Economics Department Discussion Papers Series

ISSN 1473 – 3307

The Role of Heterogeneity in a model of Strategic Experimentation

Kaustav Das

Paper number 17/03

URL Repec page: http://ideas.repec.org/s/exe/wpaper.html

The Role of Heterogeneity in a model of Strategic Experimentation *

Kaustav Das†

April 11, 2017

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, are heterogeneous 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 one risky arm. There are two players and each has access to an identical 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 Markov perfect equilibrium in simple cutoff strategies. For any degree of heterogeneity, there always exist equilibria where at least one player uses a non-cutoff strategy. When the equilibrium in cutoff strategies exists, it strictly dominates any other equilibria. When there does not exist any equilibrium in cutoff strategies, for some range of heterogeneity we can identify the welfare maximising equilibrium. For very low range of heterogeneity, the ranking among the equilibria is ambiguous.

*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 Siddhartha Bandyopadhyay, Martin Cripps, Johannes Horner, Jusso Valimaki and Ben Zissimos 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. Finally, thanks to Atisha Ghosh for reading through the draft.

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

JEL Classification Numbers:C73, D83, O31 **Keywords:** Two-armed Bandit, Heterogeneous Agents, Free-Riding, Learning

1 Introduction

Economic agents are often faced with situations where they have to strike an optimal balance between generating new information by trying out a new option on the one hand, and to continue with the current option of which they have complete knowledge. This leads to trade-offs between *exploration* and *exploitation*. For example, think of two universities which currently have a yearly examination system. Both of them are contemplating if the student ratings will be positively affected by shifting to a semestral examination system. This means, a university has to decide which of these two hypotheses to investigate. If both these universities get a similar kind of student on average, then once one of them finds out that the semestral examination system positively affects student ratings, both benefit from this information generation. In this case, although the cost of trying out a new option has to be privately borne, the potential benefit generated from that has the nature of a public good. This means, if a university generates new information on the efficacy of the semester system, the other university can *free ride* on it. This gives rise to the strategic element in agents' decisions to try out a new option. In this paper, I study the behaviour of agents in such a setting under a situation where economic agents differ with respect to their abilities in exploring the new option. With respect to the university example described, this means that two universities might differ in their administrative capabilities. If the semester system is actually better, a univeristy with a better administrative capability will be able to learn this sooner than the one which has relatively worse administrative capability. In this paper, I develop a stylised model to consider such situations.

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 a 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 experiences of other agents. In the present paper, by appropriately modifying the two-armed bandit framework of ([4]), I analyse how heterogeneous players behave when they face trade-offs between exploration and exploitation in the presence of informational externalities. The players are heterogeneous in the sense that given the risky arm is good, expected time required to learn this differs among players. Players start with a common prior, which is the probability that the risky arm is good. Since, all actions and outcomes are perfectly observable, they always have a common posterior belief. We 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 exists an equilibrium where both players use a cutoff strategy. That is, each player exercises the risky option (or arm) only when the likelihood he attributes to the option being good is greater than a threshold. Otherwise, he carries on with the option of which he has complete knowledge (safe arm). This equilibrium is unique in the class of equilibria where players use cutoff strategies. In this equilibrium, whenever only one of the players experiments and the other free rides, it is always the player with lower innate ability who free rides. With respect to the aggregate payoff, the equilibrium in cutoff strategies strictly dominates any other equilibrium. If the degree of heterogeneity is low, all equilibria are in non-cutoff strategies. From this set of equilibria, we can identify an equilibrium in which the weaker player free rides for the largest possible range of beliefs. We denote this equilibrium as the *most heterogeneous* equilibrium. As the degree of heterogeneity increases, this equilibrium coincides with the equilibrium in cutoff strategies. Within this low range of heterogeneity, we can identify two sub ranges. At the higher end of this range, the *most heterogeneous equilibrium* strictly dominates any other equilibrium. However, at the lower end of this range, the ranking between equilibria is ambiguous. Hence, except for very low range of heterogeneity, we can always identify the welfare maximising equilibrium. This is unlike the model with homogeneous players ([4]) in which, even though we know how to improve the equilibrium welfare, we are not able to identify the welfare maximising equilibrium.

The formal 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. If a player accesses the good risky arm, he 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 implies that 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 while 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 $g_1 > g_2 > s$. In the basic model of the paper, at any point of time, a player can choose only one of the arms.[1] Players start with a common prior, which is the probability with which the risky arm is good.

[1]In the paper there is a section which analyses the case when players can diversify efforts between two arms.

All actions and outcomes are publicly observable. Based on these, players update their beliefs using Bayes' rule. This implies that players always have a common posterior belief.

We start by examining the social planner's problem which aims to maximise the sum of the expected payoffs of the players. The planner, in continuous time, decides to allocate players to one of the arms. The social optimum involves *specialisation* at the extremes and *diversification* for interim range of beliefs. This means that if it is very 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. Similarly, 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 interim range of beliefs, the weaker player (player 2) is allocated to the safe arm and the stronger player (player 1) is allocated to the risky arm.

For the analysis of the non-cooperative solutions, we restrict ourselves to Markovian strategies with the common posterior belief as the state variable. All equilibria are inefficient. We characterise the set of Markov perfect equilibria for all possible kinds of heterogeneity between the players. In all these equilibria, players use simple strategies. A simple strategy implies that for a particular belief, a player chooses exclusively one or the other arm. A common feature of any equilibrium is that both the beliefs where all experimentation ceases and below which only one of the players experiments are higher than the corresponding beliefs in the planner's solution. This is due to free riding by the weaker player (player 2) and also due to the fact that player 1 does not internalise player 2's benefit from his own experimentation.

We first show that there exists a threshold λ_2^* such that if λ_2 is lower than this threshold, there exists an equilibrium where both players use simple cutoff strategies, i.e., use the risky arm exclusively when the probability assigned to the risky arm being good is above some threshold or cutoff. This equilibrium is unique in the class of equilibria where players use cutoff strategies. There also exist other equilibria in non-cutoff strategies. However, the aggregate equilibrium payoff is highest for the equilibrium in cutoff strategies.

Next, we analyse the situation when the degree of heterogeneity between the players is low. In this case, all equilibria are in non-cutoff strategies. Within these equilibria, we can identify a particular equilibrium where player 2 does not choose the risky arm unless it is his dominant action to do so. This implies that in this equilibrium, player 2 free rides for the largest possible range of beliefs. We denote this equilibrium as the *most heterogeneous equilibrium*. As the degree of heterogeneity increases, this equilibrium coincides with the equilibrium in cutoff strategies. Within this low range of heterogeneity, we can further identify two sub ranges. At the higher end of this range, the aggregate equilibrium payoff is highest in the most heterogeneous equilibrium. At the lower end however, the welfare ranking among the equilibria is ambiguous.

One important takeaway from this paper is that with heterogeneous players, unless the degree of heterogeneity is low enough, we can always identify the welfare maximising equilibrium. This is unlike the model with homogeneous players ([4]). In that case, whenever in equilibrium only one player can experiment, we know that by making the players take turns in experimenting more often, the equilibrium welfare is improved. However, although it is possible to say that welfare improves with more frequent turn-taking, it is not possible to identify the welfare maximising equilibrium. Also, in the case with heterogeneous players, equilibrium behaviour in the welfare maximising equilibrium (whenever it is possible to identify) is *simplified* compared to the equilibrium behaviour which improves welfare in the model with homogeneous players. Further, as stated above, for very low degree of heterogeneity, the ranking between the equilibria is ambiguous. Hence, as soon as we introduce some heterogeneity, it is no longer true that we unambiguously gain by making players take turns in experimenting more often.

Finally, we consider two variants of the basic model. First, we analyse the case where players have the same innate ability along the risky arm but different safe arm payoffs. We show that, like the basic model, when players are sufficiently heterogeneous, an equilibrium in cutoff strategies exists and it is unique in the class of equilibria where players use cutoff strategies. A priori, this is not obvious. Innate ability of a player along the risky arm directly affects the rate at which learning takes place. The safe arm payoff does not directly affect the rate of learning. However, both these kinds of heterogeneity make the free riding opportunities different, and like in the basic model, an increase in heterogeneity shrinks (expands) the free riding opportunities of player 1 (2). This similar effect on free riding opportunities gives rise to the results.

Second, we analyse the basic model with the additional feature that players can diversify efforts between the arms. This gives rise to additional equilibria where players over some range of beliefs diversify efforts between the arms. We show that for sufficiently high degree of heterogeneity, welfare wise these equilibria are strictly dominated by the equilibrium in cutoff strategies.

Related Literature: This paper contributes to the strategic bandit literature. Some of the works that 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]). A comprehensive survey on the recent developments in learning, experimentation and strategic interactions is available in Horner and Skrzypacz ([3]). The paper which is closest to the present paper is Keller, Rady and Cripps ([4]). They find that an equilibrium in cutoff strategies never exists. The present work contributes in two ways. First, I show that with heterogeneous players, it is not only possible to have an equilibrium in cutoff strategies, but also it is welfare maximising and strictly dominates all other equilibria. Further, except for a very low degree of heterogeneity, it is always possible to identify the welfare maximising equilibrium. Finally, I find that as soon as some heterogeneity is introduced, it is no longer true that the aggregate equilibrium welfare for all beliefs can be increased by making players take turns in experimenting more often.

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. The safe arm can be accessed by only one player at a time. 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, unlike Thomas ([9]) where types are stochastically independent, the type of risky arm in the present paper is the same for both players. 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 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 in a bandit model with a safe arm and a risky arm, heterogeneity between the players in most of the cases allows us to identify the welfare maximising equilibrium.

Tonjes([10]) in her work analyses the effect of introducing heterogeneity in a model of strategic experimentation and discusses the possibility of having the results obtained formally in this paper. Further, she mentions about existence of equilibria in non-cutoff strategies under the situation when the degree of heterogeneity between the players is not high enough. However, as shown in the current paper, these equilibria always exist.

The rest of the paper is organised as follows. Section 2 lays down the details of the basic model with heterogeneous players and discusses the social planner's solution. A detailed analysis of equilibria for different ranges of heterogeneity is undertaken in section 3. Section 4 discusses two variants of the basic model. Finally, section 5 concludes the paper.

2 Two armed bandit model with heterogeneous players

The Basic 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 a lump sum payoff 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 $\lambda_1 > \lambda_2$. This implies that 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 information 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 have[2]

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

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

Simplifying the 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

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

having player i on the risky arm. 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 the planner's optimal policy $k^* = (k_1^*, k_2^*)$ is given by*

$$(k_1^*, k_2^*) = \begin{cases} (1,1) & : \quad \textit{if } p \in (p_2^*, 1) \\ & : \\ (1,0) & : \quad \textit{if } p \in (p_1^*, p_2^*] \\ & : \\ (0,0) & : \quad \textit{if } p \in (0, p_1^*] \end{cases}$$

and the value function is

$$v(p) = \begin{cases} gp + \{\frac{\frac{\lambda_1+\lambda_2}{\lambda_2}s - gp_2^*}{(1-p_2^*)[\Lambda(p_2^*)]^{\frac{r}{\lambda}}}\}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}} \equiv v_{rr} & : \quad \textit{If } p \in (p_2^*, 1], \\ & : \\ s + [\frac{\lambda_1 g + r g_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p + \{\frac{s - [\frac{\lambda_1 g + r g_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p_1^*}{(1-p_1^*)[\Lambda(p_1^*)]^{\frac{r}{\lambda_1}}}\}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}} \equiv v_{sr} & : \quad \textit{if } p \in (p_1^*, p_2^*], \\ & : \\ 2s & : \quad \textit{if } p \in (0, p_1^*]. \end{cases}$$

p_1^* is given as

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

p_2^* is such that

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

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 $k_2 = 0$. 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:

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

where $g = (g_1 + g_2)$; $\Lambda(p) = \frac{(1-p)}{p}$ and C_{rs}^P 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

$$s + [\frac{\lambda_1 g + rg_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p_1^* + C_{rs}^P(1-p_1^*)[\Lambda(p_1^*)]^{\frac{r}{\lambda_1}} = 2s$$

$$\Rightarrow C_{rs}^P = \frac{s - [\frac{\lambda_1 g + rg_1}{\lambda_1 + r} - \frac{s\lambda_1}{r+\lambda_1}]p_1^*}{(1-p_1^*)[\Lambda(p_1^*)]^{\frac{r}{\lambda_1}}}$$

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

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

Substituting the value of C_{rs}^P we have

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

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

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

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

where $g=g_1+g_2$ and $\lambda=\lambda_1+\lambda_2$. C_r^P is the integration constant.

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

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

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

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

This implies that $b_2(p_2^*,v)$ can be written as

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

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

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

This is because $\lambda_1>\lambda_2$. Let $v_{sr}(.)$ be the representation of the value function when 1 is at the risky arm and 2 is at the safe arm and v_{rr} be the same when both players are at the risky arm. 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

$$[\frac{\lambda_1g+rg_1}{\lambda_1+r}-\frac{s\lambda_1}{r+\lambda_1}]p_2^*+[\frac{s-[\frac{\lambda_1g+rg_1}{\lambda_1+r}-\frac{s\lambda_1}{r+\lambda_1}]p_1^*}{(1-p_1^*)[\Lambda(p_1^*)]^{\frac{r}{\lambda_1}}}](1-p_2^*)[\Lambda(p_2^*)]^{\frac{r}{\lambda_1}}=\frac{\lambda_1}{\lambda_2}s \tag{1}$$

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 and monotonically increasing in p_2^*, there exists a unique $p_2^* \in (p_1^*, 1)$, such that (1) holds.

The integration constant of v_{rr} is given by

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

The obtained value function is

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

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

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

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 risky arm decreases. Hence, player 1 is shifted at a higher belief.

The planner's solution is depicted in the 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.

In the next section, we carry out the analysis of the non-cooperative game.

