Strategic Experimentation with Asymmetric Players*

Kaustav Das† Nicolas Klein‡ Katharina Schmid

April 14, 2018

Abstract

We examine a two-player game with two-armed exponential bandits, where players operate different technologies for exploring the risky option. We characterise the set of Markov perfect equilibria (MPE), and show that there always exists an equilibrium in which the weaker player uses a cutoff strategy. If the degree of asymmetry between the players is high enough, there exists an MPE in cutoff strategies, which is welfare-maximising whenever it exists.

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

1 Introduction

In many instances, the information produced by one agent is interesting to other agents as well. Think e.g. of firms exploring neighbouring oil patches: If one firm strikes oil, chances are there will be oil in its neighbour's patch as well. Such games of purely informational externalities have been analysed by the strategic bandit literature,[1] which so far has only analysed the case of homogeneous agents. However, in many instances, one of the oil firms, for example, might be a big multinational firm that has access to a superior drilling technology. In this article, we aim

*We thank Sven Rady for his advice and guidance. The second author gratefully acknowledges support from the *Social Sciences and Humanities Research Council of Canada*. Part of the results presented in this paper were already contained in the third author's undergraduate thesis, entitled "Strategisches Experimentieren mit asymmetrischen Spielern," which she submitted at the University of Munich in 2009 under her maiden name Tönjes.

†Department of Economics, University of Exeter Business School, Email:K.Das@exeter.ac.uk
‡Université de Montréal and CIREQ, Email:kleinnic@yahoo.com

[1]The first paper to do so was Bolton and Harris (1999). Keller, Rady, Cripps (2005) have introduced exponential bandits, which we shall use here.

to analyse the impact of asymmetries in players' exploration technologies in a game of strategic experimentation with two-armed exponential bandits.

Specifically, players' bandits have a safe arm that generates a known positive flow payoff, and a risky arm that can either be good or bad. If it is bad, it does not generate any payoffs. If it is good, it generates a better expected flow payoff than the safe arm. The payoffs of the good risky arm arrive as lump sums realised at random times, which are exponentially distributed. Initially, players do not know if their risky arm is good or bad; they share a common prior belief about it. When the risky arm is used without a lump sum arriving, players continuously grow more pessimistic about its quality. As the type of the risky arm is assumed to be the same for both players, and players' actions and their outcomes are perfectly publicly observable, players also learn about their own risky arm from their partner's experimentation. There thus arises an informational externality, although there are no direct payoff externalities between the players.

The seminal paper by Keller, Rady, Cripps (2005) analyses this problem with homogeneous players. In the current paper, we generalise the analysis by introducing asymmetric players, in the sense that their payoff arrival rates from a good risky arm differ. This implies that, given the risky arm is good, the expected time needed to learn this differs between the players. As actions and outcomes are perfectly publicly observable, and players start out with a common prior, they will always have a common posterior belief. We characterise the set of Markov perfect equilibria with the players' common posterior belief as the state variable for all ranges of asymmetry between the players. If the degree of asymmetry between the players is sufficiently high, there exists an *equilibrium in cutoff strategies*, i.e. where both players use a cutoff strategy. That is, either player uses the risky arm if and only if the likelihood he attributes to the option being good is greater than a certain threshold. This equilibrium is unique in the class of equilibria in cutoff strategies. Whenever only one of the players experiments and the other free rides in this equilibrium, it is always the player with the weaker technology who free rides. In the case of homogeneous players (Keller, Rady, Cripps (2005)), by contrast, there never exists an equilibrium in cutoff strategies, and players swap the roles of pioneer and free-rider at least once in any equilibrium. In our setting, aggregate payoffs in the equilibrium in cutoff strategies are higher than in any other equilibrium. If the degree of asymmetry is low, at least one player uses a non-cutoff strategy in any equilibrium. In contrast to the homogeneous case (Keller, Rady, Cripps (2005)), we furthermore show that more frequent switches of arms do not unambiguously improve the equilibrium welfare with asymmetric players.

This paper contributes to the literature on strategic experimentation with bandits, a problem studied quite widely in economics, amongst others, by Keller and Rady (2010), Klein and Rady (2011) and Thomas (2017). In all of these papers, players are homogeneous. Except in Thomas

(2017) and Klein and Rady (2011), players' bandits are of the same type and *free-riding* is a common feature in all the above models except for Thomas (2017). Many variants of this problem have been studied in the literature. Rosenberg, Salomon, Vieille (2013) and Murto and Välimäki (2011), for instance, assume that switches to the safe arm are irreversible and that experimentation outcomes are private information, while Bonatti and Hörner (2011) and Heidhues, Rady, Strack (2015) investigate the case of private actions. Rosenberg, Solan, Vieille (2007) analyse the role of the observability of outcomes and the correlation between risky-arm types in a setting in which a switch to the safe arm is irreversible. Besanko and Wu (2013) use the Keller, Rady, Cripps (2005) framework to study how an R & D race is impacted by market structure. The paper closest to the present paper is Keller, Rady, Cripps (2005), who find that, with homogeneous players, there is never an equilibrium in cutoff strategies. By contrast, we show that, with heterogeneous players, an equilibrium in cutoff strategies may exist, and that it is welfare-maximising whenever it exists.

The rest of the paper is organised as follows. Section 2 sets out the model. Section 3 discusses the social planner's solution. A detailed analysis of equilibria for different ranges of heterogeneity is undertaken in Section 4. Finally, Section 5 concludes. Payoff functions are shown in Appendix A, while some proofs are relegated to the Appendix B.

2 Two armed bandit model with heterogeneous players

There are two players (1 and 2), each of whom faces a two-armed bandit in continuous time. One of the arms is safe, in that a player who uses it gets a flow payoff of $s > 0$. The risky arm can be either good or bad. Both players' risky arms are of the same type. If the risky arm is good, then a player using it receives a lump sum, drawn from a time-invariant distribution with mean $h > s$, at the jumping times of a Poisson process. The Poisson process governing player 1's arrivals has intensity $\lambda_1 = 1$, while player 2's arrive according to a Poisson process with intensity $\lambda_2 \in (\frac{s}{h}, 1)$. Thus, a good risky arm gives player 1 (2) an expected payoff flow of $g_1 = \lambda_1 h = h$ ($g_2 = \lambda_2 h$), with $g_1 > g_2 > s$. The parameters and the game are common knowledge.

The uncertainty in this model arises from the fact that players do not initially know whether their risky arms are good or bad. Players start with a common prior belief $p_0 \in (0,1)$ that their risky arms are good. Players have to decide in continuous time whether to choose the safe arm or the risky arm. At each instant, players can choose only one arm. We write $k_{i,t} = 1$ ($k_{i,t} = 0$) if player $i \in \{1,2\}$ uses his risky (safe) arm at instant $t \geq 0$. Players' actions and outcomes are publicly observable and, based on these, they update their beliefs. Players discount the future according to the common discount rate $r > 0$.

Let p_t be the players' common belief that their risky arms are good at time $t \geq 0$. Given player i's ($i \in \{1,2\}$) actions $\{k_{i,t}\}_{t \geq 0}$, which are required to be progressively measurable with respect

to the available information and to satisfy $k_i(t) \in \{0,1\}$ for all $t \geq 0$, player i's expected payoff is given by

$$\mathbb{E}\left[\int_0^\infty re^{-rt}[(1-k_{i,t})s + k_{i,t}p_t g_i]\,dt\right],$$

where the expectation is taken with respect to the processes $\{k_{i,t}\}_{t\geq 0}$ and $\{p_t\}_{t\geq 0}$. As can be seen from the objective function, there are no payoff externalities between the players. Indeed, the presence of the other player impacts a given player's payoffs only via the information that he generates, i.e. via the belief.

As mentioned in the Introduction, we will focus our analysis on Markov perfect equilibria with the players' common posterior belief as the state variable. Formally, a Markov strategy of player i is any left-continuous function $k_i : [0,1] \to \{0,1\}, p \mapsto k_i(p)$ $(i = 1,2)$ that is also piecewise continuous, i.e. continuous at all but a finite number of points.

