr}[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}} = C_1^{rs}[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

$C_1^{rr} > 0$ and $C_1^{rs} > 0$.

This gives us

$$v_1^{rr'}(p) = g_1 - C_1^{rr}[\Lambda(p)]^{\frac{r}{(\lambda_1+\lambda_2)}}[1 + \frac{r}{(\lambda_1+\lambda_2)p}] = g_1 - C_1^{rs}[\Lambda(p)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{(\lambda_1+\lambda_2)p}]$$

$$> g_1 - C_1^{rs}[\Lambda(p)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{\lambda_1 p}] = v_1^{rs1'}(p)$$

Consider the function $v_1^{rr}(p)$ such that $v_1^{rr}(\tilde{p}_1) = v_1^{rs1}(\tilde{p}_1)$. From our above conclusion we can infer that at the right ε neighborhood of $\tilde{p}_1$, $v_1^{rr}(p) > v_1^{rs1}(p)$. Further, they cannot cross again since if they have to cross then $v_1^{rs1'}()$ would have to be strictly greater than $v_1^{rr'}()$. However, as argued, this is not possible. Thus, for all $p_s < p < 1$, $v_1^{rr}(p) > v_1^{rs1}(p)$. This implies that as long as $F_1^1(p) < D_1(p)$,

$$F_1^1(p) \geq v_1^{rr}(p) > v_1^{rs}(p)$$

Let $\hat{p}_1$ be the belief where player 1 switches arms for the first time before his payoff meets D_1. It must be the case that $\hat{p}_1 > \bar{p}_2$ and at $p = \hat{p}_1$, player 1 switches from the risky arm to the safe arm. From our arguments made above, we know that for $p > \hat{p}_1$, as long as player 1 is choosing the safe arm in equilibrium, player 1's payoff will always be above $\bar{v_1}^{rs}$.

Next, consider a situation when over a range of beliefs greater that $\hat{p}_1$, player 1's payoff is below the line D_1 and player 1 is choosing the risky arm. Then over this range of beliefs, player 1's payoff $v_1^{rs}(p)$ must be strictly above $\bar{v_1}^{rs}$. This is because at the belief when player 1 switched from the safe arm to the risky arm, $v_1^{rs}(p)$ and $F_1(p)$ were equal. This implies that at that belief $v_1^{rs}(p)$ was strictly greater than $\bar{v_1}^{rs}(p)$. Hence, for all $p < 1$, $v_1^{rs} > \bar{v_1}^{rs}(p)$. Thus, we have argued for beliefs $p > \hat{p}_1$, as long as the payoff function of player 1 is below D_1, it lies strictly above $\bar{v_1}^{rs}(p)$. Hence, if p_s^1 is the belief where player 1's payoff meets the line D_1, then $p_s^1 < p_1^{'}$. This concludes the proof of the lemma.

■

G Proof of lemma 2

Proof.

Consider an equilibrium in non-cutoff strategies. In equilibrium, before the payoff of player 2 meets the best response line D_2, if player 2 is choosing the risky (safe) arm then player 1 must be choosing the safe (risky) arm. Hence, if player 2 is choosing the risky arm then his payoff is given by

$$v_2^{rs}(p) = g_2 p + C_2^{rs}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

The integration constant C_2^{rs} is determined from the value matchign condition at the belief where player 2 switches to the risky arm from the safe arm.

On the other hand if player 2 is choosing the safe arm then his payoff is given by

$$F_2(p) = s + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p + C_2^{sr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

The integration constant C_2^{sr} is determined from the value matchign condition at the belief where player 2 switches from the risky arm to the safe arm.

Suppose at a belief $\tilde{p_2} > \bar{p_1}$ player 2's payoff is below the best response line D_2 and player 2 switches from the safe arm to the risky arm. Then, the payoff to 2 at and at the left neighborhood of $\tilde{p_2}$ is given by $F_2(p)$ and at the right neighborhood of $\tilde{p_2}$ is given by $v_2^{rs}(p)$ with $F_2(\tilde{p}) = v_2^{rs}(\tilde{p})$. We will now argue that as long as $p < \frac{s}{g_2}$, $v_2^{rs}(p) < F_2(p)$.

We first show that v_2^{rs1} is strictly convex.

Consider the function

$$\hat{v_2^{rs}}(p) = g_2 p + C_2^{rs}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

with C_2^{rs} being determined from $\hat{v_2^{rs}}(\tilde{p_2}) = s$. Since $\tilde{p_2} < \frac{s}{g_2}$, the integration constant will be strictly positive and hence the function will be strictly convex. By definition,

$$v_2^{rs} \equiv v_2^{rs}(p)$$

with $v_2^{rs}(\tilde{p}) = F_2^1(\tilde{p})$. Since $F_2(\tilde{p}) > s$, we must have the integration constant of v_2^{rs} to be strictly greater than that of $\hat{v_2^{rs}}$. Hence, v_2^{rs} is strictly convex.

