Negatively Correlated Bandits*

Nicolas Klein† Sven Rady‡

This version: January 23, 2008

Abstract

We analyze a two-player game of strategic experimentation with two-armed bandits. Each player has to decide in continuous time whether to use a safe arm with a known payoff or a risky arm whose likelihood of delivering payoffs is initially unknown. The quality of the risky arms is perfectly *negatively* correlated between players. In marked contrast to the case where both risky arms are of the same type, we find that learning will be complete in any Markov perfect equilibrium if the stakes exceed a certain threshold, and that all equilibria are in cutoff strategies. For low stakes, the equilibrium is unique, symmetric, and coincides with the planner's solution. For high stakes, the equilibrium is unique, symmetric, and tantamount to myopic behavior. For intermediate stakes, there is a continuum of equilibria.

*We are grateful to Philippe Aghion, Patrick Bolton, Martin Cripps, Matthias Dewatripont, Stephen Ryan, as well as seminar participants at Bonn, Bronnbach, and Munich for helpful comments and suggestions. Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 and GRK 801 is gratefully acknowledged.

†Munich Graduate School of Economics, Kaulbachstr. 45, D-80539 Munich, Germany; email: kleinnic@yahoo.com.

‡Department of Economics, University of Munich, Kaulbachstr. 45, D-80539 Munich, Germany; email: sven.rady@lrz.uni-muenchen.de.

1 Introduction

Two-armed bandit problems as a means of modeling the trade-off between experimentation and exploitation have already received a quite extensive treatment in the literature thus far, even though most of the interest has been decision-theoretic, i.e. focussing on single-agent problems. Only recently has strategic interaction been introduced into the model: Bolton and Harris (1999, 2000) analyze the case of Brownian motion bandits, while Keller, Rady, Cripps (2005) as well as Keller and Rady (2007) analyze Poisson bandits. All of the previous literature, however, has assumed perfect *positive* correlation across bandits; what was good news to any given player was assumed to be good news for everybody else.

In the real world, however, situations abound where one man's boon is the other one's bane. Think of a suit at law, for instance: Whatever is good news for one party is bad news for the other. An appropriate model of strategic interaction in such a setup would, as a matter of course, assume correlation across bandits was negative. This we propose to do in the present paper.[1]

In this respect, our work is related to Dewatripont and Tirole (1999), who, in a moral hazard setting, which bears no resemblance to ours, pose the question whether it is socially better to adjudicate disputes through a centralized system of gathering evidence, which they assimilate to the inquisitorial system of Civil Law countries, or whether the interests of justice may be better served in a decentralized, adversarial system, as it is found in the Common Law countries. They show that, in a centralized system, it is not possible to give adequate incentives to make sure the truth is uncovered, and conclude that the Common Law system of gathering information was therefore superior. Our model provides an alternative framework to ascertain the effectiveness of information-gathering processes in a strategic setting where the parties' interests are diametrically opposed.

Rather surprisingly, in our setup, all Markov perfect equilibria are in cutoff strategies (i.e. of the form "play risky at beliefs more optimistic than some threshold, and safe otherwise"). While this structure of optimal strategies is prevalent in single-agent bandit problems, our result is in stark contrast to Keller, Rady and Cripps (2005), who find that there is *no* such equilibrium when all risky arms are of the same type.

On account of the symmetry of the situation, it is not surprising that there always exists a symmetric equilibrium, where both players use the same cutoff. This symmetric

[1]There is a decision-theoretic literature on correlated bandits which analyzes correlation across different arms of a bandit operated by a single agent; see e.g. Camargo (2007) for a recent contribution to this literature, or Pastorino (2005) for economic applications. Our focus here is quite different, though, in that we are concerned with correlation between different bandits operated by two players who interact strategically.

equilibrium is unique. What is more, we are able to show that for a large set of parameter values, there is no other equilibrium besides the symmetric one. This uniqueness result is again in sharp contrast to the multiplicity of equilibria in Keller, Rady and Cripps (2005).

When the stakes (as measured by the payoff advantage of a good risky arm over a safe one) are low, the unique equilibrium is efficient. This is due to the fact that, with low stakes, single-agent cutoff beliefs are so optimistic that the two bandit problems essentially fail to interact. Hence, the social planner will treat them as two separate problems, and will let the players behave as though they were single agents, which is then obviously consistent with equilibrium.