As only a good risky arm can yield positive payoffs in the form of lump sums, the arrival of a lump sum fully reveals the risky arm to be good. Hence, if either player receives a lump sum at a time $\tau \geq 0$, then $p_t = 1$ for all $t > \tau$. In the absence of a lump-sum arrival, the belief follows the following law of motion for a.a. t:

$$dp_t = -(k_{1,t} + \lambda_2 k_{2,t})p_t(1-p_t)\,dt.$$

3 Planner's Problem

Suppose there is a benevolent social planner, who controls the actions of both players and wants to maximise the sum of their payoffs. By standard arguments, it is without loss of generality for the planner to restrict himself to Markov strategies $(k_1(p_t), k_2(p_t))$ with the posterior belief p_t as the state variable. The Bellman equation for the planner's problem is given by

$$v(p) = 2s + \max_{k_1,k_2\in\{0,1\}} \big\{k_1[B_1(p,v) - c_1(p)] + k_2[B_2(p,v) - c_2(p)]\big\}, \tag{1}$$

where we write $v(p)$ for the planner's value function, and, like Keller, Rady, Cripps (2005), define the myopic opportunity cost of having player i play risky , $c_i(p) = s - pg_i$, and the corresponding learning benefit

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

Note that the planner's Bellman equation is linear in both k_1 and k_2, so that our restriction to

action plans $\{(k_{1,t},k_{2,t})\}_{t\geq 0}$ with $k_{i,t}\in\{0,1\}$ for all (i,t) is without loss in the planner's problem. To state the following proposition, which describes the planner's solution, we define $g=g_1+g_2$, $\lambda=1+\lambda_2$, $\mu=\frac{r}{\lambda}$, $u_1(p):=(1-p)\left(\frac{1-p}{p}\right)^r$, $u_0(p):=(1-p)\left(\frac{1-p}{p}\right)^\mu$.

Proposition 1 *The planner's optimal policy $k^*(p)=(k_1^*,k_2^*)(p)$ is given by*

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

and the value function is

$$v(p)=\begin{cases} gp+\left[\frac{\lambda}{\lambda_2}s-gp_2^*\right]\frac{u_0(p)}{u_0(p_2^*)} & \text{if } p\in(p_2^*,1],\\ s+\left[\frac{g+rg_1}{1+r}-\frac{s}{1+r}\right]p+\left[s-\left(\frac{g+rg_1}{1+r}-\frac{s}{1+r}\right)p_1^*\right]\frac{u_1(p)}{u_1(p_1^*)} & \text{if } p\in(p_1^*,p_2^*],\\ 2s & \text{if } p\in(0,p_1^*],\end{cases}$$

where p_1^* is defined as

$$p_1^*=\frac{rs}{(1+r)g_1+g_2-2s}, \tag{2}$$

and $p_2^*\in(p_1^*,\frac{s}{g_2})$ is implicitly defined by $v(p_2^*)=\frac{\lambda}{\lambda_2}s$.

Proof. Proof is by a standard verification argument. Please see the Appendix B.1 for details. ■

By the above proposition, the belief at which player 1 switches to the safe arm in the planner's solution is higher than it would be if both players' Poisson arrival rates were equal to $\lambda_1=1$. This is because, as player 2's arrival rate λ_2 decreases, the benefit from player 1's experimentation decreases.

The planner's solution is depicted in the Figure 1. The planner's value function is a smooth convex curve which lies in the range $[2s,g]$. At the belief $p_2^*(p_1^*)$, player 2 (1) switches to the safe arm.

4 Non-cooperative game

We will first analyse a player's best responses to a given Markov strategy of the other player.

Best Responses: Fix player j's strategy k_j ($j\in\{1,2\}\setminus\{i\}$). If the payoff function from player

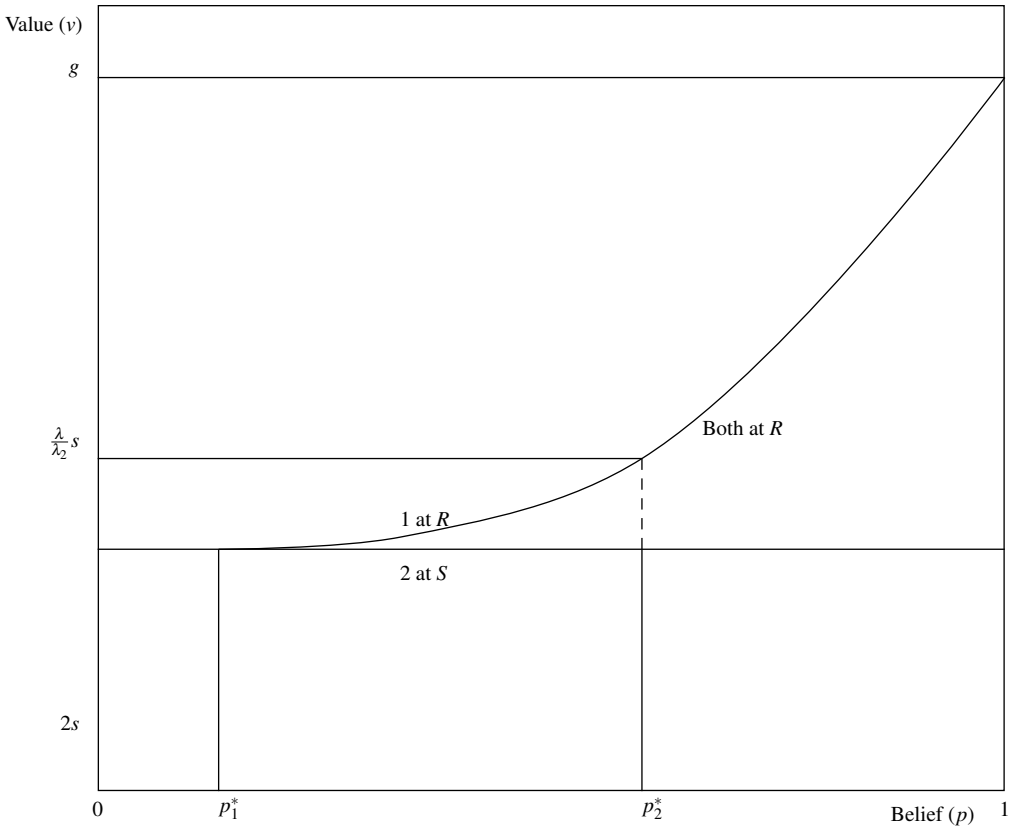

Figure 1.

i's response satisfies the following Bellman equation, player i is playing a best response:[2]

$$v_i(p) = s + k_j(p)\lambda_j b_i(p, v_i) + \max_{k_i \in \{0,1\}} k_i[\lambda_i b_i(p, v_i) - (s - g_i p)] \tag{3}$$

where

$$b_i(p, v_i) = p\frac{\{g_i - v_i - (1-p)v_i'\}}{r}.$$

As before, $\lambda_i b_i(p, v_i)$ can be interpreted as the learning benefit accruing to player i due to his own experimentation, while $\lambda_j b_i(p, v_i)$ is the learning benefit accruing to player i from player j's experimentation. The myopic opportunity cost of experimentation continues to be $c_i(p) = s - g_i p$.

[2]By standard results, on any open interval of beliefs in which player j's action choice is constant, player i's value function v_i will be continuously differentiable. At those (finitely many) beliefs at which player j's action changes, v' should be understood as the left derivative of v (since beliefs can only drift down).

For a given $k_j \in \{0,1\}$, from (3) we know that player i's payoff function satisfies the Bellman equation if and only if

$$k_i(p)\begin{cases} =1 & \text{if } \lambda_i b_i(p,v_i) > s - g_i p, \\ \in\{0,1\} & \text{if } \lambda_i b_i(p,v_i) = s - g_i p, \\ =0 & \text{if } \lambda_i b_i(p,v_i) < s - g_i p. \end{cases}$$

By rearranging we can infer that

$$k_i(p)\begin{cases} =1 & \text{if } v_i > s + k_j \frac{\lambda_j}{\lambda_i}[s - g_i p], \\ \in\{0,1\} & \text{if } v_i = s + k_j \frac{\lambda_j}{\lambda_i}[s - g_i p], \\ =0 & \text{if } v_i < s + k_j \frac{\lambda_j}{\lambda_i}[s - g_i p]. \end{cases}$$

This implies that when $k_j = 1$, player i chooses the risky arm, safe arm or is indifferent between them depending on whether his value in the (p,v) plane lies above, below, or on the line

$$D_i(p) = s + \frac{\lambda_j}{\lambda_i}[s - g_i p] \tag{4}$$

The single-agent threshold for player i is given by

$$\bar{p}_i = \frac{\mu_i s}{\mu_i s + (1+\mu_i)(g_i - s)} \tag{5}$$

where $\mu_i = \frac{r}{\lambda_i}$. In Appendix A.2, we display the ODEs the players' payoff functions satisfy, as well as their solutions, for each possible action profile. We start off by showing that, as in the homogeneous case (Keller, Rady, Cripps (2005)), no efficient equilibrium exists.