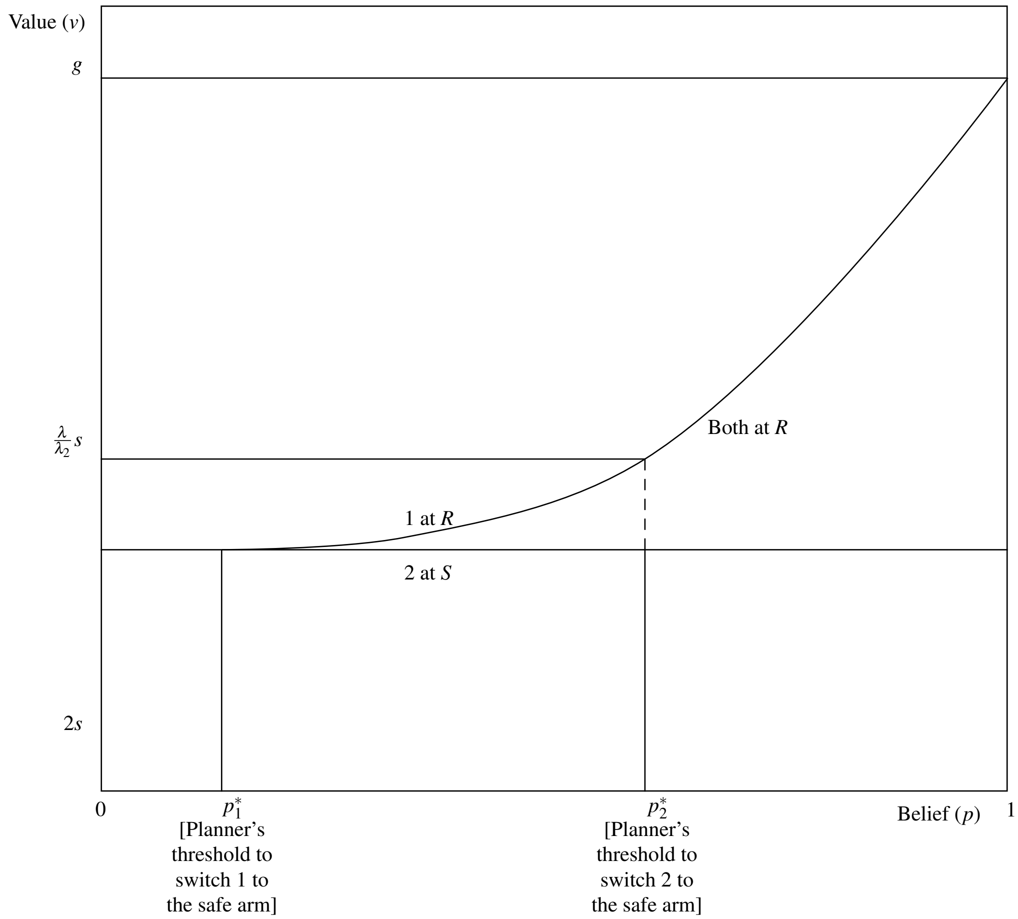

Figure 1.

3 Non-cooperative game

In this section, 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

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

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 interpret 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

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

This means 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. This means if k_2 is always equal to 0, 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 player 1'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 player 1 is always choosing the safe arm, 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 responses of the players are depicted in figure 2.

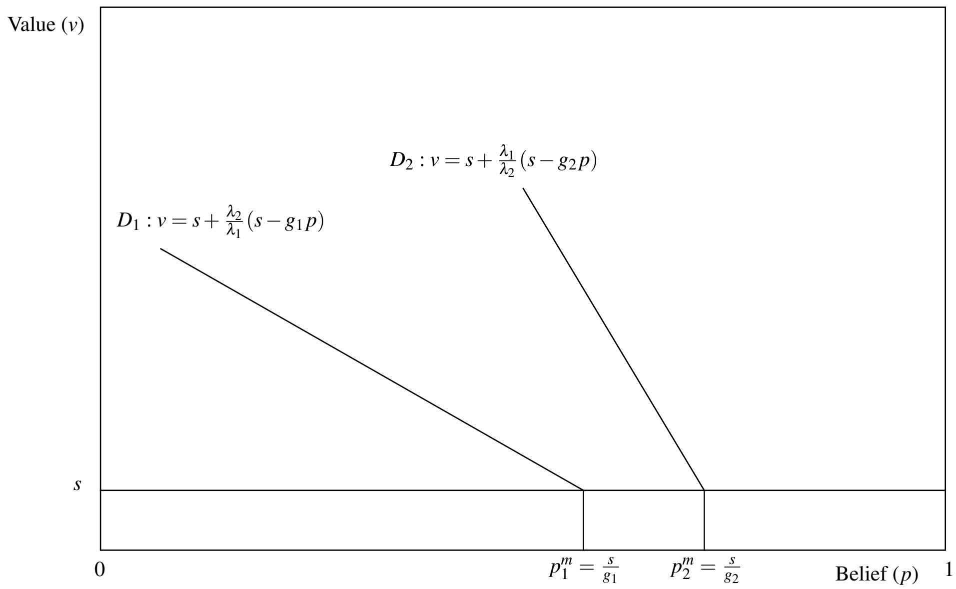

Figure 2.

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. In this region, only player 2 has free riding opportunities.

Payoffs: Before we discuss equilibrium formally, we obtain explicit solutions for the payoffs of 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 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, the ODE which v_i^{rs} should satisfy is

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

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[3] and $\Lambda(p) = \frac{1-p}{p}$.

We will now argue that no efficient equilibrium exists. To show this, we will 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 $p_l < \bar{p}_i$ and p_l is the belief where all experimentation along the risky arm ceases, 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

[3]This will take a unique value for any particular function.

solution experimentation occurs till the belief reaches the point p_1^* and $p_1^* < \bar{p_1}$. This implies that no equilibrium is efficient.

In the next subsection, we will describe the condition under which an equilibrium in simple cutoff strategies exists and will fully characterise it.

3.1 Equilibrium in cutoff strategies

In this subsection, we will identify the condition under which an equilibrium in simple cutoff strategies exists. This is unlike the case with homogeneous players ([4]) in which there does not exist any equilibrium in simple cutoff strategies. A player is said to follow a simple cutoff strategy if the risky arm is chosen for beliefs above a particular threshold and the safe arm is chosen for beliefs less that or equal to that threshold.

In the previous subsection, we have argued that in any equilibrium, all experimentation ceases at the belief $\bar{p_1}$. We will now argue that in any equilibrium, at the right ε- neighbourhood ($\varepsilon \rightarrow 0$) of $\bar{p_1}$, only player 1 can experiment along the risky arm and player 2 will always free ride. This means that player 1 is always the last player to experiment in any equilibrium.

Suppose, at the right ε- neighbourhood 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-neighbourhood 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 $\bar{p_1}$ is the lowest belief at which some experimentation takes place in any equilibrium. In that case, value of player 2 at $p = \bar{p_1}$ is equal to s. Since $\bar{p_1} < \bar{p_2}$, the continuation value to player 2 from playing the risky arm at the right-ε neighbourhood of $\bar{p_1}$ becomes less than s (refer to appendix (B) for a detailed proof). Hence, it is not an optimal response on player 2's part. This implies that in any equilibrium, player 1 is always the last player to do any experimentation along the risky arm. Formally, this means that a common characteristic of any equilibrium is that there exists a range of beliefs $(\bar{p_1}, p)$ ($p > 0$ and can be arbitrarily close to $\bar{p_1}$) such that player 1 experiments and player 2 free rides in this range.

The above arguments imply that for beliefs right above $\bar{p_1}$, in any equilibrium, player 1's payoff is given by the following function

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

with $C_1^{rsc} = \frac{s - g_1\bar{p}_1}{(1-\bar{p}_1)[\Lambda(\bar{p}_1)]^{\frac{r}{\lambda_1}}}$

Player 2's equilibrium payoff for these beliefs is given by

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

with $C_2^{rsc} = -\frac{\frac{\lambda_1}{r+\lambda_1}(g_2-s)\bar{p}_1}{(1-\bar{p}_1)[\Lambda(\bar{p}_1)]^{\frac{r}{\lambda_1}}}$

Since $C_1^{rsc} > 0$ and $C_2^{rsc} < 0$, v_1^{rsc} is a strictly convex function and F_2^c is a strictly concave function. The following lemma shows that these functions intersect the corresponding best response lines at a unique belief.

Lemma 1 *There exists a unique $p_1^{'} \in (\bar{p}_1, 1)$ and a unique $p_2^{*n} \in (\bar{p}_1, 1)$ such that $v_1^{rsc}(p_1^{'}) = D_1(p_1^{'})$ and $F_2^c(p_2^{*n}) = D_2(p_2^{*n})$*

Proof of Lemma. $v_1^{rsc}(p)$ is strictly convex and strictly increasing in p. On the other hand, D_1 is monotonically decreasing in p. At $p_1^{'} = \bar{p}_1$, $v_1^{rsc} < D_1$. At $p_1^{'} = 1$, $v_1^{rsc} > D_1$. Since both $v_1^{rsc}(p)$ and $D_1(p)$ are continuous and strictly monotonic functions, there exists a unique $p_1^{'} \in (\bar{p}_1, 1)$ such that $v_1^{rsc}(p_1^{'}) = D_1(p_1^{'})$.

$F_2^c(p)$ is strictly concave and strictly increasing in the range $p \in (\bar{p}_1, 1)$. At $p_2^{*n} = \bar{p}_1$, $F_2^c(p_2^{*n}) = s < D_2(p_2^{*n})$. At $p_2^{*n} = 1$, $F_2^c(p_2^{*n}) > D_2(p_2^{*n})$. Since, both $F_2^c(p)$ and $D_2(p)$ are continuous and monotonic functions in the range $p \in (\bar{p}_1, 1)$, and $D_2(p)$ is strictly decreasing in p, there exists a unique $p_2^{*n} \in (\bar{p}_1, 1)$ such that the curve of the function $F_2^c(p)$ intersects the line D_2, i.e $F_2^c(p_2^{*n}) = D_2(p_2^{*n})$.

This concludes the proof of the lemma.

■

Next, we argue that for any λ_1, λ_2 such that $\lambda_1 > \lambda_2 > \frac{s}{h}$, the belief at which the function $F_2^c(p)$ intersects the line D_2 is strictly higher than $\bar{p}_2$. This is shown in the following lemma.

Lemma 2 *For any $\lambda_1 < \lambda_2 < \frac{s}{h}$, we have $p_2^{*n} > \bar{p}_2$*

Proof.

The integration constant of $F_2^c(p)$ is strictly negative. This implies $F_2^c(\bar{p}_2) < s + \frac{\lambda_1}{\lambda_1+r}[g_2 - s]\bar{p}_2$. From the expression of $\bar{p}_2$, we then have

$$F_2^c(\bar{p}_2) < s + \frac{\lambda_1}{\lambda_1+r}[g_2 - s]\bar{p}_2 = s + \frac{\lambda_1}{\lambda_1+r}[g_2 - s]\frac{\mu_2 s}{(\mu_2+1)g_2 - s} \equiv f$$

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, $D_2(\bar{p_2}) > F_2^c(\bar{p_2})$. As argued before, $F_2^c(p)$ is strictly increasing and $D_2(p)$ is strictly decreasing in p. Since $D_2(p_2^{*n}) = F_2^c(p_2^{*n})$, we can conclude that $p_2^{*n} > \bar{p_2}$. This concludes the proof of the lemma.

■

We will now show that if the degree of heterogeneity between the players is high enough, then there exists an equilibrium where both players use simple cutoff strategies. A player is said to use a cutoff strategy if he uses the risky arm exclusively when the probability assigned to the risky arm being good is above some threshold or cutoff. For probabilities less than or equal to that threshold, the player uses the safe arm. From now on, we will refer to this kind of equilibrium as equilibrium in cutoff strategies. The following proposition describes this.

Proposition 2 *There exists a $\lambda_2^* \in (\frac{s}{h}, \lambda_1)$ such that if $\lambda_2 \in (\frac{s}{h}, \lambda_2^*)$, there exists an equilibrium where both players use simple cutoff strategies. In this equilibrium, player 1 experiments along the risky arm as long as the belief is strictly greater than $\bar{p}_1$ and reverts to the safe arm otherwise. For beliefs strictly greater than p_2^{*n}, player 2 experiments along the risky arm and reverts to the safe arm otherwise. This equilibrium is unique in the class of equilibria where both players use cutoff strategies.*

Proof.

The formal proof of this proposition is relegated to appendix (C). Here we provide a sketch of the proof with the help of figure 3. In this picture, the curve v_2 represents the function F_2^c (which turns out to be the equilibrium payoff of player 2 in the equilibrium in cutoff strategies) and the curve v_1 represents the function v_1^{rsc} (which turns out to be the equilibrium payoff of player 1 in the equilibrium in cutoff strategies). There are three regions which can be identified in this picture. In the area above D_2, experimenting along the risky arm is the dominant action for both players. In the area between the lines D_1 and D_2, only player 2 has free riding opportunities and experimenting along the risky arm is the dominant action for player 1. Finally, both players have free riding opportunities in the area below the line D_1.

We will first argue that for an equilibrium in cutoff strategies to exist, it must be the case that $p_2^{*n} > p_1^{'}$. It has been argued that in any equilibrium, at the right ε-neighbourhood of $\bar{p}_1$, only player 1 can experiment. This implies that if an equilibrium in cutoff strategies exists, then player 1 should experiment along the risky arm for all beliefs greater than $\bar{p}_1$. Given this, player 2 finds it beneficial to remain on the safe arm as long as his payoff in the (v, p) plane is below the best response line D_2. This means player 2 finds it beneficial to remain on the safe arm as long as the belief is less than p_2^{*n}. For both players, as the payoff function intersects the best response line, experimenting along the risky arm becomes the dominant action. Thus, if an equilibrium in cutoff strategies exists then it has to be true that for beliefs greater than p_2^{*n}, player 1 should not have any incentive to revert to the safe arm. This implies that for an equilibrium in cutoff strategies to exist, it should be the case that the belief $p_1^{'}$, at which the curve v_1^{rsc} intersects the line D_1 is lower than p_2^{*n}.

Next, we will argue that p_2^{*n} is greater than $p_1^{'}$ only when λ_2 is lower than a particular threshold. If λ_2 is very close to λ_1, then D_1 and D_2 almost coincide. In that case, as in the game with homogeneous players, $p_1^{'} > p_2^{*n}$. As λ_2 decreases, the line D_1 pivots towards left and hence, there is a decline in $p_1^{'}$. On the other hand, as λ_2 decreases, the line D_2 shifts right and the curve F_2^c becomes flatter. This implies that there is an increase in p_2^{*n}. As $\lambda_2 \rightarrow \frac{s}{\lambda_2}$, $p_2^{*n} \rightarrow 1$. On the other hand, $p_1^{'}$ is always between $\bar{p}_1$ and $\frac{s}{g_1}$. Thus, there exists a $\lambda_2^* \in (\frac{s}{\lambda_2}, \lambda_1)$, such that if $\lambda_2 < \lambda_2^*$, $p_2^{*n} > p_1^{'}$.

Finally, given that player 1 is the last one to experiment, in an equilibrium in cutoff strategies, strategy of player 1 is always fixed. In that case, player 2 also has a unique best response. This implies that the equilibrium characterised is unique in the class of equilibria where players use cutoff strategies.