When the stakes are high, equilibrium is again unique, and it amounts to the players' behaving myopically, and hence inefficiently. With the stakes high, players are so eager to play risky that there exists a range of beliefs where both are experimenting. Of course, when both are doing the same thing with the same result, there is no new information made available. Thus, the players are essentially shutting down the incremental learning process, keeping the belief at its current value and effectively freezing the problem in its current state. This, however, they are only willing to do if the current state is attractive from a myopic perspective.

If the stakes are intermediate in size, there is a continuum of equilibria. As the stakes gradually increase and we move from the low to the intermediate case, at first, given *any* initial belief, there still exists among the continuum of equilibria one that is efficient. As stakes increase further, there then appears a range of initial beliefs for which no equilibrium achieves efficiency. As we move from high stakes down to intermediate stakes, there at first always exists an equilibrium that involves *one* player behaving myopically. To achieve this, the other player has to bear the entire load of experimentation by himself when the uncertainty is greatest. As stakes gradually grow lower, however, the other player will at some point no longer be willing to bear this burden, and the equilibrium disappears.

Our strategic setting assumes that either player's actions and payoffs are perfectly observable to the other player. Hence, information that is garnered via experimentation is a public good. Therefore, intuition would suggest, and the previous literature would confirm, that levels of experimentation were depressed by an inherent free-rider problem and learning would often cease prematurely. Our analysis, however, shows that incompleteness of information gathering, which has been prevalent throughout the previous strategic experimentation literature, can be overcome by competition between two antagonistic agents. More precisely, we find that, in equilibrium, learning will be complete whenever this is efficient. Thus, whenever society places a lot of emphasis on uncovering the truth, as one may argue is the case with the justice system, our analysis would suggest an adversarial setup

was superior.[2]

Our analysis would furthermore suggest that competition was conducive to investment in risky experimentation. This is consistent with the recent literature on growth, innovation and R&D which shows that decentralization of decision structures boosts investments in innovation; see, for example, Aghion and Howitt (2006), Acemoglu et al. (2006), Aghion et al. (2006).

The rest of the paper is structured as follows. Section 2 introduces the model. Section 3 analyzes the planner's solution. Section 4 sets up the non-cooperative game. Section 5 discusses long-run properties of learning in equilibrium. Section 6 characterizes the Markov perfect equilibria of the non-cooperative game. Section 7 concludes. Proofs are provided in the appendix.

2 The Model

There are two players, 1 and 2, each of whom faces a two-armed bandit problem in continuous time. Bandits are of the exponential type studied in Keller, Rady and Cripps (2005). One arm is safe in that it yields a known payoff flow of s; the other arm is risky in that it is either good or bad. If it is bad, it never yields any payoff; if it is good, it yields a lump-sum payoff with probability λdt when used over a time of length dt. Let $g\,dt$ denote the corresponding expected payoff increment; thus, g is the product of the arrival rate λ and the average size of a lump-sum payoff. The time-invariant constants $\lambda > 0$ and $g > 0$ are common knowledge. It is also common knowledge that exactly one bandit's risky arm is good. To have an interesting problem, we assume that the expected payoff of a good risky arm exceeds that of the safe arm, whereas the safe arm is better than a bad risky arm, i.e., $g > s > 0$.

The strategic link between the two players' actions is provided by the assumption that players perfectly observe each other's actions and payoffs. Thus, as the bandits are perfectly negatively correlated, any information that is garnered about the quality of the risky arm is a public good. At the outset of the game, players have a common prior about which of the risky arms is good. Since the results of each player's experimentation are public, players share a common posterior at all times. Let p_t be the players' posterior probability assessment

[2]Our complete learning result carries over to the situation where the players' actions are publicly observable, but their payoffs are private information. In this respect, our work is related to the growing literature on strategic learning with private information; cf. Moscarini and Squintani (2004), Hopenhayn and Squintani (2006), Murto and Välimäki (2006), and Rosenberg, Solan and Vieille (2007a, b).

that player 1's risky arm is good. This common posterior will serve as the problem's state variable, as it encapsulates all relevant information about the decision problem.