Proposition 2 *In any MPE, both players play safe at all beliefs in* $[0, \bar{p}_1]$. *There is thus no efficient MPE.*

Proof. Suppose to the contrary that p_l, the infimum of the set of beliefs at which at least one player plays risky, satisfies $p_l < \bar{p}_1$. Clearly, $v_i(p_l) = s$ for both $i \in \{1,2\}$. We shall now distinguish two cases depending on whether or not there exists an $\bar{\varepsilon} > 0$ such that, in any ε-right neighbourhood of p_l with $\varepsilon \in (0, \bar{\varepsilon})$, only one player i plays risky. If there does not exist such an $\bar{\varepsilon} > 0$, i is not playing a best response, because $p_l < \bar{p}_i < \frac{s}{g_i}$ implies that the point (p_l, s) is below the diagonal D_i. In the other case, player i faces the same trade-off as a single agent, and does not play a best response either, because $p_l < \bar{p}_i$. ■

In the next subsection, we will characterise the condition under which an equilibrium in cutoff strategies exists.

4.1 Equilibrium in cutoff strategies

As we have argued in the proof of Proposition 2, there is no experimentation below the belief $\bar{p}_1$ in any equilibrium. We will now argue that, in any equilibrium, only player 1 will experiment in some right-neighbourhood of $\bar{p}_1$, implying that player 1 is the last player to experiment in any equilibrium.

By Proposition 2, we know that $v_1(\bar{p}_1) = v_2(\bar{p}_1) = s$, and thus, by continuity, both players' value functions must be below their respective diagonals D_i in some neighbourhood of $\bar{p}_1$. Thus, in any equilibrium, at most one player can play risky in some right-neighbourhood of $\bar{p}_1$. Now, suppose that player 2 is the only player to experiment in some right-neighbourhood of $\bar{p}_1$. Then, the relevant ODE (Equation 12 in Appendix A.2) gives us that $\lambda_2 \bar{p}_1(1-\bar{p}_1)v_2'(\bar{p}_1+) = \bar{p}_1\lambda_2(g_2 - s) - rc_2(\bar{p}_1) < 0$, as $\bar{p}_1 < \bar{p}_2$. Thus, player 2's value function drops below s immediately to the right of $\bar{p}_1$, which contradicts his playing a best response. We can thus conclude that there exists some belief $\hat{p}_1 > \bar{p}_1$ such that, on $(\bar{p}_1, \hat{p}_1)$, player 2 plays safe. As either player can always guarantee himself his single-agent payoff by ignoring the information he gets for free from the other player, his payoff in any equilibrium is bounded below by his single-agent payoff. Thus, in any equilibrium, $v_1 > s$ on $(\bar{p}_1, \hat{p}_1]$, and player 1 experiments, while player 2 free-rides, in this range.

Thus, for beliefs right above $\bar{p}_1$, in any equilibrium, player 1's payoff is given by

$$\bar{v}_1(p) = g_1 p + \bar{C}_1 u_1(p), \tag{6}$$

with $\bar{C}_1 = \frac{s - g_1\bar{p}_1}{u_1(\bar{p}_1)}$. Player 2's equilibrium payoff for these beliefs is given by

$$\bar{v}_2(p) = s + \frac{(g_2 - s)p}{1+r} + \bar{C}_2 u_1(p) \tag{7}$$

with $\bar{C}_2 = -\frac{(g_2-s)\bar{p}_1}{(1+r)u_1(\bar{p}_1)}$.

Since $\bar{C}_1 > 0$ and $\bar{C}_2 < 0$, $\bar{v}_1$ is strictly convex and $\bar{v}_2$ is strictly concave.[3] The following lemma shows that the functions $\bar{v}_i$ intersect the corresponding diagonals D_i at a unique belief.

Lemma 1 *There exists a unique $p_1' \in (\bar{p}_1, 1)$ such that $\bar{v}_1(p_1') = D_1(p_1')$, and a unique $p_2' \in \left(\bar{p}_2, \frac{s}{g_2}\right)$ such that $\bar{v}_2(p_2') = D_2(p_2')$.*

Proof. The function $\bar{v}_1$ is strictly increasing, while D_1 is strictly decreasing. Furthermore, $\bar{v}_1(\bar{p}_1) < D_1(\bar{p}_1)$ and $\bar{v}_1(1) > D_1(1)$. Since both $\bar{v}_1$ and D_1 are moreover continuous, there exists a unique $p_1' \in (\bar{p}_1, 1)$ such that $\bar{v}_1(p_1') = D_1(p_1')$.

[3] $\bar{v}_1$ and $\bar{v}_2$ are obtained from Equations 13 and 15 respectively by imposing the condition $\bar{v}_i(\bar{p}_1) = s$ $(i = 1, 2)$.

As $\bar{C}_2 < 0$, we have

$$\bar{v}_2(\bar{p}_2) < s + \frac{[g_2 - s]\bar{p}_2}{1+r} = s + \frac{[g_2 - s]}{1+r}\frac{\mu_2 s}{(\mu_2+1)g_2 - s} \equiv \bar{\Psi}.$$

On the other hand, $D_2(\bar{p}_2) = s + \frac{s[g_2-s]}{\lambda_2[(\mu_2+1)g_2-s]}$. This implies

$$D_2(\bar{p}_2) - \bar{\Psi} = \frac{s[g_2 - s]}{[(\mu_2+1)g_2 - s]\,\lambda_2(1+r)} > 0.$$

Hence, $D_2(\bar{p}_2) > \bar{v}_2(\bar{p}_2)$. Furthermore, the function $\bar{v}_2$ is strictly increasing, and D_2 is strictly decreasing, on $\left(\bar{p}_2, \frac{s}{g_2}\right)$, while $\bar{v}_2\left(\frac{s}{g_2}\right) > D_2\left(\frac{s}{g_2}\right) = s$. Since both $\bar{v}_2$ and D_2 are moreover continuous, there exists a unique $p_2^{'} \in \left(\bar{p}_2, \frac{s}{g_2}\right)$ such that $\bar{v}_2(p_2^{'}) = D_2(p_2^{'})$. ■

In the following proposition, we will show that there exists an equilibrium in cutoff strategies if and only if the degree of asymmetry between the players is high enough, .

Proposition 3 *There exists a* $\lambda_2^* \in (\frac{s}{h}, 1)$ *such that there exists an equilibrium in cutoff strategies if and only if* $\lambda_2 \in (\frac{s}{h}, \lambda_2^*]$. *In this equilibrium, player* 1 *plays risky on* $(\bar{p}_1, 1]$ *and safe otherwise, while Player 2 plays risky on* $(p_2^{'}, 1]$ *and safe otherwise.*

Proof. By our previous arguments, in any equilibrium in cutoff strategies, player 1 will play risky on $(\bar{p}_1, 1]$ and safe otherwise. In response, by the definition of $p_2^{'}$, player 2 must play risky on $(p_2^{'}, 1]$ and safe otherwise, if there is an equilibrium in cutoff strategies. Indeed, below $p_2^{'}$, player 2 is playing a best response to player 1's action choice by the definition of $p_2^{'}$. Since D_2 is decreasing, it is sufficient to show that player 2's payoff function is increasing on $[p_2^{'}, 1]$ in order to show that he is also playing a best response at beliefs above $p_2^{'}$. Firstly, we note that the closed-form expression for player 2's payoff function (see Equation 11 in Appendix A.2) implies that player 2's payoff v_2 is strictly convex on $(p_2^{'}, 1)$, as $v_2(p_2^{'}) = D_2(p_2^{'}) > g_2 p_2^{'}$, where the inequality follows from $p_2^{'} < \frac{s}{g_2}$ (see Lemma 1). Furthermore, the relevant ODEs (Equations 14 and 10 in Appendix A.2) show that $v_2(p_2^{'}) = D_2(p_2^{'})$ implies smooth pasting at $p_2^{'}$. As moreover $\bar{v}_2^{\prime} > 0$ (as $\bar{C}_2 < 0$ and $u_1^{\prime} < 0$), we can conclude that player 2's value function is strictly increasing on $(p_2^{'}, 1)$ as well, and hence that player 2 is playing a best response at beliefs above $p_2^{'}$.