■

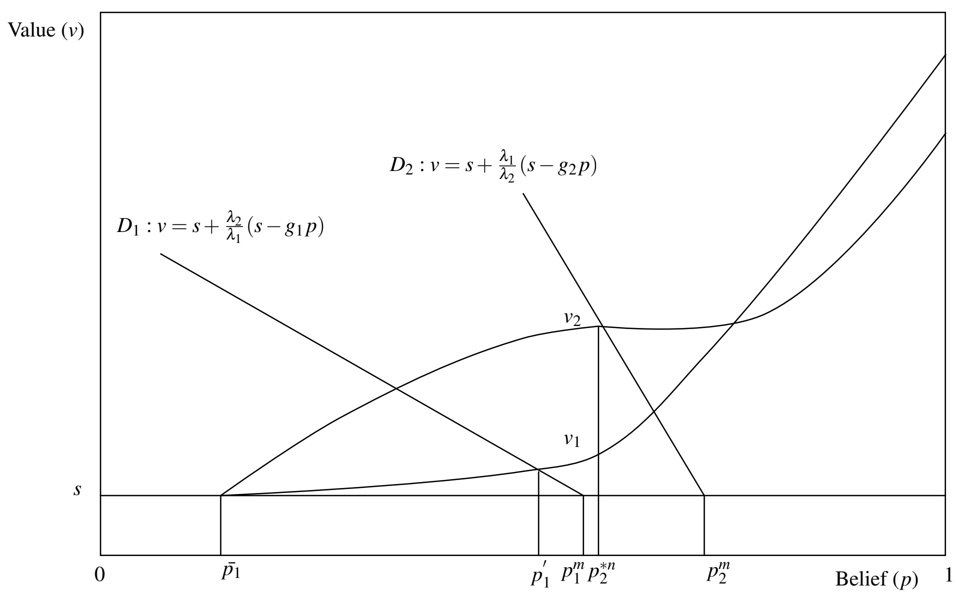

Figure 3.

The intuition behind the result described in the preceding proposition can be understood as follows. In any equilibrium, player 1 is the last player to experiment at beliefs right above $\bar{p}_1$. Hence, in order to guarantee the existence of an equilibrium in cutoff strategies, there should be a sufficient range of beliefs where only player 2 has free riding opportunities. This will ensure that experimenting along the risky arm becomes a dominant action for player 2 at a belief when player 1 has no incentive to free ride. In the case of homogeneous players, free riding opportunities are the same. This implies that a player directly moves from a region where both can free ride to his own dominant action region. Since, players cross over from the free riding region to the dominant action region at different beliefs, it is always the case that whenever a player crosses over, the other player is still in the free riding region. Further, the last player to experiment can never be the first player to cross over. This is the reason why there is no equilibrium in cutoff strategies. As explicitly shown in [4], in any equilibrium, for some range of beliefs players have to take turns in experimenting. However, as players start becoming heterogeneous, their free riding opportunities start becoming different. In particular, the set beliefs over which player 1 can free ride shrinks[4] and that for player 2 expands. Hence, there emerges a set of beliefs where only player 2 can free ride. As the extent of heterogeneity between the players exceeds a threshold, the set of beliefs where only player 2 can free ride expands to an extent such that player 2 who free rides at beliefs right above $\bar{p}_1$, moves to the dominant action region at a belief when player 1 has already moved to his dominant action region. This ensures the existence of the equilibrium in cutoff strategies.

We conclude this subsection by providing the formal characterisation of the equilibrium in cutoff strategies. For $p \in [0, \bar{p}_1]$, both players' payoffs are equal to s. For $p \in (\bar{p}_1, p_2^{*n}]$, player 1's payoff is given by $v_1^{rsc}(p)$ and that of player 2 is given by $F_2^c(p)$. For $p > p_2^{*n}$, player i's ($i = 1,2$) payoff is given by

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

where C_1^{rrc} is determined from $v_1^{rrc}(p_2^{*n}) = v_1^{rsc}(p_2^{*n})$ and C_2^{rrc} is determined from $v_2^{rrc}(p_2^{*n}) = F_2^c(p_2^{*n})$.

In the following subsection, we will characterise equilibria in which at least one of the players uses non-cutoff strategy. We will compare these equilibria with the one in cutoff strategies.

3.2 Equilibria in non-cutoff strategies

In the previous subsection, we have identified the condition under which there is an equilibrium where both players use cutoff strategies. In the current subsection we will discuss about equilibria where players use non-cutoff strategies. We begin our analysis on this by arguing that for any degree of het-

[4]It never becomes empty.

erogeneity between the players, there always exist equilibria where players use non-cutoff strategies. The following lemma describes this.

Lemma 3 *For any $\lambda_2 \in (\frac{s}{h}, \lambda_1)$, there exist equilibria where both players use non-cutoff strategies.*

Proof.

In any equilibrium, at the right ε-neighbourhood of $\bar{p}_1$, only player 1 can experiment along the risky arm. This implies that in order to prove this lemma, we need to show that in any equilibrium, there always exists a range of beliefs over which player 2 experimenting along the risky arm and player 1 free riding by choosing the safe arm constitute mutual best responses. If this is true then we can make players take turns in experimenting along the risky arm before their payoffs intersect their respective best response lines. This will guarantee the existence of an equilibrium where players use non-cutoff strategies. We show this in following two steps.

The first step is to show that for any degree of heterogeneity between the players, there exist free riding opportunities for player 1. This follows from the fact that for all values of $\lambda_2 \in (\frac{s}{h}, \lambda_1)$, the best response line D_1 is always above the line $D_1^L = s + \frac{\frac{s}{h}}{\lambda_1}(s - g_1 p)$. Since D_1^L is negatively sloped and intersects the horizontal line $v = s$ at $p = \frac{s}{g_1} > \bar{p}_1$, for any degree of heterogeneity, there is free riding opportunities for player 1.

The next step shows that for any belief p if player 2's equilibrium payoff is greater than s and below the best response line D_2, then given player 1 is choosing the safe arm, it is the best response of player 2 to choose the risky arm. The formal proof is given in appendix (D). The intuition behind this is as follows. Since at p the payoff to player 2 is greater than s, it must be the case that at beliefs just below p, some player is experimenting. In that case, given that player 1 is choosing the safe arm, player 2 has incentive to choose the risky arm. If he does not experiment, then nothing happens and he gets a flow payoff of s. On the other hand, if he experiments then there will either be a breakthrough or not. In the first case the payoff is strictly higher than s and in the latter case the belief is going to drift downwards and lead to experimentation by some other players which in turn will give a flow payoff greater than s. This shows that the expected payoff to player 2 by experimenting at p is greater than s. This is true even if the belief $p < \bar{p}_2$, the single person threshold of player 2. This shows that there is some kind of *encouragement* effect possible for player 2 in equilibrium. However, we will see later that this encouragement effect for player 2 need not be welfare maximising. From our analysis this far, we know that in any equilibrium there is a range of beliefs starting from $\bar{p}_1$, where player 1 experiments and 2 free rides. Hence, we can find a range of beliefs where both players' payoffs are greater than s and below their respective best response lines. This shows that there are beliefs where player 2 experimenting and player 1 free riding can be mutual best responses. This concludes the proof of the lemma. ■

From the above lemma it can be inferred that in any equilibrium in non-cutoff strategies, when both players have free riding opportunities, players switch arms at least once. We will now characterise these equilibria in non-cutoff strategies. First we will consider the case when the equilibrium in cutoff strategies exists.

3.2.1 Characterisation of equilibria in non-cutoff strategies when the equilibrium in cutoff strategies exists

Suppose $\lambda_2 \in (\frac{s}{h}, \lambda_2^*)$. From our analysis this far, we know that the equilibrium in cutoff strategies exists and there are other equilibria where players use non-cutoff strategies. We will first show that compared to the equilibrium in cutoff strategies, the intensity of experimentation is always lower in any equilibrium in non-cutoff strategies. The following lemma will help us to infer this.

Lemma 4 *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 and $p_s^2 > \bar{p}_1$ be the belief at which the payoff of player 2 meets the line D_2. Then, we have $p_s^1 < p_1^{'}$ and $p_s^2 > p_2^{*n}$.*

The proof of the lemma is relegated to appendix (F). From the above lemma, it can be inferred that the intensity of experimentation in the equilibrium in cutoff strategies is always higher than that in any equilibrium in non-cutoff strategies. This can be observed as follows. In the equilibrium in non-cutoff strategies, both players experiment for beliefs above p_2^{*n}, both use safe arms for beliefs below $\bar{p}_1$ and for the beliefs in between, player 1 experiments and player 2 free rides. Consider any equilibrium in non-cutoff strategies. From lemma (4) we know that $p_s^1 < p_s^2$. Hence, for all beliefs in $(\bar{p}_1, p_s^2)$ only one of the players experiment along the risky arm. In both kinds of equilibrium, all experimentation ceases at the belief $\bar{p}_1$. Since, $p_s^2 > p_2^{*n}$, the range of beliefs over which both players experiment along the risky arm is higher in the equilibrium in cutoff strategies. Further, in the equilibrium in cutoff strategies, whenever only one player experiments, it is player 1 who does so. However, in the equilibrium in non-cutoff strategies, when only one player chooses the risky arm, players take turns in experimenting. This shows that the intensity of experimentation in these equilibria is unambiguously lower than that in the equilibrium in cutoff strategies.

The following proposition characterises a generic equilibrium in non-cutoff strategies.

Proposition 3 *Any equilibrium in non-cutoff strategies where players switch arms k times ($k \geq 1$ and k finite) in the region where both can free ride, can be characterised by thresholds $\{\tilde{p}_j\}_{j=1}^{j=k}$, p_s^1 and p_s^2. $\bar{p}_1 < \tilde{p}_1 < \tilde{p}_2 .... < \tilde{p}_k < p_s^1 < p_s^2$. $\tilde{p}_1$ is the belief where players switch arms for the first time and at the beliefs $\{\tilde{p}_j\}$ ($j = 2,..,k$) players further switch arms. Let v_1 and v_2 be the equilibrium payoffs of 1 and 2 respectively. p_s^1 is the belief where v_1 intersects D_1 and p_s^2 is the belief where v_2 intersects D_2. For $p \in (\bar{p}_1, \tilde{p}_1)$, $v_1 = v_1^{rsc}$ and $v_2 = F_2^c$. For $p \in (\tilde{p}_j, \tilde{p}_{j+1}]$ ($1 \leq j \leq k-1$) if player i ($i = 1,2$) chooses the safe arm then $v_i = F_i^j = s + \frac{\lambda_l}{\lambda_l + r}(g_i - s) + C_i^j(1-p)[\Lambda(p)]^{\frac{r}{\lambda_l}}$, where ($l = 1,2; l \neq i$) C_i^j is determined from $F_i^j(\tilde{p}_j) = v_i(\tilde{p}_j)$. If player i chooses the risky arm then $v_i = v_i^j = g_i p + C_i^j(1-p)[\Lambda(p)]^{\frac{r}{\lambda_i}}$. C_i^j is determined from $v_i^j(\tilde{p}_j) = v_i(\tilde{p}_j)$. If j is odd(even), then for $p \in (\tilde{p}_j, \tilde{p_{j+1}}]$, player 1 chooses the safe (risky) and player 2 chooses the risky (safe) arm. If k is odd then at p_s^1 player 1 switches to the risky arm and player 2 switches to the safe arm.*

Appendix (G)lays out the formal proof of the above proposition. It can be observed that p_s^1 and p_s^2 are determined endogeneously and their values depend on how many times players switch arms.

The comparison between the equilibrium in cutoff strategies and an equilibrium in non-cutoff strategies is depicted in figure $4(a)$. Here we have restricted ourselves to those equilibria in non-cutoff strategies where before experimenting along the risky arm becomes the dominant action, players switch arms only once[5]. Figure $4(b)$ and $4(c)$ depict the actions of players in the equilibrium in cutoff strategies and equilibrium in non-cutoff strategies respectively. These equilibria correspond to the ones depicted in figure $4(a)$.

The black curves v_1 and v_2 depict the payoffs to player 1 and 2 respectively in the equilibrium in cutoff strategies. In the equilibrium in non-cutoff strategies, payoffs are same as before for beliefs less than or equal to $\tilde{p}_1$. At $\tilde{p}_1$, players switch arms. Blue curve depicts the payoff to player 1 and the red curve depicts the payoff to player 2 for $p > \tilde{p}_1$, in the equilibrium in non-cutoff strategies. As argued, the blue curve meets the line D_1 at a belief p_s^1, which is strictly less than $p_1^{'}$. In the region $(\tilde{p}_1, p_s^1]$, player 2 experiments and player 1 free rides. At p_s^1, 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 p_s^2, player 2 shifts to the risky arm again. As argued, $p_s^2 > p_2^{*n}$.

It is important to have a welfare comparison between the equilibrium in cutoff strategies and an equilibrium in non-cutoff strategies. We have already established that the intensity of experimentation in the equilibrium in cutoff strategies is strictly higher than that in any equilibrium in non-cutoff strategies. This leads us to establish that welfare wise, the equilibrium in cutoff strategies is always better than any equilibrium in non-cutoff strategies. The following proposition describes this.

[5]Lemma 4 implies that the qualitative characteristics of p_s^1 and p_s^2 are the same in any equilibrium in non-cutoff strategies. Hence, it is without loss of generality to have this restriction.

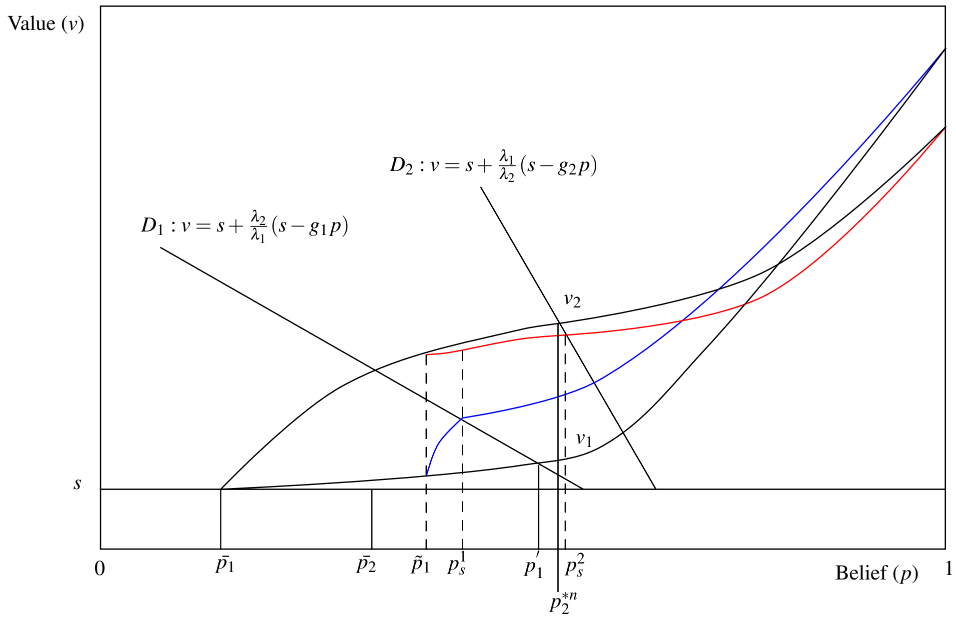

Figure $4(a)$.