Each player chooses actions $\{k_t\}_{t\geq 0}$ such that $k_t \in \{0,1\}$ is measurable with respect to the information available at time t, with $k_t = 1$ indicating use of the risky arm, and $k_t = 0$ use of the safe arm. The player's total expected discounted payoff, expressed in per-period units, is

$$\mathrm{E}\left[\int_0^\infty r\, e^{-r\,t}\ [(1-k_t)s + k_t p_t g]\ dt\right],$$

where the expectation is taken over the stochastic processes $\{k_t\}$ and $\{p_t\}$, and r is the players' common discount rate.

The belief jumps to 1 if there has been a breakthrough on player 1's bandit, and to 0 if there has been a breakthrough on player 2's bandit, where in either case it will remain ever after. If the players choose the actions k_1 and k_2 over the time interval $[t, t+\Delta]$ and there is no breakthrough on either bandit, Bayes' rule gives us

$$p_{t+\Delta} = \frac{p_t e^{-\lambda k_1 \Delta}}{p_t e^{-\lambda k_1 \Delta} + (1-p_t)e^{-\lambda k_2 \Delta}},$$

and so the belief solves the ordinary differential equation

$$\dot{p} = -(k_1 - k_2)\lambda p(1-p).$$

Note that for $k_1 = k_2 = 1$, the belief remains unchanged up to the first breakthrough on a risky arm.

3 The Planner's Problem

In this section, we shall be examining a benevolent utilitarian social planner's behavior in our setup. Proceeding exactly as Keller, Rady and Cripps (2005), we can write the Bellman equation for the maximization of the average payoff from the two bandits as

$$u(p) = s + \max_{(k_1,k_2)\in\{0,1\}^2} \left\{ k_1 \left[B_1(p,u) - \frac{c_1(p)}{2} \right] + k_2 \left[B_2(p,u) - \frac{c_2(p)}{2} \right] \right\},$$

where $B_1(p,u) = \frac{\lambda}{r} p[\frac{g+s}{2} - u(p) - (1-p)u'(p)]$ measures the expected benefit of playing risky arm 1, $B_2(p,u) = \frac{\lambda}{r}(1-p)[\frac{g+s}{2} - u(p) + pu'(p)]$ the expected benefit of playing risky arm 2, $c_1(p) = s - pg$ the opportunity cost of playing risky arm 1, and $c_2(p) = s - (1-p)g$ the opportunity cost of playing risky arm 2. Thus, the planner's problem is linear in both k_1 and k_2, and the planner is maximizing separately over k_1 and k_2.

If it is optimal to set $k_1 = k_2 = 0$, then the value function works out as $u(p) = s$. If it is optimal to set $k_1 = k_2 = 1$, then the Bellman equation reduces to $u(p) = \frac{\lambda}{r}\left[\frac{g+s}{2} - u(p)\right] + \frac{g}{2}$, and so $u(p) = u_{11} = \frac{1}{2}\left(g + \frac{\lambda}{\lambda+r}\, s\right)$.

If it is optimal to set $k_1 = 0$ and $k_2 = 1$, then the Bellman equation amounts to the first-order ODE

$$\lambda p(1-p)u'(p) - [r + \lambda(1-p)]u(p) = -\tfrac{1}{2}\left\{[r + \lambda(1-p)]s + (r+\lambda)(1-p)g\right\}.$$

This has the solution

$$u(p) = \tfrac{1}{2}[s + (1-p)g] + Cp^{\frac{r+\lambda}{\lambda}}(1-p)^{-\frac{r}{\lambda}},$$

where C is some constant of integration.

Finally, if it is optimal to set $k_1 = 1$ and $k_2 = 0$, then the Bellman equation is tantamount to the first-order ODE

$$\lambda p(1-p)u_1'(p) + (r + \lambda p)u_1(p) = \tfrac{1}{2}\left\{(r + \lambda p)s + (r+\lambda)pg\right\},$$

which is solved by

$$u(p) = \tfrac{1}{2}(s + pg) + C\,(1-p)^{\frac{r+\lambda}{\lambda}}p^{-\frac{r}{\lambda}}.$$

Note that whenever $k_1 = k_2$, the value function is flat as the planner does not care which arm is good. For the same reason, the problem is symmetric around $p = \frac{1}{2}$. All the planner cares about is the uncertainty that stands in the way of his realizing the upper bound on the value function, $\frac{g+s}{2}$. Hence, intuitively, the planner's value function will admit its global minimum at $p = \frac{1}{2}$, where the uncertainty is starkest.