Thus, the candidate strategy profile is indeed an equilibrium if and only if player 1's strategy is a best response to player 2's. This requires player 2 to choose the safe arm for all beliefs at which player 1's payoff is below D_1. Thus, it remains to determine under what conditions $p_2^{'} \geq p_1^{'}$.

We will first argue that $p_1^{'}$ ($p_2^{'}$) is increasing (decreasing) in λ_2. Recall that $p_1^{'}$ is the point of intersection of the function $\bar{v_1}$ and the line D_1. As λ_2 decreases, the line D_1 rotates anticlockwise around the point $(\frac{s}{g_1}, s)$. Since $\bar{v_1}$ is independent of λ_2, $p_1^{'}$ decreases as λ_2 decreases. On the other

hand, as λ_2 decreases, the line D_2 shifts to the right and becomes steeper. By direct computation, one shows that $\bar{v_2}$ becomes flatter as λ_2 decreases. This implies that $p_2^{'}$ increases.

Consider the case $\lambda_2 \downarrow \frac{s}{h}$. Then, $D_2 \to s + \frac{(s-sp)}{\lambda_2}$. Thus, the belief $\hat{p}$ such that $D_2(\hat{p}) = s$ will tend to 1. Moreover, $\bar{v}_2 \to s$. Hence, $p_2^{'} \to 1$. However, D_1 still intersects the line s at $p = \frac{s}{g_1}$, implying that $p_1' \leq \frac{s}{g_1} < p_2^{'}$.

Next, we consider the case $\lambda_2 \uparrow 1$ and argue that there exists a left neighborhood of 1 such that, for all λ_2 in this neighborhood, $p_2^{'} < p_1^{'}$. By the ODEs for the $(1,0)$-region, $\bar{v}_2 > \bar{v}_1$ for all beliefs in $(\bar{p}_1, \check{p}]$, where $\check{p} = \frac{rs}{rg_1 + (g_1 - g_2)} < \frac{s}{g_1}$. Note that $\check{p} \uparrow \frac{s}{g_1}$ as $\lambda_2 \uparrow 1$. Furthermore, recall that p_1' is implicitly defined by

$$(1+\lambda_2)(g_1 p_1' - s) + \bar{C}_1 u_1(p_1') = 0,$$

where we note that $\bar{C}_1$ and u_1 are both independent of λ_2. This implies that (1) $p_1' < \frac{s}{g_1}$ for all $\lambda_2 \in [\frac{s}{h}, 1]$ (as $\bar{C}_1 > 0$ and $u_1 < 0$ for $p < 1$), and (2) that p_1' is a continuous function of λ_2 (by the Implicit Function Theorem). Therefore $\hat{p} = \max_{\lambda_2 \in [\frac{s}{h}, 1]} p_1' < \frac{s}{g_1}$. Thus, we can choose $\underline{\lambda}_2 \in (\frac{s}{h}, 1)$ such that, for all $\lambda_2 \in [\underline{\lambda}_2, 1]$, $\check{p} > \hat{p}$, and therefore $\bar{v}_2 > \bar{v}_1$ on $(\bar{p}_1, p_1']$. It thus follows that, for $\lambda_2 \in [\underline{\lambda}_2, 1]$, $\tilde{p}_2 < p_1'$, where $\tilde{p}_2$ is the belief where the function $\bar{v_2}$ intersects the line D_1. As $p_2' \downarrow \tilde{p}_2$ for $\lambda_2 \uparrow 1$, we can conclude that there exists some $\hat{\underline{\lambda}}_2 \in (\frac{s}{h}, 1)$ such that, for all $\lambda_2 \in (\hat{\underline{\lambda}}_2, 1)$, $p_2' < p_1'$. Thus, by monotonicity of $p_1^{'}$ and $p_2^{'}$ in λ_2, there exists a unique $\lambda_2^* \in (\frac{s}{h}, 1)$ such that $p_2^{'} \geq p_1^{'}$ if and only if $\lambda_2 \in (\frac{s}{h}, \lambda_2^*]$. ■

Appendix B.2 shows that the belief $p_2^{'}$ where player 2 switches to the safe arm in the above equilibrium is strictly greater than p_2^*, the threshold in the planner's solution. This shows that for $p \in (p_2^*, p_2^{'}]$, player 2 inefficiently free-rides.

The equilibrium in cutoff strategies is depicted in Figure 2. In this equilibrium, both players's payoffs are equal to s for $p \leq \bar{p}_1$. For $p \in (\bar{p}_1, p_2^{'}]$, player i's $(i = 1,2)$ payoff is $\bar{v}_i(p)$. For $p > p_2^{'}$, player i's payoff is given by

$$v_i^r(p) = g_i p + C_i^r u_0(p)$$

with $C_i^r = \frac{\bar{v}_i(p_2^{'}) - g_i p_2^{'}}{u_0(p_2^{'})}$.[4] Player 1's equilibrium payoff function is (strictly) convex (on $(\bar{p}_1, 1)$); it is smooth, except for a kink at $p_2^{'}$. Player 2's payoff function is strictly concave on $(\bar{p}_1, p_2^{'})$ and strictly convex on $(p_2^{'}, 1)$; it has an inflection point at $p_2^{'}$. It is smooth except for a kink at $\bar{p}_1$.

Recall the intuition for why there is no equilibrium in cutoff strategies in the homogeneous case (Keller, Rady, Cripps (2005)): There is a region of beliefs at which safe and risky are mutually best responses, and hence both players cannot be using the same cutoff. Now, the player applying the most pessimistic cutoff will have lower payoffs than the free-rider, as he will have to bear higher myopic opportunity costs in exchange for the same learning benefit. Therefore, the free-rider will cross into the region where risky is dominant at a more pessimistic belief than the pioneer, implying

[4] These payoffs are obtained from 11 by imposing the condition $v_i^r(p_2^{'}) = \bar{v}_i(p_2^{'})$ $(i = 1,2)$.

that the roles of free-rider and pioneer must switch at least once.

With heterogeneous players, by contrast, the respective regions in which risky is dominant no longer coincide for both players. Now, a higher value for player 2 no longer implies that he will cross into the region where playing risky is dominant at a more pessimistic belief than player 1. Geometrically, the diagonals D_1 and D_2 no longer coincide, as Figure 2 shows. Indeed, as the proof of Proposition 3 shows, the condition for existence of an equilibrium in cutoff strategies is precisely that player 2 will enter the region in which risky is dominant at a more optimistic belief than player 1 does, even though the latter's payoff function is lower at each belief. This is possible if and only if the region in which risky is dominant for player 2 is relatively small enough compared to that of player 1, i.e. if and only if λ_2 is small enough compared to $\lambda_1 = 1$.

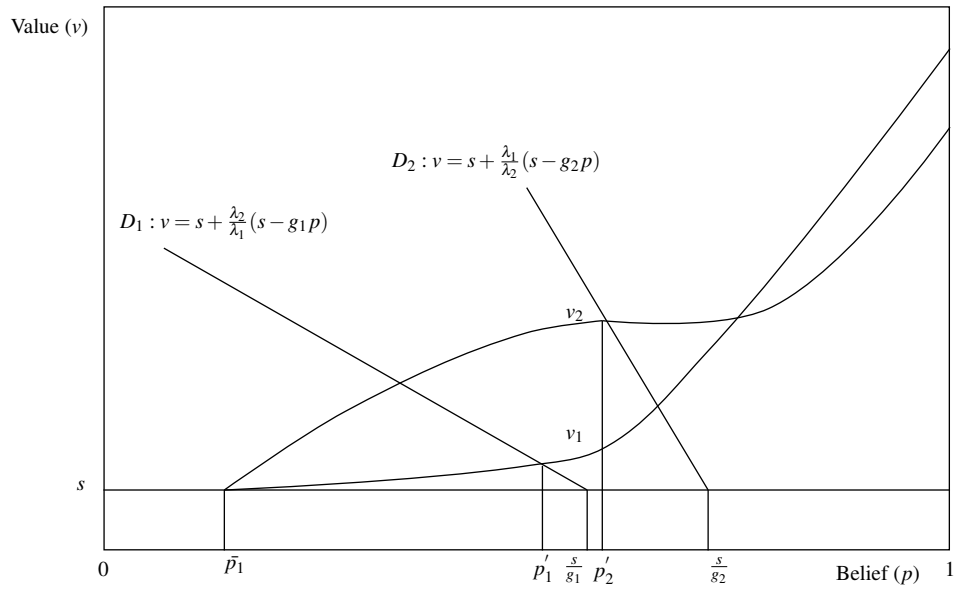

Figure 2.