We have

$$v_2^{rs}(p) = g_2 p + C_2^{rs}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

and

$$F_2(p) = s + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p + C_2^{sr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

This gives us

$$F_2'(\tilde{p_2}) = \frac{\lambda_1}{\lambda_1 + r}(g_2 - s) - C_2^{sr}[\Lambda(\tilde{p_2})]^{\frac{r}{\lambda_1}}[1 + \frac{r}{\lambda_1 \tilde{p_2}}]$$

and

$$v_2^{rs'}(\tilde{p}_2) = g_2 - C_2^{rs}[\Lambda(\tilde{p}_2)]^{\frac{r}{\lambda_2}}[1+\frac{r}{\lambda_2\tilde{p}_2}]$$

From the value matching condition at $p = \tilde{p}_2$, we have

$$C_2^{rs}[\Lambda(\tilde{p}_2)]^{\frac{r}{\lambda_2}} - C_2^{sr}(1-\tilde{p}_2)[\Lambda(\tilde{p}_2)]^{\frac{r}{\lambda_1}} = \frac{s+\frac{\lambda_1}{\lambda_1+r}(g_2-s)\tilde{p}_2 - g_2\tilde{p}_2}{(1-\tilde{p}_2)}$$

Next, since $[1+\frac{r}{\lambda_2\tilde{p}_2}] > [1+\frac{r}{\lambda_1\tilde{p}_2}]$ and $C_2^{sr} > 0$, we have

$$v_2^{rs'}(\tilde{p}_2) = g_2 - C_2^{rs}[\Lambda(\tilde{p}_2)]^{\frac{r}{\lambda_2}}[1+\frac{r}{\lambda_2\tilde{p}_2}] < g_2 - C_2^{rs}[\Lambda(\tilde{p}_2)]^{\frac{r}{\lambda_2}}[1+\frac{r}{\lambda_1\tilde{p}_2}] = \hat{v_2}$$

Hence,

$$F_2^{1'}(\tilde{p}_2) - \hat{v_2} = \{\frac{\lambda_1}{\lambda_1+r}(g_2-s) - g_2\} + \frac{[1+\frac{r}{\lambda_1}]\{-g_2\tilde{p}_2 + s + \frac{\lambda_1}{\lambda_1+r}(g_2-s)\tilde{p}_2\}}{(1-\tilde{p}_2)}$$

$$= \frac{r}{\lambda_1}[\frac{s}{\tilde{p}_2} - g_2] > 0$$

since $p_s < \frac{s}{g_2}$.

Since $F_2^1(p)$ is strictly concave and v_2^{rs1} is strictly convex, for all $p < \frac{s}{g_2}$, $F_2(p)$ will be strictly above $v_2rs1(p)$.

Let $\hat{p}_2 \in (\bar{p}_1, p_s^2)$ be the belief where player 2 switches arms for the first time. It must be the case that $\hat{p}_2 > \bar{p}_2$ and at $p = \hat{p}_2$, player 2 switches from the safe arm to the risky arm. From our arguments made above, we know that as long as $p > \hat{p}_2$, as long as player 2 is choosing the risky arm, player 2's payoff will be strictly below $\bar{F}_2(p)$.

Next, consider a situation when over a range of beliefs greater than $\hat{p}_2$, player 2's payoff is below the line D_2 and player 1 is choosing the safe arm. Then, over this range of beliefs, player 2's payoff $F_2(p)$ must be strictly below $\bar{F}_2(p)$. This is because at the belief when player 2 switched from the risky arm to the safe arm, $F_2(p) = v_2^{rs}(p)$. This implies that $F_2(p)$ was strictly greater than $\bar{F}_2(p)$. Hence, for all $p < 1$, $F_2(p) < \bar{F}_2(p)$. Thus, we have argued that for beliefs $p > \hat{p}_2$, as long as the payoff function of player 2 is below the line D_2, it lies strictly below $\bar{F}_2(p)$. Hence, $p_s^2 > p_2^{*n}$. This concludes the proof of the lemma

■

H Welfare comparison

Lemma 4 *Consider two equilibria in non-cutoff strategies and a range of beliefs* $[p_l, p_h]$. *In equilibrium* 1 *for all p in*$[p_l, p_h]$ *and at the left* ε *neighborhood of* p_l, *player* 1 *chooses the risky arm and player* 2 *chooses the safe arm. In equilibrium* 2*, for all* $p \in [p_l, p_h]$, *player* 1 *chooses the safe arm and player* 2 *chooses the risky arm. At* $p = p_l$, *if the aggregate payoff in equilirbium* 1 *is greater than or equal to that in equilibrium* 2*, then the aggregate payoff in equilibrium* 1 *is strictly higher than that in equilibrium* 2 *for all* $p \in (p_l, p_h]$.

Proof.