It is clear that $(k_1, k_2) = (1, 0)$ will be optimal in a neighborhood of $p = 1$, and $(k_1, k_2) = (0, 1)$ in a neighborhood of $p = 0$. What is optimal at beliefs around $p = \frac{1}{2}$ depends on which of the two possible plateaus s and u_{11} is higher. This in turn depends on the size of the stakes involved, i.e. on the value of information as measured by the ratio $\frac{g}{s}$, and on the parameters λ and r that govern the speed of resolution of uncertainty and the planner's impatience. In fact, $s > u_{11}$ if and only if $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$. This is the case we consider first.

Proposition 3.1 *If $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$, it is optimal for the planner to use $k_1 = k_2 = 0$ on $]1-p^*, p^*[$, $(k_1, k_2) = (0, 1)$ on $[0, 1-p^*[$, and $(k_1, k_2) = (1, 0)$ on $]p^*, 1]$, where $p^* = \frac{rs}{(r+\lambda)g - \lambda s} > \frac{1}{2}$. The choice of actions at $1-p^*$ and p^* is of no consequence. The planner's value function is*

$$u(p) = \begin{cases} \frac{1}{2}\left\{s + (1-p)g + (s - p^*g)\left(\frac{p}{1-p^*}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{1-p}{p^*}\right)^{-\frac{r}{\lambda}}\right\} & \text{if} \quad p \le 1-p^*, \\ s & \text{if} \quad 1-p^* \le p \le p^*, \\ \frac{1}{2}\left\{s + pg + (s - p^*g)\left(\frac{1-p}{1-p^*}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{p}{p^*}\right)^{-\frac{r}{\lambda}}\right\} & \text{if} \quad p \ge p^*. \end{cases}$$

Figure 1 illustrates the result. Interestingly, p^* equals the cut-off belief for the single-agent problem in Keller, Rady and Cripps (2005). Thus, when the (social) value of information, as measured by $\frac{g}{s}$, is so low that the single-agent cutoff $p^* > \frac{1}{2}$, it is optimal for the planner to let the players behave as though they were single players solving two separate, completely unconnected, problems.

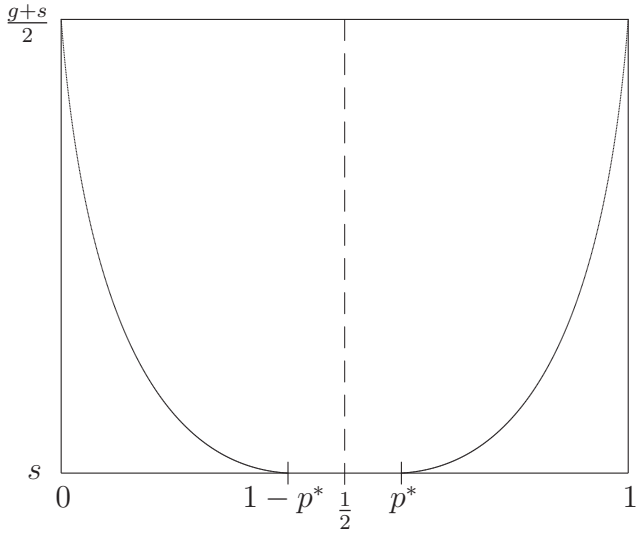

Figure 1: The planner's value function for $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$.

Conditional on there not being a breakthrough, the belief will evolve according to

$$\dot{p} = \begin{cases} \lambda p(1-p) & \text{if} \quad p < 1-p^*, \\ 0 & \text{if} \quad 1-p^* \leq p \leq p^*, \\ -\lambda p(1-p) & \text{if} \quad p > p^*. \end{cases}$$

Let us suppose risky arm 1 is good. If the initial belief $p_0 < 1-p^*$, then the posterior belief will converge to $1-p^*$ with probability 1 as there cannot be a breakthrough on risky arm 2. If $1-p^* \leq p_0 \leq p^*$, the belief will remain unchanged at p_0. If $p_0 > p^*$, the belief will converge either to 1 or to p^*. If t^* is the length of time needed for the belief to reach p^* conditional on there not being a breakthrough on risky arm 1, the probability that the belief will converge to p^* is $e^{-\lambda t^*}$. By Bayes' rule, we have $\frac{1-p_t}{p_t} = \frac{1-p_0}{p_0 e^{-\lambda t}}$ in the absence of a breakthrough, and so $e^{-\lambda t^*} = \frac{1-p_0}{p_0}\frac{p^*}{1-p^*}$. The belief will therefore converge to p^* (and learning will remain incomplete) with probability $\frac{1-p_0}{p_0}\frac{p^*}{1-p^*}$, and to 1 (and hence the truth) with the counter-probability. Analogous results hold when risky arm 2 is good.