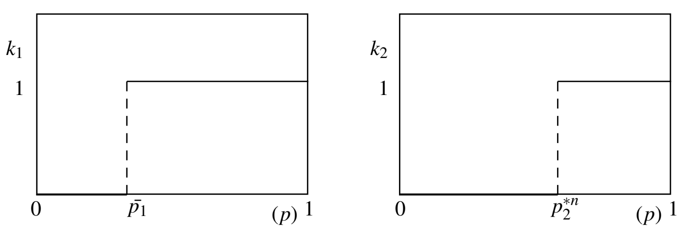

Figure $4(b)$: Actions of players in the equilibrium in cutoff strategies.

Proposition 4 *Consider the equilibrium in cutoff strategies and any equilibrium in non-cutoff strategies. Let v^c_{agg} be the aggregate equilibrium payoff in the equilibrium in cutoff strategies and v^{nc}_{agg} be the aggregate equilibrium payoff in an arbitrary equilibrium in non-cutoff strategies. For all $p \in (\bar{p}_1, 1)$, $v^c_{agg} \geq v^{nc}_{agg}$ with strict inequality at some beliefs.*

The proof of this proposition is relegated to appendix (H). This proposition implies that if the degree of heterogeneity between the players is higher than a particular threshold, the equilibrium in cutoff strategies is always the welfare maximising equilibrium. The intuition behind this is as follows. In the planner's solution, which is the action profile that yields maximum welfare, if a player has to free ride, then it is always player 2 who is made to do so. In the equilibrium in cutoff strategies, we in a way, achieve the second best. This is in the sense that although the aggregate range of beliefs where

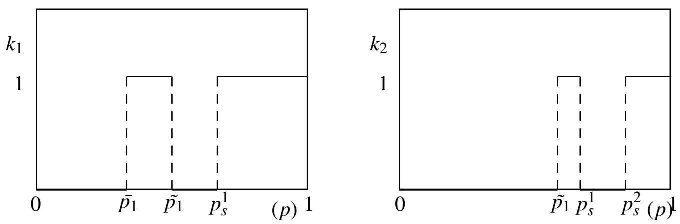

Figure $4(c)$: Actions of players in the equilibrium in non-cutoff strategies.

some experimentation takes place and also the range of beliefs where both players experiment shrink, in this equilibrium, it is only player 2 who free rides. However, as we have seen, in any equilibrium in non-cutoff strategies, both players free ride. Further, compared to the equilibrium in cutoff strategies, in any equilibrium in non-cutoff strategies, the range of beliefs where both players experiment is lower and the total range of beliefs over which some experimentation takes place does not change. This explains, why the equilibrium in cutoff strategies is the welfare maximising equilibrium.

In the following part, we turn our focus on characterising equilibria when there does not exist any equilibrium in cutoff strategies.

3.2.2 Characterisation of equilibria when the equilibrium in cutoff strategies does not exist

Consider the situation when all equilibria are in non-cutoff strategies. This happens when the value of λ_2 is in the range (λ_2^*, λ_1). From our analysis this far we know that in such a situation, $p_2^{*n} < p_1^{'}$. This allows us to infer that there does not exist any equilibrium in cutoff strategies and all equilibria are in non-cutoff strategies. The proposition below characterises a generic equilibrium in this situation.

Proposition 5 *When $\lambda_2 \in (\lambda_2^*, \lambda_1)$, a generic equilibrium is characterised by thresholds $\{\tilde{p}_j\}_{j=1}^{j=k}$, $k \geq 1$ (k finite) such that $\bar{p}_1 < \tilde{p}_1 \leq p_2^{*n}$ and $\tilde{p}_1 < \tilde{p}_2 < ... < \tilde{p}_k \leq p_s$ where $p_s = \min\{p_s^1, p_s^2\}$. Let v_1 and v_2 be the equilibrium payoffs of 1 and 2 respectively. p_s^i is the belief where the payoff function of player i meets the best response line D_i. If $p_s = p_s^1$, then for $p \in (p_s, p_s^2]$, player 1 experiments along the risky arm and player 2 chooses the safe arm. For all $p > p_s^2$, both players experiment along the risky arm. If $p_s = p_s^2$, then for $p \in (p_s, p_s^1]$, player 2 experiments along the risky arm and player 1 chooses the safe arm. For all $p > p_s^1$, both players experiment along the risky arm. For $p \in (\tilde{p}_j, \tilde{p}_{j+1}]$ $(1 \leq j \leq k-1)$ if player i $(i = 1, 2)$ chooses the safe arm then $v_i = F_i^j = s + \frac{\lambda_l}{\lambda_l + r}(g_i - s) + C_i^j(1 -$*

$p)[\Lambda(p)]^{\frac{r}{\lambda_l}}$, where $(l=1,2; l \neq i)$ C_i^j is determined from $F_i^j(\tilde{p}_j) = v_i(\tilde{p}_j)$. If player i chooses the risky arm then $v_i = v_i^j = g_i p + C_i^j(1-p)[\Lambda(p)]^{\frac{r}{\lambda_i}}$. C_i^j is determined from $v_i^j(\tilde{p}_j) = v_i(\tilde{p}_j)$. If j is odd(even), then for $p \in (\tilde{p}_j, \tilde{p_{j+1}}]$, player 1 chooses the safe (risky) and player 2 chooses the risky (safe) arm.

The proof of the above proposition is relegated to appendix (I). From the class of equilibria characterised in the above proposition, we can identify a particular equilibrium, where player 2 does not use the risky arm unless it becomes his dominant action. This corresponds to the equilibrium where for all $p \in (\bar{p}_1, p_2^{*n}]$ $(p_2^{*n} < p_1^{'})$, player 1 experiments along the risky arm and player 2 chooses the safe arm. There is a range of beliefs (p_2^{*n}, p_s^1), where player 1 free rides and for all $p > p_s^1$, both players experiment along the risky arm. Thus, as per the characterisation in proposition (5), for this particular equilibrium, $k=1$ and $\tilde{p_1} = p_s = p_s^2 = p_2^{*n}$. We will denote this equilibrium as the most heterogeneous equilibrium. In this equilibrium, player 2 uses a cutoff strategy but player 1 does not. The figure below depicts this equilibrium.

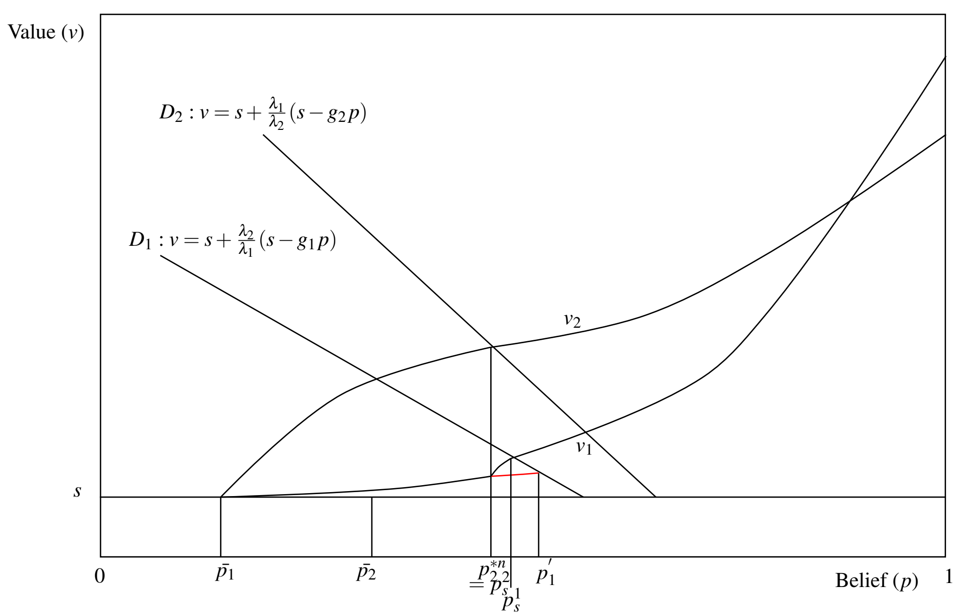

Figure 5.

The most heterogeneous equilibrium, that is the equilibrium in which player 2 does not experiment along the risky arm unless it becomes his dominant action, coincides with the equilibrium in cutoff strategies as $\lambda_2 \to \lambda_2^*$ from the above. This is described in the following lemma.

Lemma 5 *As $\lambda_2 \to \lambda_2^*$ from the above, the most heterogeneous equilibrium coincides with the equilibrium where both players use cutoff strategies.*

Proof.

In the most heterogeneous equilibrium, player 2 uses a cutoff strategy with p_2^{*n} as the threshold belief. Player 1 experiments along the risky arm for $p \in (\bar{p}_1, p_2^{*n}]$ and for $p > p_s^1$. For $p \in (p_2^{*n}, p_s^1]$, player 1 free rides. As λ_2 becomes smaller, p_2^{*n} approaches towards $p_1^{'}$ and $p_2^{*n} \to p_1^{'}$ as $\lambda_2 \to \lambda_2^*$. Hence, the range of beliefs over which player 1 free rides shrinks and eventually as $\lambda_2 \to \lambda_2^*$, this region becomes empty. This implies that in the equilibrium where player 2 does not choose the risky arm until it becomes his dominant action, player 1's strategy becomes a threshold type strategy as $\lambda_2 \to \lambda_2^*$. This concludes the proof of the lemma.

■

From the above lemma it is clear that the most heterogeneous equilibrium eventually coincides with the equilibrium in cutoff strategies as players become sufficiently heterogeneous. It is important that we analyse how the most heterogeneous equilibrium compare with other equilibria in terms of aggregate payoff.

We will argue that when λ_2 exceeds λ_2^*, there exists a range of values of λ_2 such that the most heterogeneous equilibrium strictly dominates all other equilibria in terms of the aggregate payoff. The following proposition describes this.

Proposition 6 *Let v_{agg}^{nch} be the aggregate payoff in the most heterogeneous equilibrium and v_{agg}^{nc} be that for any other equilibrium in non-cutoff strategies. There exists a $\bar{\lambda}_2 \in (\lambda_2^*, \lambda_1)$, such that for all $\lambda_2 \in (\lambda_2^*, \bar{\lambda}_2)$, for all $p > \bar{p}_1$, we have $v_{agg}^{nch} \geq v_{agg}^{nc}(p)$ with strict inequality at some beliefs.*

The formal proof of the proposition is relegated to appendix (J). We discuss the intuition behind this result. In any equilibrium, except the most heterogeneous one, players switch arms at least once at a belief below p_2^{*n}. This implies that the belief where the payoff of player 1 (2) intersects D_1 (D_2) is lower (higher) than the corresponding belief in the most heterogeneous equilibrium. When $\lambda_2 = \lambda_2^*$, the most heterogeneous equilibrium coincides with the equilibrium in cutoff strategies, and hence, the range of beliefs over which both players experiment along the risky arm shrinks for any other equilibrium. Hence, given that $\lambda_2 > \lambda_2^*$, there must be some range of values of λ_2 for which it is true that the range of beliefs over which both players experiment is highest in the most heterogeneous equilibrium.

Next, from the proof of proposition (4) we can infer that for beliefs greater than $\bar{p}_1$ and less than equal to p_2^{*n}, the aggregate payoff in the most heterogeneous equilibrium is highest. This is because for this range of beliefs, in the most heterogeneous equilibrium player 2 free rides. However, in any other equilibrium, for some beliefs in this range, player 1 free rides. Since all experimentation ceases at $\bar{p}_1$, the claim follows. This is true for any value of λ_2.

In the most heterogeneous equilibrium, there is a range of beliefs for which player 1 free rides. This range is $(p_2^{*n}, p_s^1]$ where p_s^1 is the belief where player 1's payoff in the most heterogeneous equilibrium intersects the line D_1. As argued, in any other equilibrium player 1's payoff will intersect D_1 at a belief $p_s^{1'} < p_s^1$. This implies that for $\lambda_2 > \lambda_2^*$, there exist some beliefs in the range (p_2^{*n}, p_s^1) where in any other equilibrium player 2 free rides, while in the most heterogeneous equilibrium player 1 free rides. However, for $\lambda_2 = \lambda_2^*$, $p_2^{*n} = p_s^1 = p_1^{'}$. Since, at p_2^{*n}, the aggregate payoff is highest in the most heterogeneous equilibrium, given that $\lambda_2 > \lambda_2^*$, there exist some values of λ_2 such that for all $p \in (p_2^{*n}, p_s^1)$, the aggregate payoff in the most heterogeneous equilibrium is highest. These allow us to conclude that for values of λ_2 sufficiently close to λ_2^* but greater than λ_2^*, welfare wise the most heterogeneous equilibrium strictly dominates any other equilibrium.

For other degree of heterogeneity the ranking between equilibria becomes ambiguous. This happens because with the introduction of any degree of heterogeneity, as argued above, for beliefs less than or equal to p_2^{*n}, the most heterogeneous equilibrium gives the highest aggregate payoff. However, it might happen that in some other equilibrium the range of beliefs over which both players experiment is higher, which in effect can make the aggregate payoff for those beliefs higher than that in the most heterogeneous equilibrium. Further, as argued above, compared to the most heterogeneous equilibrium, in any other equilibrium, there is always a range of beliefs where player 2 free rides in the latter equilibrium while player 1 free-rides in the former. For low degree of heterogeneity, this can make the aggregate payoff in the latter to be higher than the former for some beliefs in this range. Hence, once any degree of heterogeneity is introduced, unlike in the model with homogeneous players ([4]), it is not true that we unambiguously gain by making players take turns in experimenting more often.

4 Some Variants of the basic model

In this section, we will consider two variants of the basic model analysed in the paper. First, we consider the case when the heterogeneity between the players is due to different certain payoffs they obtain along the safe arm.

4.1 Heterogeneity in the safe arm payoffs

In this subsection, we analyse a model which is a variant of the basic model considered in the paper. In the current model we consider, both players have identical innate abilities in exploring the risky arm. That is $\lambda_1 = \lambda_2 = \lambda > 0$. However, they differ with respect to the payoff 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 $g = \lambda h$.

In all other respects, the model is similar to the basic model considered in the paper.