4.2 Equilibria in non-cutoff strategies

In the previous subsection, we have identified a necessary and sufficient condition for the existence of an equilibrium in cutoff strategies. In this subsection, we will analyse equilibria where at least one of the players uses a non-cutoff strategy. Such equilibria will always exist, as the following proposition shows. To state the proposition, we let v_i be player i's equilibrium payoff. For both players $n \in \{1,2\}$, we define p_S^n as the (unique) point of intersection of v_n and D_n.[5] Let $p_S^i = \min\{p_S^1, p_S^2\}$.

[5]The uniqueness of $p_S^n \in (\bar{p}_1, \frac{s}{g_n})$ follows from (10).

Proposition 4 *For any $\lambda_2 \in (\frac{s}{h}, 1)$, there exists a continuum of Markov perfect equilibria in which at least one player uses a non-cutoff strategy. For each integer $k > 1$ and each sequence of threshold beliefs $(\tilde{p}_i)_{i=1}^k$ such that $\bar{p}_1 < \tilde{p}_1 < \cdots < \tilde{p}_k = p_S^i$, there exists an equilibrium such that both players play safe at all beliefs $p \leq \bar{p}_1$; player 1 plays risky and player 2 plays safe in $(\bar{p}_1, \tilde{p}_1] \cup \bigcup_{i \in 2\mathbb{N} \wedge i<k} (\tilde{p}_i, \tilde{p}_{i+1}]$, while player 1 plays safe and player 2 plays risky in $\bigcup_{i \in 2\mathbb{N} \wedge i \leq k} (\tilde{p}_{i-1}, \tilde{p}_i]$; on $(p_S^i, p_S^j]$, player i plays risky and player j plays safe, while both players play risky on $(p_S^j, 1]$. The same strategies with $k = 1$ also describe an equilibrium in which only player 2 uses a cutoff strategy if and only if $p_2^{'} = p_S^2 < p_S^1$.*

On $[0, \bar{p}_1]$, both players' value function is s. For even $i < k$, on $(\tilde{p}_i, \tilde{p}_{i+1}]$, player 1's (2's) value function is given by (13) ((15)), while on $(\tilde{p}_{i-1}, \tilde{p}_i]$, player 2's (1's) value function is given by (13) ((15)); on $(p_S^i, p_S^j]$, player i's (j's) payoff is given by (13) ((15)). On $(p_S^j, 1]$, both players' payoffs are given by (11). The constants of integration are determined by value matching.

Proof. That the proposed strategies are mutually best responses immediately follows from our discussion at the top of Section 4. That such equilibria always exist follows immediately from the continuity of players' payoff functions and the fact that $D_i(\bar{p}_1) > s$ for both $i \in \{1,2\}$. ■

As $\bar{p}_1 < \bar{p}_2$, the proposition implies that there exist equilibria in which player 2 experiments below his single-agent threshold $\bar{p}_2$. Indeed, by being the last player to experiment on $(\bar{p}_1, \tilde{p}_1]$, player 1 provides an *encouragement effect* to player 2, as the latter is willing to play risky on $(\tilde{p}_1, \tilde{p}_2]$ only because he knows that, should his experimentation not be successful, he will get to free-ride on player 1's experimentation once the belief will have dropped to $\tilde{p}_1$.

If the equilibrium in cutoff strategies exists, it allows player 2 to take maximal advantage of player 1's free-riding efforts. Therefore, the equilibrium in cutoff strategies is the worst (best) equilibrium for player 1 (player 2). Thus, in any equilibrium in which players swap the roles of pioneer and free-rider at least once, player 1's (2's) payoff will hit D_1 (D_2) at a more pessimistic (optimistic) belief than in the equilibrium in cutoff strategies, as the following proposition shows.

Proposition 5 *Consider any equilibrium described in Proposition 4. Suppose $p_S^1 > \bar{p}_1$ is the belief at which the equilibrium payoff of player 1 meets the line D_1 and $p_S^2 > \bar{p}_1$ be the belief at which the equilibrium payoff of player 2 meets the line D_2. Then, we have $p_S^1 < p_1^{'}$ and $p_S^2 > p_2^{'}$.*

Proof. It is sufficient to show that $\bar{v}_1 < v_1$ and $\bar{v}_2 > v_2$ on $(\tilde{p}_1, p_S^j]$, where v_n is player n's equilibrium payoff function.

Note that $\bar{v}_n(\tilde{p}_1) = v_n(\tilde{p}_1)$ for both $n \in \{1,2\}$ and suppose that $\bar{v}_2(\tilde{p}_{i-1}) \geq v_2(\tilde{p}_{i-1})$ and $\bar{v}_1(\tilde{p}_{i-1}) \leq v_1(\tilde{p}_{i-1})$ for some $i \in \{2, \cdots, k\}$. Suppose that $i-1 \geq 1$ is odd, and let v_1^{rr} be player 1's payoff from deviating to playing risky on $(\tilde{p}_{i-1}, \tilde{p}_i]$. By construction, $v_1^{rr}(\tilde{p}_{i-1}) = v_1(\tilde{p}_{i-1}) \geq \bar{v}_1(\tilde{p}_{i-1})$. Suppose to the contrary that there exists a belief $p \in (\tilde{p}_{i-1}, \tilde{p}_i]$ such that $\bar{v}_1(p) = v_1^{rr}(p)$. The relevant ODEs ((12) and (10)) imply that $v_1^{rr'}(p-) > \bar{v}_1'(p-)$. As $v_1^{rr}(\tilde{p}_{i-1}) = v_1(\tilde{p_{i-1}}) \geq \bar{v}_1(\tilde{p}_{i-1})$, this

implies that there exists a $\hat{p} \in [\tilde{p}_{i-1}, \tilde{p}_i)$ such that $v_1^{rr}(\hat{p}) = \bar{v}_1(\hat{p})$ and $v_1^{rr'}(\hat{p}+) < \bar{v}_1'(\hat{p}+)$, a contradiction to (12) and (10). By the same token, suppose that there exists a belief $p \in (\tilde{p}_{i-1}, \tilde{p}_i]$ such that $v_2(p) = \bar{v}_2(p)$. As $s > pg_2$, the relevant ODEs ((12) and (14)) imply that $\bar{v}_2'(p-) > v_2'(p-)$. As $\bar{v}_2(\tilde{p}_{i-1}) \geq v_2(\tilde{p}_{i-1})$, this implies that there exists a $\hat{p} \in [\tilde{p}_{i-1}, \tilde{p}_i)$ such that $v_2(\hat{p}) = \bar{v}_2(\hat{p})$ and $v_2'(\hat{p}+) > \bar{v}_2'(\hat{p}+)$, a contradiction to (12) and (14).

Now, let $i-1 \geq 2$ be even. Note that our previous step implies that $\bar{v}_2(\tilde{p}_{i-1}) > v_2(\tilde{p}_{i-1})$ and $\bar{v}_1(\tilde{p}_{i-1}) < v_1(\tilde{p}_{i-1})$. Suppose that there exists a $p \in (\tilde{p}_{i-1}, \tilde{p}_i]$ such that $v_n(p) = \bar{v}_n(p)$ for an $n \in \{1,2\}$. As $(k_1, k_2) = (1,0)$ on $(\tilde{p}_{i-1}, \tilde{p}_i]$, this immediately implies that $v_n(\tilde{p}_{i-1}) = \bar{v}_n(\tilde{p}_{i-1})$, a contradiction.