First consider the case when at $p = p_l$, the aggregate payoffs in both the equilibria are equal. We denote the aggregate payoff in equlibrium 1 and 2 by v_{12} and v_{21} respectively. When player 1 chooses the risky arm and player 2 chooses the safe arm, the aggregate equilibrium payoffs of the players are given by

$$v_{12} = s + [\frac{\lambda_1 g + rg_1}{\lambda_1 + r} - \frac{s\lambda_1}{\lambda_1 + r}]p + C_{12}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

When player 1 chooses the safe arm and player 2 chooses the risky arm, the aggregate equilibrium payoffs of the players are given by

$$v_{21} = s + [\frac{\lambda_2 g + rg_2}{\lambda_2 + r} - \frac{s\lambda_2}{\lambda_2 + r}]p + C_{21}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

We denote $[\frac{\lambda_1 g + rg_1}{\lambda_1 + r} - \frac{s\lambda_1}{\lambda_1 + r}] = A$ and $[\frac{\lambda_2 g + rg_2}{\lambda_2 + r} - \frac{s\lambda_2}{\lambda_2 + r}] = B$.

At $p = p_l$, the aggregate payoffs of the players are the same. This implies that at $p = p_l$ we have

$$(1-p_l)[C_{21}[\Lambda(p)]^{\frac{r}{\lambda_1}} - C_{12}[\Lambda(p)]^{\frac{r}{\lambda_1}}] = p[A-B]$$

$$\Rightarrow [C_{21}[\Lambda(p)]^{\frac{r}{\lambda_2}} - C_{12}[\Lambda(p)]^{\frac{r}{\lambda_1}}] = \frac{p_l}{1-p_l}[A-B] > 0$$

The derivatives of the aggregate payoffs are given by

$$v_{12}^{'} = A - C_{12}[\Lambda(p)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{\lambda_1}\frac{1}{p}]$$

$$v_{21}^{'} = B - C_{21}[\Lambda(p)]^{\frac{r}{\lambda_2}}[1 + \frac{r}{\lambda_2}\frac{1}{p}]$$

This gives us

$$v'_{12} - v'_{21} = [A - B] + C_{21}[\Lambda(p)]^{\frac{r}{\lambda_2}}[1 + \frac{r}{\lambda_2}\frac{1}{p}] - C_{12}[\Lambda(p)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{\lambda_1}\frac{1}{p}] > 0$$

This shows that for all $p \in (p_l, p_h)$, $v_{12} > v_{21}$.

Next, consider the case when at $p = p_l$, the aggregate payoff in equilibrium 1 is strictly higher than that in equilibrium 2. In that case let the aggregate payoff in equilibrium 2 be denoted by v^2_{21} where

$$v^2_{21} = s + [\frac{\lambda_2 g + rg_2}{\lambda_2 + r} - \frac{s\lambda_2}{\lambda_2 + r}]p + C^2_{21}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

Since $v^2_{21}(p_l) < v_{21}(p_l)$, we have $C^2_{21} < C_{21}$. This implies that for all $p \in (p_l, p_h]$ we have $v^2_{21} < v_{21}$. This proves that for all $p \in (p_l, p_h]$, $v_{12} > v^2_{21}$ ■

Lemma 5 *Consider two equilibria and a range of beliefs $[p_l, p_h]$. In one equilibrium (Equilibrium 1) both players choose the risky arm for all $p \in [p_l, 1]$. In the other equilibrium, both players choose risky arm for $p \in [p_h, 1]$ and only one player chooses the risky arm for $p \in [p_l, p_h]$. If the aggregate equilibrium payoff in equilibrium 1 at $p = p_l$ is at least as large as that in equilibrium 2, then the aggregate payoff in the former equilibrium is strictly higher than that in the later for all $p \in (p, 1)$*

Proof. We will show this only for the case when only one player chooses the risky arm, it is player 1 who does it. The other case can be shown in similar manner. First consider the case when $v_{rr}(p_l) = v_{sr}(p_l)$.

$$v_{rr} = gp + C_{rr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$
$$v_{sr} = g_1 p + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p + C_{sr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

From $v_{rr}(p_l) = v_{sr}(p_l)$, we have

$$C_{sr}[\Lambda(p)]^{\frac{r}{\lambda_1}} - C_{rr}[\Lambda(p)]^{\frac{r}{\lambda}} = \frac{p_l}{(1-p_l)}[\frac{rg_2 + \lambda_1 s}{\lambda_1 + r}] > 0$$

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

$$v_{sr}^{'} = g_1 + \frac{\lambda_1}{r+\lambda_1}(g_2 - s) - C_{sr}[\Lambda(p)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{\lambda_1}\frac{1}{p}]$$

This gives us

$$v_{rr}^{'} - v_{sr}^{'} = g_2 - \frac{\lambda_1}{r+\lambda_1}(g_2 - s) + C_{sr}[\Lambda(p)]^{\frac{r}{\lambda_1}}[1 + \frac{r}{\lambda_1}\frac{1}{p}] - C_{rr}[\Lambda(p)]^{\frac{r}{\lambda}}[1 + \frac{r}{\lambda}\frac{1}{p}]$$

Since $[1 + \frac{r}{\lambda_1}\frac{1}{p}] > [1 + \frac{r}{\lambda}\frac{1}{p}]$

$$v_{rr}^{'} - v_{sr}^{'} > 0$$

This proves that for all $p \in (p_l, 1)$, $v_{rr} > v_{sr}$.

■