We first demonstrate the Planner's solution. The qualitative nature of the solution is same as in the basic model of the paper.

Planner's Solution: The planner's objective is to maximise the sum of the expected discounted 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 $v(p)$ should satisfy

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

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

The following lemma demonstrates the optimal solution of the planner.

Lemma 6 *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 $\mu = \frac{r}{\lambda}$ and $k^*(p) = (k^*_1(p), k^*_2(p))$ is the optimal solution of the planner. For all beliefs greater than p^*_{22}, $k^*(p) = (1,1)$, i.e both players are made to choose the risky arm. For beliefs greater than p^*_{12} and less than or equal to p^*_{22}, $k^*(p) = (1,0)$ implying 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}, $k^*(p) = (0,0)$, which means both players are made to choose the safe arm.

The detailed proof of this lemma is in appendix (K).

Next, we turn our attention to the analysis of the non-cooperative game.

Non-cooperative game:

We restrict ourselves to markovian strategies with the common posterior as the state variable. Let the strategy of player i be denoted by 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

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

and

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

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

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

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

This means given that the other player is choosing the risky arm, choosing the risky arm constitutes the best response for player 1 as long as in the (p,v) 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 the best response for player 2 as long as in the (p,v) plane, player 2's value lies above the line

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

On his own (equivalent to when player 2 is always choosing the safe arm), 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, on his own (equivalent to when player 1 is always choosing the safe arm), 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.

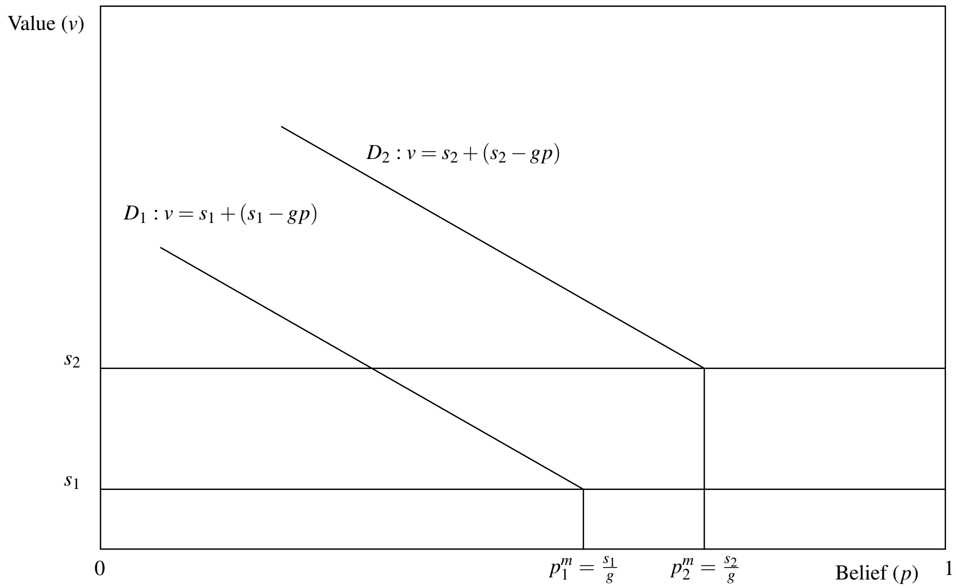

Figure 6.

From figure 6, it can be observed that heterogeneity in safe arm payoffs makes the best response lines of the players to diverge from each other. When $s_1 = s_2$, the lines coincide with each other and when $s_2 = g$, the lines are the farthest apart from each other. As argued in the basic model, in equilibrium, no experimentation will take place for beliefs below $\bar{p_{1s}}$. Also, for beliefs just above $\bar{p_{1s}}$ only player 1 can experiment. The proposition below argues that when players are sufficiently heterogeneous, there exists an equilibrium where both players use simple cutoff strategies and this equilibrium is unique in the class of equilibria where players use cutoff strategies.

Proposition 7 *There exists a $s_2^* \in (s_1, g)$ such that for all $s_2 \in (s_2^*, g)$, there exists an equilibrium where both players use simple cutoff strategies. Player* 1 *chooses the risky arm for all beliefs greater than $\bar{p_{1s}}$ and chooses the safe arm otherwise. There exists a unique $p_2^{*ns} \in (\bar{p_{1s}}, 1)$ such that player* 2 *chooses the risky arm for beliefs greater than p_2^{*ns} and safe arm otherwise. This equilibrium is unique in the class of equilibria where players use cutoff strategies.*

The formal proof of this proposition is given in appendix (L). It can be seen from figure 6 that as players' safe arm payoffs diverge from each other, the range of beliefs over which only player 2 can free ride gets expanded. As explained in the basic model, this allows for the existence of the equilibrium where both players use cutoff strategies. This equilibrium is depicted in figure 7. The equilibrium in cutoff strategies exists only if the curve v_1 intersects D_1 at a belief lower than the one at which the curve v_2 intersects D_2. This is possible only if the players are sufficiently heterogeneous.

The takeaway message from this proposition is that the qualitative effect of heterogeneity on equilibrium behaviour is robust to the kind of heterogeneity considered. This is because any kind of heterogeneity between the players makes the free riding opportunities of the players different. As players's safe arm payoffs become more different, free riding opportunities of player 1 shrink and that of player 2 expand. It is this similar effect on the free riding opportunities which gives rise to similar qualitative results.

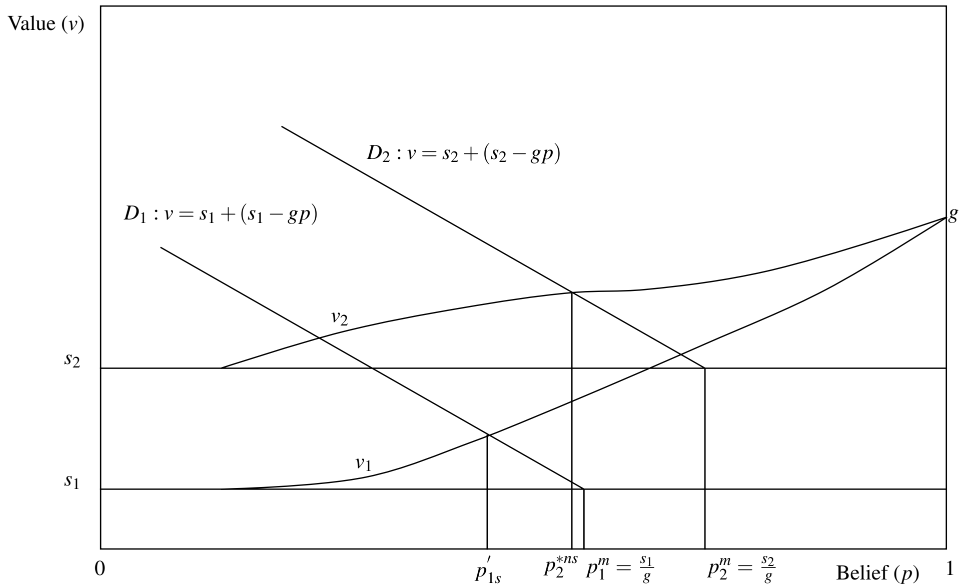

Figure 7.

In the following subsection, we consider the case when players are allowed to diversify efforts between the arms.

4.2 Players can diversify efforts between the arms

In this subsection, we analyse the basic model with an additional feature that at any belief, players are allowed to diversify their efforts between the arms. Hence, we will have $k_i \in [0,1]$, where k_i denotes the proportion of total effort player $i(i = 1,2)$ puts at the risky arm. The equilibria described in the basic model of the paper will still continue to exist for this modified model. However, there will emerge other equilibria, where over a range of beliefs, both players will diversify efforts between the arms. Before we demonstrate such equilibria, it is important that we discuss some properties of such an equilibrium.

First, it is easy to observe that in an equilibrium, if a player $i(i = 1,2)$ for some range of beliefs is diversifying efforts between the arms, then it must be the case that for this range of beliefs, the payoff for this player is below his best response line D_i. This is because once the payoff has crossed the best response line D_i, experimenting along the risky arm is the dominant action for player i. Further, if a player i is diversifying efforts between the arms over a range of beliefs, then it must be the case that for that range of beliefs, the other player is also diversifying efforts. This is because, when a player i's payoff is below his best response line D_i, if the other player is choosing a pure action, then player i's strict best response is to choose a pure action. Hence, in an equilibrium where players diversify efforts, both players should do it for the same range of beliefs. This implies that in an equilibrium where players diversify efforts over a range of beliefs, 1 (2)'s payoff should be below $D_1 (D_2)$. We will denote such an equilibrium as an *interior* equilibrium.

Next, it can be inferred that unlike in ([4]), no symmetric interior equilibrium exists in a model with heterogeneous players. This is because all experimentation ceases at the belief $\bar{p}_1$ and in any equilibrium, at beliefs just above $\bar{p}_1$ only player 1 can experiment. The following lemma describes a generic property of an interior equilibrium.

Lemma 7 *A common feature of any interior equilibrium is that there exist range(s) of beliefs where players diversify efforts between the arms. Over this range of beliefs, player* 1*'s payoff is given by*

$$W_1 = g_1 + \mu_1(g_1 - s) + \mu_1 s(1-p)\log\Lambda(p) + C_1^i(1-p)$$

Player 2*'s payoff is given by*

$$W_2 = g_2 + \mu_2(g_2 - s) + \mu_2 s(1-p)\log\Lambda(p) + C_2^i(1-p)$$

The integration constants C_1^i and C_2^i are determined from the value matching conditions at the belief where diversification of efforts begin.

Proof.

In an equilibrium, if a player is diversifying efforts between the arms, then the player must be indifferent between choosing the risky arm and the safe arm. From (2) we can infer that W_1, the payoff of player 1 over the range of beliefs where efforts are diversified will satisfy

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

This gives us the expression for W_1. Similarly, from (3) we obtain the expression for W_2.

From (2) we can infer that

$$W_1(p) = s + k_2\frac{\lambda_2}{\lambda_1}(s - g_1 p) \Rightarrow k_2 = \frac{(W_1(p) - s)}{(s - g_1 p)}\frac{\lambda_1}{\lambda_2}$$

Similarly, from (3) we get

$$k_1 = \frac{(W_2(p) - s)}{(s - g_2 p)} \frac{\lambda_2}{\lambda_1}$$

This concludes the proof of the lemma.

■ In any interior equilibrium, diversification of efforts begins at a belief which is strictly higher than $\bar{p_1}$. We can identify a set of interior equilibria where diversification starts from $\bar{p_1} + \varepsilon$, ($\varepsilon > 0$ and $\varepsilon \to 0$) and continues until a player's payoff intersects his best response line. As $\lambda_2 \to \lambda_1$, all these equilibria converge to the symmetric equilibrium in ([4]). Let us denote the set of these equilibria by Φ

The following lemma describes the properties of the beliefs where payoffs intersect their corresponding best response lines in an interior equilibrium.

Lemma 8 *Let p_s^{1i} be the belief at which the payoff of player 1 intersects the line D_1 and p_s^{2i} be the belief at which the payoff of player 2 intersects the line D_2. We have $p_s^{1i} < p_1^{'}$ and $p_s^{2i} > p_2^{*n}$.*

Proof.

From lemma (4), without loss of generality we can restrict ourselves to those equilibria where before players start diversifying efforts, only player 2 free rides. The detailed proof of this lemma is given in appendix (M).

■ The above lemma implies that when the equilibrium in cutoff strategies exists, the range of beliefs over which in an interior equilibrium both players experiment, shrinks. This is however, not sufficient to guarantee that whenever the equilibrium in cutoff strategies exists, it strictly dominates any interior equilibrium. It might happen that over the range of beliefs where players diversify efforts, the aggregate payoff in an interior equilibrium is higher than that in the equilibrium in cutoff strategies. The following proposition describes that for sufficiently high degree of heterogeneity, all interior equilibria are strictly welfare wise dominated by the equilibrium in cutoff strategies.

Proposition 8 *Whenever the equilibrium in cutoff strategies exists, welfare wise it strictly dominates any equilibrium in Φ. Further, there exists a λ_2^{**} ($\frac{s}{h} < \lambda_2^{**} \leq \lambda_2^{*}$) such that for all $\lambda_2 < \lambda_2^{**}$, the aggregate equilibrium payoff is highest in the equilibrium in cutoff strategies.*

The formal proof of the proposition is relegated to appendix (N). This shows that even if players are allowed to diversify efforts between the arms, for sufficient degree of heterogeneity between the players, welfare wise the equilibrium in cutoff strategies strictly dominates all other equilibria.

5 Conclusion

This paper has exhaustively characterised the equilibria in a two armed bandit model when players are heterogeneous. Except for very low degree of heterogeneity, we can always identify the welfare maximising equilibrium. As heterogeneity between the agents increases, equilibrium in cutoff strategies exists. When the equilibrium in cutoff strategies exists, it is the welfare maximising equilibrium. For sufficient degree of heterogeneity, this result holds even if we allow players to diversify efforts between the arms. Further, the effect on equilibrium behaviour due to heterogeneity is robust to the kind of heterogeneity adopted.

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] Horner, J., Skrzypacz, A., 2016 “Learning, Experimentation and Information Design ”, *Working paper, Stanford University*.

[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 $v(p)$ in this range satisfies

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

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

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

This gives us

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

This gives us

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

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

$$v(p) = v_{sr} = s + [\frac{\lambda_1 g + r g_1}{\lambda_1 + r} - \frac{s\lambda_1}{r + \lambda_1}]p + C_{rs}^P(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

This gives us

$$[g - v - v^{'}(1-p)] = \frac{r(g-s) - r g_1}{\lambda_1 + r} + \frac{r}{\lambda_1}\frac{1}{p}C_{rs}^P(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

Hence,

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

Thus,

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

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

On the other hand we have

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

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

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

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

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

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

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

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

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

B If player 2 is the last player to experiment at $p = \bar{p}_1$, then for beliefs right above $\bar{p}_1$, 2's payoff becomes strictly lower than s

Suppose player 2 is the last player to experiment at the belief $\bar{p}_1$. This implies that at $p = \bar{p}_1$, 2's payoff is equal to s and at beliefs right above $\bar{p}_1$, 2's payoff is given by

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

Value matching condition at $p = \bar{p}_1$ implies that $C_2 = \frac{s - g_2\bar{p}_1}{(1-\bar{p}_1)[\Lambda(\bar{p}_1)]^{\frac{r}{\lambda_2}}}$. Hence, at $p = \bar{p}_1$, the right derivative of player 2's payoff is given by

$$g_2 - \frac{(s - g_2\bar{p}_1)}{(1-\bar{p}_1)}\{1 + \frac{r}{\lambda_2}\frac{1}{\bar{p}_1}\}$$

The sign of the above expression is same as

$$g_2 - g_2\bar{p}_1 - (s - g_2\bar{p}_1)\{1 + \frac{r}{\lambda_2}\frac{1}{\bar{p}_1}\}$$

$$= g_2 - s + \frac{rg_2}{\lambda_2} - \frac{sr}{\lambda_2\bar{p}_1}$$

This expression is negative as long as $\bar{p}_1 < \frac{\mu_2 s}{(\mu_2+1)g_2 - s} = \bar{p}_2$. Since this is always true, the payoff of 2 at beliefs right above $\bar{p}_1$ is strictly less than s.