On $(\tilde{p}_k, p_S^j]$, a similar argument to the case of even (odd) $i-1$ applies if $j = 2$ ($j = 1$), so that we can conclude that $\bar{v}_1 < v_1$ and $\bar{v}_2 > v_2$ on $(\bar{p}_1, p_S^j]$, and hence $p_S^1 < p_1'$ and $p_S^2 > p_2'$. ■

Propositions 4 and 5 imply that an equilibrium in which only player 2 uses a cutoff strategy (the equilibria corresponding to $k = 1$ in Proposition 4) exists if and only if $\lambda_2 > \lambda_2^*$, as the following corollary shows. In the limit $\lambda_2 \downarrow \lambda_2^*$, this equilibrium coincides with the equilibrium in cutoff strategies.

Corollary 1 *There exists an equilibrium in which only player 2 uses a cutoff strategy if and only if* $\lambda_2 > \lambda_2^*$.

Proof. If $\lambda_2 \leq \lambda_2^*$, $p_1' \leq p_2'$, by the proof of Proposition 3. Suppose to the contrary that the equilibrium in which only player 2 uses a cutoff exists. By Proposition 5, $p_S^1 < p_1' \leq p_2' < p_S^2$, a contradiction to the characterisation of this equilibrium in Proposition 4.

Now, suppose $\lambda_2 > \lambda_2^*$. By the proof of Proposition 3, $p_1' > p_2'$. It thus remains to show that $p_S^1 > p_2'$. Yet, player 1's payoff from the conjectured equilibrium strategies at p_2' is given by $\bar{v}_1(p_2') < D_1(p_2')$, the inequality being immediately implied by $p_1' > p_2'$, we have $p_S^1 > p_2'$, and, by Proposition 4, the equilibrium exists. ■

Suppose $\lambda_2 \in (\frac{s}{h}, \lambda_2^*]$. This implies that the equilibrium in cutoff strategies exists. Propositions 4 and 5 allow us to compare the experimentation intensities. First, observe that in the equilibrium in cutoff strategies, both players experiment for beliefs greater than p_2'. Since $p_S^2 > p_2'$ (by Proposition 5), the range of beliefs where both players experiment is greater in the equilibrium in cutoff strategies than in any other equilibrium. Next, in the equilibrium in cutoff strategies, whenever only one player experiments, it is the player with the higher payoff arrival rate, player 1. In any other equilibrium, however, there is a range of beliefs where player 2 plays the role of the lonely pioneer. Since, in any equilibrium, all experimentation ceases at $\bar{p}_1$, the intensity of experimentation is thus highest in the equilibrium in cutoff strategies. The following proposition is thus no surprise.

Proposition 6 *Suppose $\lambda_2 \leq \lambda_2^*$ and let v_{agg}^c be the aggregate equilibrium payoff in the equilibrium in cutoff strategies and v_{agg}^{nc} be the aggregate equilibrium payoff in an arbitrary equilibrium in non-cutoff strategies. Then, $v_{agg}^c \geq v_{agg}^{nc}$, with the inequality strict on $(\tilde{p}_1, 1)$.*

Proof. If player i ($i = 1,2$) experiments and player j ($j = 1,2,\ j \neq i$) free rides then the players' aggregate equilibrium payoff is given by $v_{agg} = v_i + v_j$, with v_i satisfying the ODE (12) and v_j satisfying the ODE (14). If both players experiment then $v_{agg} = v_1 + v_2$ and v_n ($n = 1,2$) satisfy the ODE (10).

From proposition (4), we know that $v_{agg}^c(\tilde{p}_1) = v_{agg}^{nc}(\tilde{p}_1)$. Suppose $v_{agg}^c(\tilde{p}_{i-1}) \geq v_{agg}^{nc}(\tilde{p}_{i-1})$ for some $i \in \{2,3,....,k\}$. Suppose first that $i-1 \geq 1$ is odd. If there exists a $p \in (\tilde{p}_{i-1}, \tilde{p}_i]$ such that $v_{agg}^c(p) = v_{agg}^{nc}(p)$, then by the ODEs (12) and (14), we can conclude that $v_{agg}^{c'}(p-) > v_{agg}^{nc'}(p-)$. This implies there exists a $\hat{p} \in [\tilde{p}_{i-1}, p)$ such that $v_{agg}^c(\hat{p}) = v_{agg}^{nc}(\hat{p})$ and $v_{agg}^{c'}(\hat{p}+) < v_{agg}^{nc'}(\hat{p}+)$, a contradiction to ODEs (12) and (14).

Suppose $i-1 \geq 2$ is even. Then from the previous step we can infer that $v_{agg}^c(\tilde{p}_{i-1}) > v_{agg}^{nc}(\tilde{p}_{i-1})$. In both kinds of equilibria, if $i-1$ is even, $(k_1,k_2) = (1,0)$ on $(\tilde{p}_{i-1}, \tilde{p}_i]$. This implies that we have $v_{agg}^c(p) > v_{agg}^{nc}(p)$ for all $p \in (\tilde{p}_{i-1}, \tilde{p}_i]$. Thus, for all $p \in (\tilde{p}_1, \tilde{p}_k]$, $v_{agg}^c(p) > v_{agg}^{nc}(p)$.

As $\lambda_2 \leq \lambda_2^*$, we have $\tilde{p}_k = p_S^1$. An argument similar to that for even $i-1$ shows that $v_{agg}^c > v_{agg}^{nc}$ on $p \in (p_S^1, p_2^{'}]$. Now, suppose that there exists a $\hat{p} \in (p_2^{'}, p_S^2]$ such that $v_{agg}^c(\hat{p}) = v_{agg}^{nc}(\hat{p})$. By the ODEs (12) and (10), this implies $v_{agg}^{c'}(\hat{p}-) > v_{agg}^{nc'}(\hat{p}-)$. This leads to a contradiction by the same argument as above. As $(k_1,k_2) = (1,1)$ prevails in both equilibria on $(p_S^2, 1)$, the claim follows. ■

The comparison between the equilibrium in cutoff strategies and an equilibrium in which players swap roles once is depicted in figure $4(a)$.[6] Figures $4(b)$ and $4(c)$ depict the actions of players in the equilibrium in cutoff strategies and the 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 in Figure 4 (a) depict the payoffs to player 1 and 2 respectively in the equilibrium in cutoff strategies. In the equilibrium in non-cutoff strategies, payoffs coincide for beliefs less than or equal to $\tilde{p}_1$. At $\tilde{p}_1$, players switch arms. The blue curve depicts the payoff to player 1 and the red curve depicts the payoff to player 2 in the equilibrium in non-cutoff strategies for $p > \tilde{p}_1$. 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 switches to the risky arm and player 2 switches to the safe arm. When the red curve meets the line D_2 at $p_S^2 > p_2^{'}$, player 2 switches to the risky arm again.

By Corollary 1, the equilibrium in which only player 2 uses a cutoff strategy (i.e. the equilibrium corresponding to the case $k = 1$ in Proposition 4, which is depicted in Figure 5) exists if and only if $\lambda_2 > \lambda_2^*$. If $\lambda_2 \leq \lambda_2^*$, meanwhile, the equilibrium in cutoff strategies exists (see

[6] Lemma 5 implies that the qualitative characteristics of p_s^1 and p_s^2 are the same in any equilibrium in non-cutoff strategies.

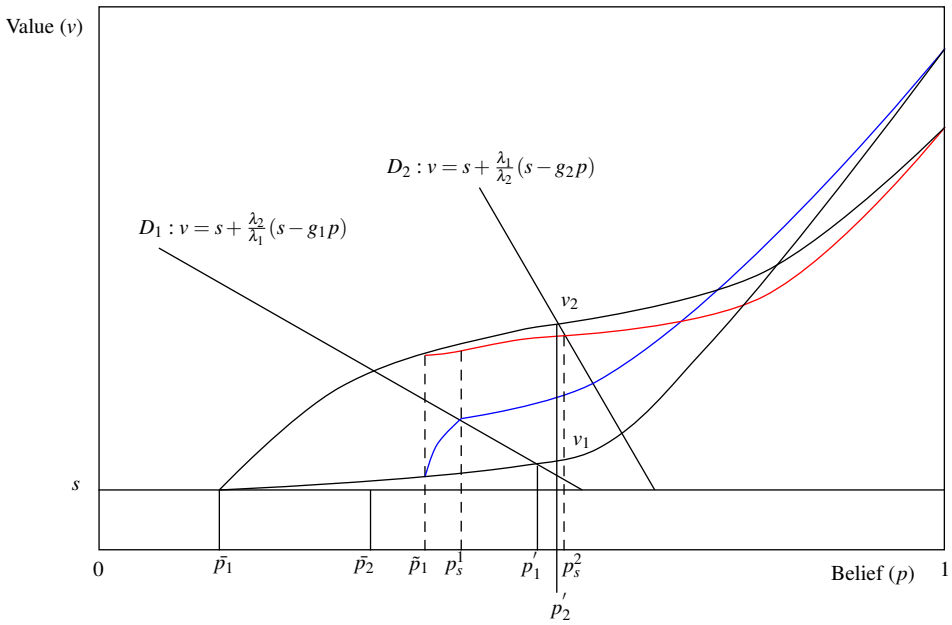

Figure 4(a).