C Proof of proposition (2)

In any equilibrium, all experimentation ceases at the belief $\bar{p}_1$ and player 1 is the last player to experiment at beliefs right above $\bar{p}_1$. Hence, if there is an equilibrium in cutoff strategies, then player 1 should experiment for all beliefs greater than $\bar{p}_1$. This implies that player 1 can never free ride. For beliefs when player 2 free rides, the payoff to 1 should be given by v_1^{rsc} and 2's payoff should be F_2^c. Since player 1 should never have any incentive to free ride, it must be the case that player 1's equilibrium payoff intersects D_1 at a belief which is lower than the belief where 2's payoff intersects D_2. Hence, the requirement to have an equilibrium in cutoff strategies is to have $p_1^{'} < p_2^{*n}$.

Consider the situation when $\lambda_2 \to \lambda_1$ from the below. In that case, D_2 almost coincides with D_1. F_2^c lies above v_1^{rsc} for all $p < \frac{s}{g_1}$. Hence, $p_2^{*n} < p_1^{'}$.

When $\lambda_2 \to \frac{s}{h}$ from the above, the line D_2 becomes $v = s + \frac{\lambda_1}{\lambda_2}(s - sp)$. This implies that the line D_2 intersects the horizontal line $v = s$ at $p = 1$. Also, the function F_2^c approximately becomes $F_2^c \approx s$. Hence, the belief at which F_2^c intersects the line D_2 is $p = 1$. On the other hand, D_1 becomes $s + \frac{s}{\lambda_1}[s - g_1 p]$. Thus, D_1 still intersects the horizontal line $v = s$ at $p = \frac{s}{g_1}$. This implies that $p_1^{'} \in (0,1)$. Hence, $p_1^{'} < p_2^{*n}$.

As λ_2 decreases, the line D_1 pivots left around the point $(\frac{s}{g_1}, s)$. v_1^{rsc} is independent of λ_2. Hence, $p_1^{'}$ decreases. On the other hand as λ_2 decreases, D_2 shifts right and F_2^c becomes flatter. Hence, p_2^{*n} increases. Since D_1, D_2, v_1^{rsc} and F_2^c are all continuous functions of λ_2, there exists a λ_2^* such that for all $\lambda_2 \in (\frac{s}{h}, \lambda_2^*)$, $p_1^{'} < p_2^{*n}$. This concludes the proof of the proposition.

D Condition for player 2 choosing the risky arm when player 1 is choosing the safe arm

If player 1 is choosing the safe arm, from the value matching condition we know that the payoff to player 2 at p from experimenting along the risky arm is given by

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

Player 2 finds it beneficial to choose the risky arm as long as

$$\lambda_2 b_2^n(p, v_2) > (s - g_2 p)$$

where $\lambda_2 b_2^n(p, v_2) = \lambda_2 p \frac{g_2 - v_2 - (1-p)v_2^{'}}{r}$. From the expression of v_2 at this point we know that the right derivative of v_2 is given by $v_2^{'} = g_2 - C_2[\Lambda(p)]^{\frac{r}{\lambda_2}}\{1 + \frac{r}{\lambda_2}\frac{1}{p}\}$. This gives us

$$\lambda_2 b_2^n(p, v_2) = C_2(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

Thus, $\lambda_2 b_2^n(p, v_2) > s - g_2 p \Rightarrow$

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

$$\Rightarrow v_2 > s$$

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

At $p = p_2^{*n}$, 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 just below p_2^{*n}, the sum of payoffs will be higher than the cost of experimentation. This implies that the net social benefit becomes higher from 2's experimentation. Since, the social planner maximises the net social benefit, we have $p_2^* < p_2^{*n}$.

F Proof of lemma 4

In any equilibrium, for beliefs right above $\bar{p}_1$, only player 1 experiments. Also, we are considering that range of heterogeneity such that the equilibrium in cutoff strategies exists. This implies, it must be the case that in an equilibrium in non-cutoff strategies, before their payoffs intersect the respective best response lines, players switch arms at least once. Let $p_s > \bar{p}_1$ be the belief where players switch arms for the first time. This means that for $p \in (\bar{p}_1, p_s]$, player 1's payoff is given by v_1^{rsc} and player 2's payoff is given by F_2^c. For beliefs just above p_s, player 1's payoff is given by $F_1^1 = s + \frac{\lambda_2}{\lambda_2 + r}(g_1 - s)p + C_1^1(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$ and player 2's payoff is given by $v_2^{rs1} = g_2 p + C_2^1(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$. The integration constants for both players are determined from the value matching conditions at $p = p_s$.

First, we 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. This shows that the function is strictly convex.

From the value matching condition at $p = p_s$, $v_2^{rs1}(p_s) = F_2^c(p_s)$. Since $F_2^c(p_s) > s$, we must have the integration constant of v_2^{rs1} to be strictly greater than that of v_2^{rs}. Hence, v_2^{rs1} is strictly convex.

From the expressions of $F_2^c(p)$ and $v_2^{rs1}(p)$, we have

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

and

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

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

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

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

$$v_2^{rs1'}(p_s) = g_2 - C_2^1[\Lambda(p_s)]^{\frac{r}{\lambda_2}}[1+\frac{r}{\lambda_2 p_s}] < g_2 - C_2^1[\Lambda(p_s)]^{\frac{r}{\lambda_2}}[1+\frac{r}{\lambda_1 p_s}] = \hat{v_2'}$$

Hence,

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

$$= \frac{\frac{r}{\lambda_1}[\frac{s}{p_s} - g_2]}{(1-p_s)} > 0$$

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

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

Next, we show that $F_1^1(p)$ lies strictly above $v_1^{rsc}(p)$ as long as $F_1^1(p) < D_1(p)$.

At $p = p_s$, player 2 shifts to the risky arm. Hence, for beliefs right above 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 has 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

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

where $v_1^{rr}(p)$ is the payoff to player 1 if he chooses the risky arm. The integration constant of v_1^{rr} is determined from $v_1^{rr}(p_s) = v_1^{rsc}(p_s)$.

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

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

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

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

Thus, we have $C_1^{rr} > 0$.

This gives us

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

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

Hence, we can infer that at beliefs right above p_s, $v_1^{rr}(p) > v_1^{rsc}(p)$. Further, they cannot cross again since if they have to cross then $v_1^{rsc'}(.)$ needs to be strictly greater than $v_1^{rr'}(.)$. However, as argued, this is not possible. Thus, for all $p_s < p < 1$, $v_1^{rr}(p) > v_1^{rsc}(p)$. This implies that as long as $F_1^1(p) < D_1(p)$,

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

If players do not switch arms any more, then as long as player 1's payoff is below the line D_1, it will be strictly above v_1^{rsc} and as long as player 2's payoff is below the line D_2, it will strictly below F_2^c. We will now argue that this will be the case even if players switch arms more than once. For some belief above p_s, if player 1 reverts to the risky arm and player 2 free rides, then 1's payoff will be given by v_1^{rs} such that the integration constant is determined from the value matching condition at the next switching belief. Since at the switching belief, $v_1^{rs} = F_1^1 > v_1^{rsc}$, v_1^{rs} will be convex, and will always be above v_1^{rsc}. From our previous analysis in the current lemma, we can infer that if again player 1 free rides, the payoff will be strictly above v_1^{rs} and hence, above v_1^{rsc}. Thus, if players switch arms at least once, player 1's payoff before intersecting D_1 will be strictly above v_1^{rsc}.

Similarly, for some beliefs above p_s, if player 2 starts free riding again, then 2's payoff will be given by F_2 such that the integration constant will be determined from the value matching condition at the next switching belief. Since at the switching belief, $F_2 = v_2^{rs1} < F_2^c$, F_2 will always be strictly concave and hence, will always be strictly below F_2^c. From our previous analysis in the current lemma, we can infer that if again player 2 starts experimenting along the risky arm, his payoff will be strictly below F_2 and hence, below F_2^c. Thus, if players switch arms at least once, payoff of player 2 before intersecting the line D_2 will be strictly below F_2^c. Hence, we can infer that $p_s^1 < p_1^{'}$ and $p_s^2 > p_2^{*n}$. This concludes the proof of the lemma.

G Proof of proposition 3

We prove this proposition by constructing a generic equilibrium in non-cutoff strategies for a particular $k \geq 1$ (k finite). From our previous conclusions, we know that for beliefs right above $\bar{p}_1$, in any equilibrium, player 1's payoff is given by $v_1^{rsc}(p)$ and that of player 2 is given by $F_2^c(p)$. The belief where players switch arms for the first time, both players' payoffs should be below their respective best response lines. Consider an arbitrary belief $\tilde{p}_1$ such that $\bar{p}_1 < \tilde{p}_1 < p_1^{'}$. Suppose players switch arms for the first time at $p = \tilde{p}_1$. Thus, for beliefs right above $\tilde{p}_1$, player 1 free rides by choosing the safe arm and player 2 experiments along the risky arm. Hence, for beliefs right above $\tilde{p}_1$, 1's payoff is given by $F_1^1(p) = s + \frac{\lambda_2}{\lambda_2+r}(g_1 - s)p + C_1^1(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$. The integration constant C_1^1 is determined from $F_1^1(\tilde{p}_1) = v_1^{rsc}(\tilde{p}_1)$. Player 2's payoff for beliefs right above $\tilde{p}_1$ is given by $v_2^1(p) = g_2 p + C_2^1(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$. The integration constant C_2^1 is determined from $v_2^1(\tilde{p}_1) = F_2^c(\tilde{p}_1)$. At $\tilde{p}_1$, $v_1^{rsc}(\tilde{p}_1) < D_1(\tilde{p}_1)$ and $F_2^c(\tilde{p}_1) < D_2(\tilde{p}_1)$. Since payoffs and best response lines are continuous functions, there exists a belief $\tilde{p}_2 > \tilde{p}_1$ such that $v_2^1(\tilde{p}_2) < D_2(\tilde{p}_2)$ and $F_1^1(\tilde{p}_2) < D_1(\tilde{p}_2)$. From lemma 4, we know that $\tilde{p}_2 < p_1^{'}$. This is the belief where players switch arms for the second time. For beliefs right above $\tilde{p}_2$, player 1 experiments along the risky arm and player 2 free rides by choosing the safe arm. Hence, 1's payoff for these beliefs is given by $v_1^2(p) = g_1 p + C_1^2(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$. As before, the integration constant is determined from the value matching condition at $p = \tilde{p}_2$. For these beliefs, player 2's payoff is given by $F_2^2(p) = s + \frac{\lambda_1}{\lambda_1+r}(g_2 - s)p + C_2^2(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$. The integration constant is determined from the value matching condition at $p = \tilde{p}_2$.

In the same manner, we can choose $\tilde{p}_3 < \tilde{p}_4 < ... < \tilde{p}_k$ such that $\tilde{p}_3 > \tilde{p}_2$ and $\tilde{p}_k < p_1^{'}$. At each of these beliefs, players switch arms. Since in any equilibrium player 1 is the last one to experiment at $p = \bar{p}_1$, at beliefs right above $\tilde{p}_j$ ($k \geq j \geq 1$), player 1 free rides by choosing the safe arm and player 2 experiments along the risky arm if j is odd. On the other hand if j is even, the opposite happens. Hence, if k is odd, then for beliefs right above $\tilde{p}_k$, player 1 free rides and player 2 experiments. Thus, at p_s^1 (which is determined endogenously) players switch arms again since for beliefs right above p_s^1, experimenting along the risky arm is the dominant action for player 1. At p_s^2 (again determined endogenously), player 2 reverts to the risky arm.

If k is even, then for all beliefs above $\tilde{p}_k$, player 1 experiments along the risky arm. As before, player 2 reverts to the risky arm at $p = p_s^2$. This concludes the proof of the proposition.

H Proof of proposition (4)

In an equilibrium, if for a range of beliefs, player 1 chooses the risky arm and player 2 chooses the safe arm, then the aggregate payoff is given by

$$v_{12} = s + g_1 p + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p + C_{12}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}} \equiv s + Ap + C_{12}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

where $A = g_1 + \frac{\lambda_1}{\lambda_1+r}(g_2 - s)$ and C_{12} is an integration constant.

If for a range of beliefs, player 2 experiments along the risky arm and player 1 chooses the safe arm, then the aggregate payoff is given by

$$v_{21} = s + g_2 p + \frac{\lambda_2}{\lambda_2 + r}(g_1 - s)p + C_{21}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}} \equiv s + Bp + C_{21}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

where $B = g_2 + \frac{\lambda_2}{\lambda_2+r}(g_1 - s)$ and C_{21} is an integration constant.

For the equilibrium in cutoff strategies, the aggregate payoff $v^c_{agg}(p)$ is given by

$$v^c_{agg}(p) = 2s \text{ for } p \leq \bar{p_1}$$

$$v^c_{agg}(p) = v^c_{12}(p) = s + Ap + C^c_{12}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}} \text{ for } p \in (\bar{p_1}, p_2^{*n}]$$

with $v^c_{12}(\bar{p_1}) = 2s$. Since $\bar{p_1} < \frac{s}{g_1 + \frac{\lambda_1}{\lambda_1+r}(g_2-s)}$, we have $C^c_{12} > 0$. For $p > p_2^{*n}$, we have

$$v^c_{agg}(p) = v^c_{rr}(p) = g_1 p + g_2 p + C^c_{rr}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

with $v^c_{rr}(p_2^{*n}) = v^c_{12}(p_2^{*n})$

Consider an arbitrary equilibrium in non-cutoff strategies. Let the aggregate payoffs be given by $v^{nc}_{agg}(p)$. From proposition (3), we can infer that for $p \leq \tilde{p_1}$, $v^{nc}_{agg}(p) = v^c_{agg}(p)$. At $p = \tilde{p_1}$, players switch arms. This means for $p \in (\tilde{p_1}, \tilde{p_2}]$, we have $v^{nc}_{agg}(p) = v^1_{21}(p) = s + Bp + C^1_{21}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$. Since value matching condition is satisfied for both players at $p = \tilde{p_1}$, we have

$$C^1_{21}(1-\tilde{p_1})[\Lambda(\tilde{p_1})]^{\frac{r}{\lambda_2}} - C^c_{12}(1-\tilde{p_1})[\Lambda(\tilde{p_1})]^{\frac{r}{\lambda_1}} = (A-B)\frac{\tilde{p_1}}{(1-\tilde{p_1})} > 0$$

As $C^c_{12} > 0$, we have $C^1_{21} > 0$

We have

$$v_{agg}^{c'}(\tilde{p_1}) - v_{agg}^{nc'}(\tilde{p_1}) = (A-B) + C_{21}^1(1-\tilde{p_1})[\Lambda(\tilde{p_1})]^{\frac{r}{\lambda_2}}[1+\frac{r}{\lambda_2}] - C_{12}^c(1-\tilde{p_1})[\Lambda(\tilde{p_1})]^{\frac{r}{\lambda_1}}[1+\frac{r}{\lambda_1}]$$

Since, $(A-B)>0$, $[1+\frac{r}{\lambda_2}] > [1+\frac{r}{\lambda_1}]$ and $C_{21}^1 > 0$, we have

$$v_{agg}^{c'}(\tilde{p_1}) - v_{agg}^{nc'}(\tilde{p_1}) > (A-B) + [1+\frac{r}{\lambda_1}]C_{21}^1(1-\tilde{p_1})[\Lambda(\tilde{p_1})]^{\frac{r}{\lambda_2}} - C_{12}^c(1-\tilde{p_1})[\Lambda(\tilde{p_1})]^{\frac{r}{\lambda_1}} > 0$$

This is true whenever v_{12}^c and v_{21}^1 intersect. Hence, we have $v_{21}^1(p) < v_{12}^c(p)$ for all $p \in (\tilde{p_1}, \tilde{p_2}]$. Hence, for all $p \in (\tilde{p_1}, \tilde{p_2}]$, we have $v_{agg}^c(p) > v_{agg}^{nc}(p)$.

At $p = \tilde{p_2}$, players switch arms. This implies that for $p \in (\tilde{p_2}, \tilde{p_3}]$, the aggregate payoff in the equilibrium in non-cutoff strategies is given by

$$v_{agg}^{nc}(p) = v_{12}^2(p) = s + Ap + C_{12}^2(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

such that $v_{12}^2(\tilde{p_2}) = v_{21}^1(\tilde{p_2})$. Since $v_{12}^c(\tilde{p_2}) > v_{21}^1(\tilde{p_2})$, we have $v_{12}^2(\tilde{p_2}) < v_{12}^c(\tilde{p_2})$. Thus, $\forall p \in (\tilde{p_2}, \tilde{p_3}]$, we have $v_{12}^2(p) < v_{12}^c(p)$. This implies $\forall p \in (\tilde{p_2}, \tilde{p_3}]$, $v_{agg}^c(p) > v_{agg}^{nc}(p)$.

Next, since $C_{21}^1 > 0$, in the same way as before we can argue that $v_{12}^{2'}(\tilde{p_2}) > v_{21}^{1'}(\tilde{p_2})$. Hence, $\forall p \in (\tilde{p_2}, \tilde{p_3}]$, $v_{21}^1(p) < v_{12}^2(p)$. Thus, at $p = \tilde{p_3}$, $v_{21}^1(p) < v_{12}^2(p)$.

At $p = \tilde{p_3}$, players switch arms. This implies that for $p \in (\tilde{p_3}, \tilde{p_4}]$, in the equilibrium in non-cutoff strategies, we have

$$v_{agg}^{nc}(p) = v_{21}^3(p) = s + Bp + C_{21}^3(1-p)[\Lambda(p)]^{\frac{r}{\lambda_2}}$$

such that $v_{21}^3(\tilde{p_3}) = v_{12}^2(\tilde{p_3})$. Since, $v_{12}^2(\tilde{p_3}) > v_{21}^1(\tilde{p_3})$, we have $v_{21}^3(\tilde{p_3}) > v_{21}^1(\tilde{p_3})$. This leads us to infer that $C_{21}^3 > C_{21}^1 > 0$. Hence, as before we can show that $\forall p \in (\tilde{p_3}, \tilde{p_4}]$, $v_{21}^3(p) < v_{12}^2(p) < v_{12}^c(p)$. Thus, $\forall p \in (\tilde{p_3}, \tilde{p_4}]$, we have $v_{agg}^c(p) > v_{agg}^{nc}(p)$.

Proceeding in this manner recursively, we can show that $\forall p \in (\tilde{p_1}, p_s^1]$, $v_{agg}^c > v_{agg}^{nc}$. For $p \in (p_s^1, p_s^2)$, in the equilibrium in non-cutoff strategies, player 1 experiments along the risky arm and player 2 chooses the safe arm. Hence, $\forall p \in (p_s^1, p_s^2]$, we have

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

such that $v_{12}^s(p_s^1) = v_{agg}^{nc}(p_s^1)$. Since $v_{agg}^{nc}(p_s^1) < v_{agg}^c(p_s^1) = v_{12}^c(p_s^1)$, we have $v_{12}^s(p) < v_{12}^c(p)\ \forall p > p_s^1$. Thus, $\forall p \in (p_s^1, p_2^{*n}]$, $v_{agg}^{nc}(p) < v_{agg}^c(p)$.

At $p = p_2^{*n}$, in the equilibrium in cutoff strategies, player 2 switches to the risky arm, with player

1 remaining at the risky arm. Hence, for all $p > p_2^{*n}$, we have

$$v_{agg}^c(p) = v_{rr}^c(p) = g_1 p + g_2 p + C_{rr}^c(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

such that $v_{rr}^c(p_2^{*n}) = v_{12}^c(p_2^{*n})$. For player 2, at $p = p_2^{*n}$, smooth pasting condition is satisfied. Further, from our earlier reasoning we know that at $p = p_2^{*n}$, for player 1, $v_1^{rrc'}(p) > v_1^{rsc'}(p)$. Hence, we can conclude that for $p > p_2^{*n}$, $v_{rr}^c(p) > v_{12}^c(p)$. Thus, for $p \in (p_2^{*n}, p_s^2]$, $v_{agg}^c(p) > v_{agg}^{nc}(p)$. Finally, at $p = p_s^2$, in the equilibrium in non-cutoff strategies, player 2 reverts to the risky arm. Hence, for $p > p_s^2$, we have

$$v_{agg}^{nc}(p) = v_{rr}^{nc}(p) = g_1 p + g_2 p + C_{rr}^{nc}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

with $v_{rr}^{nc}(p_s^2) = v_{12}^s(p_s^2)$. Since, $v_{12}^s(p_s^2) < v_{rr}^c(p_s^2)$, we have $v_{rr}^{nc}(p_s^2) < v_{rr}^c(p_s^2)$. Thus, for all $p \in (p_s^2, 1)$, $v_{agg}^{nc}(p) < v_{agg}^c(p)$.

Hence, we have shown that for all $p \in (\tilde{p}_1, 1)$, the aggregate payoff in the equilibrium in cutoff strategies is strictly greater than that in any equilibrium in non-cutoff strategies. This concludes the proof.

I Proof of proposition (5)

We will prove this proposition by constructing a generic equilibrium. Since we consider that range of heterogeneity such that the equilibrium in cutoff strategies does not exist, we have $\bar{p}_1 < p_2^{*n} < p_1^{'}$. In any equilibrium, for beliefs right above $\bar{p}_1$, only player 1 can experiment along the risky arm. Hence, payoff to player 2 for beliefs right above $\bar{p}_1$ is given by F_2^c and that to player 1 is given by v_1^{rsc}. We choose a $\tilde{p}_1 \in (\bar{p}_1, p_2^{*n}]$. At this belief, both players' payoffs are below their respective best response lines. Thus, in equilibrium, players can swap arms at $p = \tilde{p}_1$. As explained in lemma (4) and proposition (3), the continuation payoff of 1 for beliefs just above $\tilde{p}_1$ is given by $F_1^1(p)$ and it lies above v_1^{rsc}. Similarly, the continuation payoff of player 2 for beliefs just above $\tilde{p}_1$ is given by $v_2^1(p)$ and it lies below $F_2^c(p)$. Let p_s^{11} be the belief where $F_1^1(p)$ intersects D_1 and p_s^{21} be the belief where player $v_2^1(p)$ intersects D_2. Clearly, $p_s^{11} < p_1^{'}$ and $p_s^{21} > p_2^{*n}$. Let $p_{s1} = \min\{p_s^{11}, p_s^{21}\}$. We can choose a $\tilde{p}_2 \in (\tilde{p}_1, p_{s1})$. Since at $p = \tilde{p}_2$, $v_2^1(p)$ is below D_2 and $F_1^1(p)$ is below D_1, in equilibrium players can again swap arms. In similar way, we can choose the remaining $k-2$ beliefs where players swap arms and similar to proposition (3), we can define the payoffs of players for beliefs within these thresholds. Depending on the choice of the k threshold beliefs, p_s^1, p_s^2 and hence, p_s are endogenously determined. This concludes the proof of the proposition.

J Proof of proposition (6)

In the most heterogeneous equilibrium, for all $p \in (\bar{p}_1, p_2^{*n}]$, player 1 experiments along the risky arm and player 2 chooses the safe arm. Hence, for $p \in (\bar{p}_1, p_2^{*n}]$, we have

$$v_{agg}^{nch}(p) = v_{12}^{c}(p) = s + Ap + C_{12}^{c}(1-p)[\Lambda(p)]^{\frac{r}{\lambda_1}}$$

such that $v_{12}^{c}(\bar{p}_1) = 2s$. It has already been argued in the proof of proposition (4) that $C_{12}^{c} > 0$. For all $p \in (p_2^{*n}, p_s^1]$, we have

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

The integration constant C_{21}^{nch} is determined from $v_{12}^{c}(p_2^{*n}) = v_{21}^{nch}(p_2^{*n})$. As shown earlier in the proof of proposition (4), $v_{21}^{nch}(p)$ lies below $v_{12}^{c}(p)$ $\forall p \in (p_2^{*n}, p_s^1]$.

For all $p > p_s^1$, we have

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

The integration constant is determined from $v_{rr}^{nch}(p_s^1) = v_{21}^{nch}(p_s^1)$.

Consider any other equilibrium in non-cutoff strategies. In this equilibrium, players switch arms at least once before choosing the risky arm becomes a dominant action for any player. From our previous analysis, we can infer that compared to the most heterogeneous equilibrium, the belief where the payoff of player 1 intersects the best response line D_1 is lower and the belief where the payoff of player 2 intersects the best response line D_2 is higher. Hence, in such an equilibrium if $p_s^{1'}$ is the belief where payoff of 1 and $p_s^{2'}$ is the belief where player2's payoff intersect D_1 and D_2 respectively, then we have $p_s^{1'} < p_s^1$ and $p_s^{2'} > p_2^{*n}$. Let $\tilde{p}_1 (\bar{p}_1 < \tilde{p}_1 < p_2^{*n})$ be the belief where in the other equilibrium in non-cutoff strategies players switch arms for the first time. From the proof of proposition (4) we know that for all $p \in (\tilde{p}_1, p_2^{*n}]$, $v_{agg}^{nch}(p) > v_{agg}^{nc}(p)$.

Next, we show that when the equilibrium in cutoff strategies does not exist, for sufficiently low values of λ_2, it is always the case that the range of beliefs over which both players experiment along the risky arm is highest in the most heterogeneous equilibrium. To see this, observe that for $\lambda_2 = \lambda_2^*$, we have $p_2^{*n} = p_1^{'}$. Hence, for values of λ_2 just above λ_2^*, in the most heterogeneous equilibrium, both players experiment along the risky arm for $p > p_s^1$ and for the other equilibrium in non-cutoff strategies, both players experiment along the risky arm for $p > p_s^{2'} > p_2^{*n} \approx p_1^{'} > p_s^1$. Hence, we can find a $\lambda_2 = \bar{\lambda}_2^1 \in (\lambda_2^*, \lambda_1)$ such that for all $\lambda_2 \in (\lambda_2^*, \bar{\lambda}_2^1]$, we will continue to have $p_s^{2'} > p_1^{'} > p_s^1$.

In that case, for all these values of λ_2, for all equilibria except the most heterogeneous equilibrium, $(p_s^{2'},1)$ is the range of beliefs over which both players experiment along the risky arm. In the most heterogeneous equilibrium, $(p_s^1,1)$ is the range of beliefs over which both players experiment along the risky arm. Hence, for $\lambda_2 \in (\lambda_2^*, \bar{\lambda}_2^1]$, the most heterogeneous equilibrium has the highest range of beliefs over which both players experiment along the risky arm.

We will now argue that when there does not exist any equilibrium in non-cutoff strategies, for certain low values of λ_2, $v_{agg}^{nch}(p) > v_{agg}^{nc}(p)$ for all $p \in (p_2^{*n}, p_s^1]$. To show this, the non-trivial case is the one when in the other equilibrium in non-cutoff strategies, there exists a range of beliefs in $(p_2^{*n}, p_s^1]$ where player 1 experiments along the risky arm and player 2 chooses the safe arm. Let p_l be the lowest belief in $(p_2^{*n}, p_s^1]$ where this happens. As explained earlier, we have $v_{agg}^{nch}(p_l) > v_{agg}^{nc}(p_l)$. For $\lambda_2 = \lambda_2^*$, the range $(p_2^{*n}, p_s^1]$ does not exist. Since the payoffs are continuous functions, there exists a $\bar{\lambda}_2^2 \in (\lambda_2^*, \lambda_1)$, such that for all $\lambda_2 \in (\lambda_2^*, \bar{\lambda}_2^2)$, we have $\forall p \in (p_2^{*n}, p_s^1]$, $v_{agg}^{nch}(p) > v_{agg}^{nc}(p)$.

Define $\bar{\lambda}_2 = \min\{\bar{\lambda}_2^1, \bar{\lambda}_2^2\}$. Then, for all $\lambda_2 \in (\lambda_2^*, \bar{\lambda}_2]$, $p_s^{2'} > p_1^{'} > p_s^1$ and for all $p \in (\bar{p}_1, p_s^1]$, $v_{agg}^{nch}(p) > v_{agg}^{nc}(p)$. Since in the most heterogeneous equilibrium, both players experiment along the risky arm for $p > p_s^1$, as argued in the proof of proposition (4), for all $p \in (p_s^1, 1)$, $v_{agg}^{nch}(p) > v_{agg}^{nc}(p)$. This concludes the proof of the proposition.

K Proof of lemma (6)

We prove this in two steps. First, from the proposed policy or solution we compute the planner's payoff. Then, by a verification argument we show that this computed payoff solves the Bellman equation of the planner.