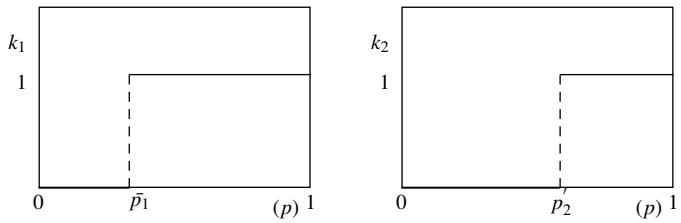

Figure 4(b): Actions of players in the equilibrium in cutoff strategies.

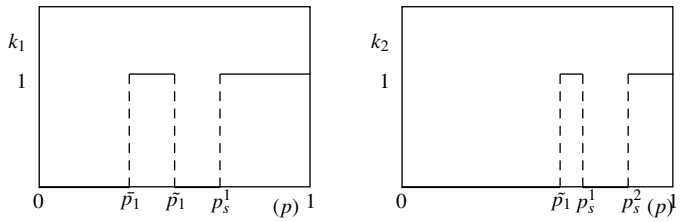

Figure 4(c): Actions of players in the equilibrium in non-cutoff strategies.

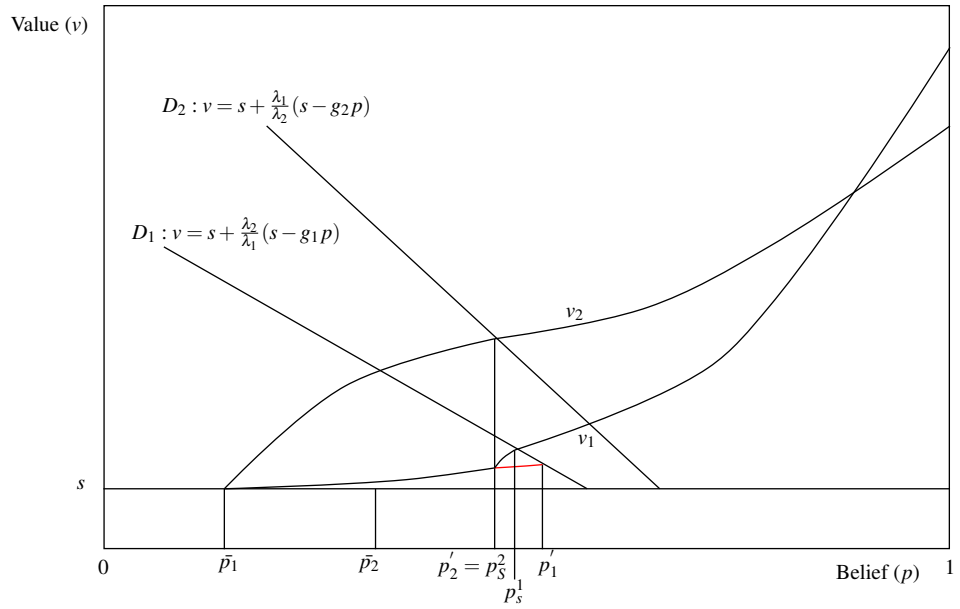

Figure 5.

Proposition 3). Thus, there always exists an equilibrium in which player 2 plays a cutoff strategy. By an argument similar to that in the proof of Proposition 6, one can show that, on $(\bar{p}_1, p_2^{'}]$, the equilibrium of Corollary 1, which is the only equilibrium in which player 1 is experimenting throughout this range, strictly welfare-dominates all other equilibria. Yet, the belief region in which $(k_1, k_2) = (1,1)$ prevails may be larger in equilibria in which neither player uses a cutoff strategy, making a welfare ranking for all beliefs hard to establish.

5 Conclusion

In this paper, we have characterised the set of Markov perfect equilibria in a two-armed bandit model with heterogeneous players. We have shown that there always exists an equilibrium in which the weaker player uses a cutoff strategy. If the heterogeneity is stark enough, there exists an equilibrium in cutoff strategies. If such an equilibrium exists, it is welfare-optimal.

We have restricted players to using one arm only at any given instant t. By the linearity of the players' Bellman equations, our equilibria would remain equilibria if we allowed players to select experimentation intensities $k_{i,t} \in [0,1]$. There might, however, be more equilibria in this case.

We have focussed our attention on asymmetries in the players' conditional lump-sum arrival rates given that their risky arm is good. Such asymmetries pertain to both payoffs and the learning process. Our analysis has relied heavily on the characterisation of players' best responses via the diagonals D_i (see Equation (4)), which was pioneered by Keller, Rady, Cripps (2005) for the homogeneous-player case. We expect that a similar approach could, *mutatis mutandis*, be used to study other kinds of asymmetries, e.g. pertaining to players' safe-arm payoffs s_i or risky-arm

payoffs h_i. We should expect a similar result to our Proposition 3 to hold in these cases, namely that there existed an equilibrium in cutoff strategies if and only if the heterogeneity was stark enough. The analysis of otherwise symmetric players that hold different priors might be more complex, however, as the informativeness of a given amount of experimentation would now differ between players. We commend these questions for future research.

References

Besanko, D., Wu, J., 2013: "The impact of market structure and learning on the tradeoff between R & D competition and cooperation ", *The Journal of Industrial Economics LXI*, 166 – 201

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

Bonatti, A., Hörner, J., 2011: "Collaborating ", *American Economic Review* 101, 632 – 663

Heidhues, P., Rady, S., Strack, P., 2015: "Strategic experimentation with private payoffs ", *Journal of Economic Theory* 159, 531 – 551

Hörner, J., Skrzypacz, A., 2016: "Learning, Experimentation and Information Design ", *Working paper, Stanford University*.

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

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

Klein, N., 2013: "Strategic Learning in Teams ", *Games and Economic Behavior* 82, 632 – 657.

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

Malueg, D., Tsutsui, S., 1997: "Dynamic R & D Competition with Learning ", *The RAND Journal of Economics* 28, 751 – 772

Murto, P., Välimäki, J., 2011: "Learning and Information Aggregation in an Exit Game ", *Review of Economic Studies* 78, 1426 – 1461.

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

Rosenberg, D., Salomon, A., Vieille, N., 2013: "On games of Strategic Experimentation ", *Games and Economic Behavior* 82, 31 – 51.

Rosenberg, D., Solan, E., Vielle, N., 2007: "Social Learning in One-arm Bandit Problems ", *Econometrica* 75, 1511 – 1611.

Thomas, C., 2017: "Experimentation with Congestion ", *mimeo, University of Texas Austin*.

APPENDIX

A Ordinary Differential Equations

A.1 ODEs in the planner's problem

Clearly, if $(k_1,k_2)=(0,0)$ is played at a belief p, the planner's payoff function satisfies $v(p)=2s$.

If the planner plays $k_1=k_2=1$ on an open set of beliefs, his payoff function on this set satisfies

$$v(p)=2s+B_1(p,v)-c_1(p)+B_2(p,v)-c_2(p),$$

which is equivalent to the ODE

$$\lambda p(1-p)v^{'}(p)+(r+\lambda p)v(p)=(r+\lambda)pg. \tag{8}$$

This is solved by

$$v(p)=gp+Cu_0(p)$$

where C is a constant of integration.

By the same token, the ODE for $(k_1,k_2)=(1,0)$ is given by

$$p(1-p)v^{'}(p)+(r+p)v(p)=r(s+pg_1)+pg. \tag{9}$$

This is solved by

$$v(p)=s+\left[\frac{g+rg_1}{1+r}-\frac{s}{1+r}\right]p+Cu_1(p).$$

A.2 ODEs of players in the non-cooperative game

If $k_1=k_2=0$, both players' payoff functions satisfy $v_i(p)=s$.