First, consider the range $p \in (0, p_1^{**}]$. According to the conjectured solution, $k_1^*(p) = k_2^*(p) = 0$. This implies $v(p) = s_1 + s_2$. For $p \in (p_1^{**}, p_2^{**}]$, according to the conjectured solution, $k_1^*(p) = 1$ and $k_2^*(p) = 0$. This implies that $v(p)$ satisfies

$$v^{'}(p) + \frac{(r+\lambda)}{\lambda p(1-p)} = \frac{rs_2}{\lambda p(1-p)} + \frac{rg + 2\lambda g}{\lambda(1-p)}$$

The solution to the above equation is

$$v(p) = s_2 + \{\frac{2\lambda p + rg}{r+\lambda} - \frac{s\lambda}{r+\lambda}\}p + C_s(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

Since at $p = p_1^{**}$, the value matching and the smooth pasting conditions are satisfied, we obtain $p_1^{**} = \frac{\mu s_1}{(2+\mu)g - (s_1+s_2)}$.

According to the conjectured solution, for $p > p_2^{**}$, both players use the risky arm. Hence,

$$v(p) = 2gp + C_r(1-p)[\Lambda(p)]^{\frac{r}{2\lambda}}$$

At $p = p_2^{**}$, player 2 is shifted to the safe arm. This means at $p = p_2^{**}$, both value matching and smooth pasting conditions are satisfied. Also, at $p = p_2^{**}$, we have

$$b(p,v) = \frac{\lambda p}{r}\{2g - v - v^{'}(1-p)\} = s_2 - gp$$

$b(p,v)$ can be expressed as $\frac{1}{2}[v-2gp]$. This gives us $v(p_2^{**}) = 2s > s_1 + s_2$. Since the value matching condition is satisfied at $p = p_2^{**}$, we must have

$$s_2 + \{\frac{2\lambda g + rg}{r+\lambda} - \frac{s\lambda}{r+\lambda}\}p_2^{**} + C_s(1-p_2^{**})[\Lambda(p_2^{**})]^{\frac{r}{\lambda}} = 2s_2$$

Thus, p_2^{**} should satisfy

$$\{\frac{2\lambda g + rg}{r+\lambda} - \frac{s\lambda}{r+\lambda}\}p_2^{**} + C_s(1-p_2^{**})[\Lambda(p_2^{**})]^{\frac{r}{\lambda}} = s_2$$

We will now show that there exists a $p_2^{**} \in (p_1^{**}, 1)$ such that the above is satisfied.

At $p_2^{**} = p_1^{**}$, we have L.H.S = $s_1 < s_2$. At $p_2^{**} = 1$, L.H.S = $g + \frac{\lambda}{r+\lambda}(g-s) > g > s_2$. Since L.H.S is continuous and monotonically increasing, there exists a unique $p_2^{**} \in (p_1^{**}, 1)$ such that the required condition is satisfied.

We will now verify that the obtained value function satisfies the Bellman equation.

First, consider the range of beliefs $p \in (p_2^{**}, 1)$. According to the conjectured solution, we have

$$v(p) = 2gp + C_r(1-p)[\Lambda(p)]^{\frac{r}{2\lambda}}$$

Thus, we have

$$b(p,v) = \frac{\lambda p}{r}\{2g - v - v^{'}(1-p)\} = \frac{1}{2}[v - 2gp]$$

The value function is strictly increasing in p. At $p = p_2^{**}$, $v(p) = 2s_2 > 2s_1$. This gives us $b(p,v) - (s_1 - gp) > 0$ and $b(p,v) - (s_2 - gp) > 0$. Hence, for $p > p_2^{**}$, it is optimal for the planner to make both players use the risky arm.

Consider $p \in (p_1^{**}, p_2^{**}]$. Given the value function obtained for this range of beliefs from the conjectured solution, we have

$$b(p,v) = v - s_2 - gp$$

For $p > p_1^{**}$, $v > s_1 + s_2$. This implies $b(p,v) - (s_1 - gp) = v - (s_1 + s_2) > 0$. Hence, for the considered range of beliefs, it is optimal to make player 1 to choose the risky arm. However, for the considered range of beliefs, we have $v < 2s_2$. This implies $b(p,v) - (s_2 - gp) = v - 2s_2 < 0$. This shows that it is optimal to make player 2 to choose the safe arm. Further, at $p = p_1^{**}$, $b(p,v) = s_1 - gp$. This implies that at this belief, the planner is indifferent between making player 1 to choose the safe arm or the risky arm. Similarly, at $p = p_2^{**}$, $b(p,v) = s_2 - gp$. Hence, at this belief the planner is indifferent between making player 2 to choose the safe or the risky arm. This completes the verification argument.

L Proof of proposition (7)

In any equilibrium, for beliefs just above $\bar{p_{1s}}$, only player 1 can experiment along the risky arm. This means for beliefs just above $\bar{p_{1s}}$, payoff of player 1 is given by

$$v_{s1}^{rsc}(p) = gp + C_{s1}^{rsc}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

The integration constant C_{s1}^{rsc} is determined from the value matching condition $v_{s1}^{rsc}(\bar{p_{1s}}) = s_1$.

Player 2's payoff is given by

$$F_{s2}^{c}(p) = s_2 + \frac{\lambda}{\lambda + r}(g - s_2)p + C_{s2}^{rsc}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

The integration constant C_{s2}^{rsc} is determined from the value matching condition $F_{s2}^{c}(\bar{p_{1s}}) = s_2$.

First, we show that there exists a unique $p_{1s}^{'} \in (\bar{p_{1s}}, 1)$ such that $v_{s1}^{rsc}(p_{1s}^{'}) = D_1(p_{1s}^{'})$. v_{s1}^{rsc} is strictly convex and increasing in p. D_1 on the other hand is monotonically decreasing in p. Since $\bar{p_{1s}} < \frac{s_1}{g}$, we have $v_{s1}^{rsc}(\bar{p_{1s}}) < D_1(\bar{p_{1s}})$ and $v_{s1}^{rsc}(1) > D_1(1)$. Hence, there exists a unique $p_{1s}^{'} \in (\bar{p_{1s}}, 1)$ such that $v_{s1}^{rsc}(p_{1s}^{'}) = D_1(p_{1s}^{'})$. It is to be observed that $p_{1s}^{'}$ is independent of s_2.

Next, we will argue that there exists a unique $p_2^{*ns} \in (\bar{p_{1s}}, 1)$ such that $F_{2s}^{c}(p_2^{*ns}) = D_2(p_2^{*ns})$. F_{2s}^{c} is strictly concave and increasing in p. D_2 is monotonically decreasing in p. $F_{2s}^{c}(\bar{p_{1s}}) < D_2(\bar{p_{1s}})$ and $F_{2s}^{c}(1) > D_2(1)$. Thus, there exists a unique $p_2^{*ns} \in (\bar{p_{1s}}, 1)$ such that $F_{2s}^{c}(p_2^{*ns}) = D_2(p_2^{*ns})$.

The equilibrium in cutoff strategies exists only if $p_2^{*ns} > p_{1s}^{'}$.

When $s_2 \to g$ from below, D_2 intersects the horizontal line $v = s_2$ at $p = 1$. Also, $F_{2s}^{c}(p) \approx s_2$. Hence, $p_2^{*ns} \approx 1 > p_{1s}^{'}$. When $s_2 = s_1$, we have the homogeneous players model and from ([4]) we know that $p_2^{*ns} < p_{1s}^{'}$. As s_2 increases, the line D_2 shifts right and $F_{2s}^{c}(p)$ becomes flatter. Hence, p_2^{*ns}

increases. Thus, there exists a unique $s_2^* \in (s_1, g)$ such that for all s_2 greater than s_2^*, $p_2^{*ns} > p_{1s}^{'}$. From our arguments in the basic model, we can infer that there exists an equilibrium where both players use cutoff strategies and this equilibrium is unique in the class of equilibria where players use cutoff strategies. This concludes the proof of the proposition.

M Proof of lemma (8)

Let p_I be the belief where players start diversifying efforts. This implies that at $p = p_I$, we have

$$W_1(p_I) = v_1^{rsc}(p_I) = g_1 p + C_1^{rsc}(1-p)[\Lambda(p)]^{\frac{r}{\lambda}}$$

We will show that $W_1^{'}(p_I) > v_1^{rsc'}(p_I)$.

From the expression of $W_1(p)$, we have

$$W_1^{'}(p_I) = -\mu_1 s[\log\Lambda(p_I) + \frac{1}{p_I}] - C_1^i$$

Thus, we have

$$\begin{gathered}(1-p_I)\{W_1^{'}(p_I) - V_1^{rsc'}(p_I)\} \\ = \{-g_1 - \mu_1(g_1 - s) - \mu_1 s(1-p_I)\log\Lambda(p_I) - C_1^i(1-p_I)\} + \{g_1 p_I + C_1^{rsc}(1-p_I)[\Lambda(p_I)]^{\frac{r}{\lambda_1}}\} \\ +\mu_1(g_1 - s) + \frac{r}{\lambda_1}C_1^{rsc}\frac{(1-p_I)}{p_I}[\Lambda(p_I)]^{\frac{r}{\lambda_1}} - \mu_1 s\frac{(1-p_I)}{p_I} \\ = \mu_1\{\frac{v_1^{rsc}(p_I) - s}{p_I}\}\end{gathered}$$

Since $p_I > \bar{p}_1$, we have $v_1^{rsc}(p_I) > s$. This means that $(1-p_I)\{W_1^{'}(p_I) - V_1^{rsc'}(p_I)\} > 0$.

Since this is true whenever these functions intersect, for the range of beliefs when players diversify efforts, $W_1(p)$ always lies above $v_1(p)$. This proves that $p_s^{1i} < p_1^{'}$.

Similarly, we have

$$\begin{gathered}(1-p_I)\{W_2^{'}(p_I) - F_2^{c'}(p_I)\} \\ = \{-g_2 - \mu_2(g_2 - s) - \mu_2 s(1-p_I)\log\Lambda(p_I) - C_2^i(1-p_I)\} + \{s + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s)p_I + C_2^{rsc}(1-p_I)[\Lambda(p_I)]^{\frac{r}{\lambda_1}} \\ +C_2^{rsc}(1-p_I)[\Lambda(p_I)]^{\frac{r}{\lambda_1}}\frac{r}{\lambda_1}\frac{1}{p_I} + \mu_2(g_2 - s) + g_2 - s - \frac{\lambda_1}{\lambda_1 + r}(g_2 - s) - \mu_2\frac{(1-p_I)}{p_I}s \\ = \frac{r}{\lambda_1 p_I}\{F_2^c(p_I) - D_2(p_I)\}\end{gathered}$$

Since for the range of beliefs when players diversify efforts, payoffs of players are below their corresponding best response lines, we have $F_2^c(p_I) < D_2(p_I)$. Thus, we have $(1-p_I)\{W_2^{'}(p_I) - F_2^{c'}(p_I)\} < 0$. Since this is true whenever these functions intersect, for the range of beliefs when players diversify efforts, $W_2(p)$ always lies below $F_2^c(p)$. This proves that $p_s^{2i} > p_2^{*n}$. This concludes the proof of the lemma.

N Proof of proposition (8)

Suppose $\lambda_2 < \lambda_2^*$ such that the equilibrium in cutoff strategies exists.

Consider a range of beliefs $(p_L, p_H]$ over which in an interior equilibrium players diversify efforts. Over this range of beliefs, in the equilibrium in cutoff strategies, either player 1 experiments and 2 free rides or both players experiment. From our previous analysis, we know that for a range of beliefs, in the equilibrium in cutoff strategies when only one player experiments, the aggregate equilibrium payoff is given by $v_{12}^c(p)$. From the expression of $v_{12}^c(p)$ we know that

$$v_{12}^{c'} = g_1 + \frac{\lambda_1}{\lambda_1 + r}(g_2 - s) - C_{12}^c[\Lambda(p)]^{\frac{r}{\lambda_1}}\{1 + \frac{r}{\lambda_1}\}$$

The aggregate payoff for this range of beliefs in the interior equilibrium is given by

$$W_{12}(p) = W_1(p) + W_2(p) = (g_1 + g_2) + \mu_1(g_1 - s) + \mu_2(g_2 - s) + (\mu_1 + \mu_2)s(1-p)\log\Lambda(p) + C_{12}^I(1-p)$$

From the value matching conditions at $p = p_L$, we have $W_{12}(p_L) = v_{12}^c(p_L)$

$$\Rightarrow W_{12}^{'} = -(\mu_1 + \mu_2)s[\log\Lambda(p) + \frac{1}{p}] - C_{12}^I$$

Thus, we have

$$\begin{aligned}
&(1-p_L)\{v_{12}^{c'}(p_L) - W_{12}^{`}(p_L)\} \\
&= -\frac{r}{\lambda_1}\frac{1}{p_L}\{v_{12}^c(p_L) - [s + \frac{\lambda_1}{\lambda_2}(s - g_2 p) + s]\} \\
&= -\frac{r}{\lambda_1}\frac{1}{p_L}\{v_{12}^c(p_L) - [D_2(p_L) + s]\}
\end{aligned}$$

Consider any equilibrium in the set Φ. Since p_L is very close to $\bar{p_1}$, payoff of player 1 at $p = p_L$ is approximately equal to s. Since at this belief the payoff of player 2 is strictly less than $D_2(p_L)$, $\{v_{12}^c(p_L) - [D_2(p_L) + s]\} < 0$. Hence, $(1-p_L)\{v_{12}^{c'}(p_L) - W_{12}^{`}(p_L)\} > 0$. Hence, the aggregate payoff in the equilibrium in cutoff strategies is strictly higher than that in the interior equilibrium.

Now consider any interior equilibrium outside the set Φ. If $\lambda_2 \to \frac{s}{h}$, then Payoff of player 2 $\to s$ and $[D_2(p_L)+s] \to s - g_1 p + s + g_1$. For $p \in (p_L, p_H]$, $s - g_1 p > 0$. Since the payoff of player 1 is always strictly less than g_1, we have $\{v_{12}^c(p_L) - [D_2(p_L)+s]\} < 0$. Thus, there exists a $\lambda_2^{*'} \in (\frac{s}{h}, \lambda_1)$ such that for all $\lambda_2 < \lambda_2^{*'}$, $\{v_{12}^c(p_L) - [D_2(p_L)+s]\} < 0$. Define $\lambda_2^{**} = \min\{\lambda_2^*, \lambda_2^{*'}\}$. Hence, for all $\lambda_2 < \lambda_2^{**}$ the aggregate payoff in the equilibrium in cutoff strategies for $p \in (p_L, p_H]$ will be strictly higher than that in any interior equilibrium. This concludes the proof of the proposition.