If $k_1 = k_2 = 1$ prevails on an open set of beliefs in the non-cooperative game, both players' value function for beliefs in this set satisfies

$$\lambda p(1-p)v_i'(p) + (r+\lambda p)v_i(p) = (r+\lambda)pg_i. \tag{10}$$

This is solved by

$$v_i = g_i p + Cu_0(p) \tag{11}$$

where C is a constant of integration.

If $k_i = 1$ and $k_j = 0$, player i's payoff function satisfies

$$\lambda_i p(1-p)v_i'(p) + (r+\lambda_i p)v_i(p) = (r+\lambda_i)pg_i. \tag{12}$$

Solving this, we get

$$v_i(p) = g_i p + Cu_i(p) \tag{13}$$

where $u_i(p) = (1-p)[\frac{(1-p)}{p}]^{\mu_i}$ and $\mu_i = \frac{r}{\lambda_i}$. Player j's payoff function satisfies

$$\lambda_i p(1-p)v_j'(p) + (r+\lambda_i p)v_j(p) = rs + \lambda_i p g_j. \tag{14}$$

This is solved by

$$v_j = s + \frac{\lambda_i}{\lambda_i + r}(g_j - s)p + Cu_i(p). \tag{15}$$

B Other Proofs

B.1 Proof of Proposition 1

The function v satisfies $v = 2s$ on $[0, p_1^*]$, $v = 2s + B_1 - c_1$ on $(p_1^*, p_2^*]$ and $v = 2s + B_1 - c_1 + B_2 - c_2$ on $(p_2^*, 1]$;[7] thus, v is the payoff function associated with the policy k^*.[8] We shall first show that v is of class C^1, (strictly) increasing and (strictly) convex (on $(p_1^*, 1)$).

One computes that, for $p \in (p_1^*, p_2^*)$, $B_1(p,v) - c_1(p) = \psi(p)$, where ψ is defined as

$$\psi(p) = -s + pg_1 + \frac{1}{r}p\left[g - s - \frac{g + rg_1}{1+r} + \frac{s}{1+r} + \frac{r}{p}\left(s - p_1^*\left(\frac{g+rg_1}{1+r} - \frac{s}{r+1}\right)\right)\frac{u_1(p)}{u_1(p_1^*)}\right].$$

Direct computation shows that $u_1'' > 0$ and $s - p_1^*\left(\frac{g+rg_1}{1+r} - \frac{s}{1+r}\right) > 0$, so that ψ, and hence $v|_{(p_1^*, p_2^*)}$ is

[7] We suppress arguments whenever this is convenient.

[8] In Appendix A.1, we display the ODEs that v satisfies for each range of beliefs and the corresponding general form of v for that range. The specific value of v is obtained by value matching.

strictly convex. One furthermore shows by direct computation that $\psi(p_1^*) = \psi'(p_1^*) = 0$, implying that $v|_{(0,p_2^*)}$ is of class C^1.

We shall now show that p_2^* is well-defined. Indeed, by definition, $x = p_2^*$ must satisfy

$$\left[\frac{g+rg_1}{1+r} - \frac{s}{1+r}\right]x + \left[s - \left(\frac{g+rg_1}{1+r} - \frac{s}{1+r}\right)p_1^*\right]\frac{u_1(x)}{u_1(p_1^*)} = \frac{s}{\lambda_2}.$$

The left-hand side of this equation is strictly increasing in x for $x > p_1^*$ and equal to $s < \frac{s}{\lambda_2}$ at $x = p_1^*$. Furthermore, at $x = \frac{s}{g_2}$, the left-hand side exceeds $\left[\frac{g+rg_1}{1+r} - \frac{s}{1+r}\right]\frac{s}{g_2} > \frac{s}{\lambda_2}$. By continuity, the equation thus admits of a unique root $p_2^* \in (p_1^*, \frac{s}{g_2})$.

As $p_2^* < \frac{s}{g_2}$, $\frac{\lambda}{\lambda_2}s - p_2^* g > 0$, and $v|_{[p_2^*,1]}$ is strictly convex as well. It remains to show that $v|_{[p_2^*,1]}$ is also strictly increasing. By convexity, it is sufficient to show smooth pasting at p_2^*. By the ODE for the region (p_1^*, p_2^*) (Equation 9 in Appendix A.1), we have $p_2^*(1-p_2^*)v'(p_2^*-) = \left[rs + rp_2^* g_1 + p_2^* g - (r+p_2^*)\frac{\lambda}{\lambda_2}s\right]$. By the ODE for the $(p_2^*, 1)$-region (Equation 8 in Appendix A.1), we find $p_2^*(1-p_2^*)v'(p_2^*+) = \left[(r+\lambda)p_2^* g - (r+\lambda p_2^*)\frac{\lambda}{\lambda_2}s\right]/\lambda$, and hence $v'(p_2^*+) = v'(p_2^*-)$.

It remains to show that v solves the Bellman equation, i.e. that $B_i \leq c_i$ for both $i \in \{1,2\}$ on $[0, p_1^*]$; $B_1 \geq c_1$ and $B_2 \leq c_2$ on $(p_1^*, p_2^*]$; and $B_i \geq c_i$ for both $i \in \{1,2\}$ on $(p_2^*, 1]$. First, let $p \in [0, p_1^*]$. In this case, $v = 2s$, and $B_i \leq c_i$ if and only if $p \leq \frac{rs}{rg_i + \lambda_i(g-2s)}$, which is verified for all $p \leq p_1^*$. Now, let $p \in (p_1^*, p_2^*]$. As v is strictly increasing in this range, $v = 2s + B_1 - c_1 > 2s$, and thus $B_1 > c_1$. Moreover, $v = 2s + B_1 - c_1$ implies that $B_2 = \lambda_2 B_1 = \lambda_2(v - s - pg_1) \leq s - pg_2 = c_2$ if and only if $v \leq \frac{\lambda}{\lambda_2}s$, which is verified as $p \leq p_2^*$. Finally, let $p \in (p_2^*, 1)$. In this range, we have that $g - v - (1-p)v' = \frac{r}{\lambda p}\left(\frac{\lambda}{\lambda_2}s - p_2^* g\right)\frac{u_0(p)}{u_0(p_2^*)}$, so that $B_i = \frac{\lambda_i}{\lambda}v - pg_i$, which exceeds $c_i = s - pg_i$ if and only if $v \geq \frac{\lambda}{\lambda_i}s$. By monotonicity of v, $v \geq \frac{\lambda}{\lambda_2}s > \lambda s$ in this range, which completes the proof.

B.2 To show that $p_2^* < p_2^{'}$

Recall from the proof of Proposition 2 that p_2^* was implicitly defined as the unique root of the (for $p > p_1^*$) strictly increasing function ζ, where

$$\zeta(p) = \left[g_1 + \frac{g_2 - s}{1+r}\right]p + \left[s - \left(g_1 + \frac{g_2 - s}{1+r}\right)p_1^*\right]\frac{u_1(p)}{u_1(p_1^*)} - \frac{s}{\lambda_2}.$$

By the same token, $p_2^{'}$ is implicitly defined by $\bar{v}_2(p_2^{'}) = D_2(p_2^{'})$, which is equivalent to

$$\frac{g_2 - s}{1+r}p_2^{'} + p_2^{'}g_1 - \frac{s}{\lambda_2} = \frac{g_2 - s}{1+r}\bar{p}_1\frac{u_1(p_2^{'})}{u_1(\bar{p}_1)}.$$

As $p_2^{'} > \bar{p}_2 > p_1^*$, it remains to show that

$$\zeta(p_2^{'}) = \frac{g_2 - s}{1+r}\bar{p}_1 \frac{u_1(p_2^{'})}{u_1(\bar{p}_1)} + \left[s - \left(g_1 + \frac{g_2 - s}{1+r} \right) p_1^* \right] \frac{u_1(p_2^{'})}{u_1(p_1^*)} > 0.$$

For this, it is sufficient that

$$s - \left(g_1 + \frac{g_2 - s}{1+r} \right) p_1^* > 0,$$

which follows by direct computation.