Next, we turn to the case where $u_{11} > s$ and so playing safe on both arms cannot be part of the planner's solution.

Proposition 3.2 *If* $\frac{g}{s} > \frac{2r+\lambda}{r+\lambda}$, *it is optimal for the planner to use* $k_1 = k_2 = 1$ *on* $]\bar{p}, 1-\bar{p}[$, $(k_1, k_2) = (0,1)$ *on* $[0, \bar{p}[$, *and* $(k_1, k_2) = (1,0)$ *on* $]1-\bar{p}, 1]$, *where* $\bar{p} = \frac{(r+\lambda)s}{(r+\lambda)g+\lambda s} < \frac{1}{2}$. *The choice of actions at* $\bar{p}$ *and* $1-\bar{p}$ *is of no consequence. The planner's value function is*

$$u(p) = \begin{cases} \frac{1}{2}\left\{s + (1-p)g + \left[\bar{p}g - \frac{r}{r+\lambda}s\right]\left(\frac{p}{\bar{p}}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{1-p}{1-\bar{p}}\right)^{-\frac{r}{\lambda}}\right\} & \text{if} \quad p \leq \bar{p}, \\ \frac{1}{2}\left(g + \frac{\lambda}{r+\lambda}s\right) & \text{if} \quad \bar{p} \leq p \leq 1-\bar{p}, \\ \frac{1}{2}\left\{s + pg + \left[\bar{p}g - \frac{r}{r+\lambda}s\right]\left(\frac{1-p}{\bar{p}}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{p}{1-\bar{p}}\right)^{-\frac{r}{\lambda}}\right\} & \text{if} \quad p \geq 1-\bar{p}. \end{cases}$$

Figure 2 illustrates this result.

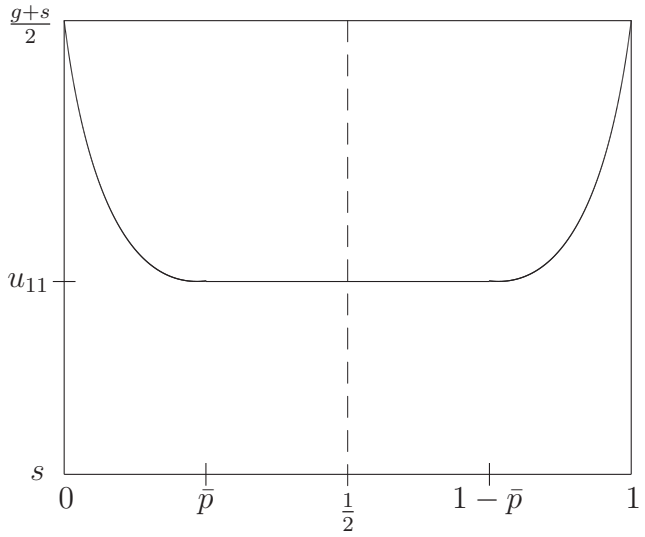

Figure 2: The planner's value function for $\frac{g}{s} > \frac{2r+\lambda}{r+\lambda}$.

The dynamics of beliefs conditional on there not being a breakthrough are now given by

$$\dot{p} = \begin{cases} \lambda p(1-p) & \text{if} \quad p < \bar{p}, \\ 0 & \text{if} \quad \bar{p} \leq p \leq 1-\bar{p}, \\ -\lambda p(1-p) & \text{if} \quad p > 1-\bar{p}. \end{cases}$$

Whenever the stakes are high, therefore, the planner shuts down *incremental* learning on $[\bar{p}, 1-\bar{p}]$. Yet he will still learn the truth with probability 1 in the long run because this interval is absorbing for the posterior belief process in the absence of a breakthrough, and he uses both risky arms forever once it is reached.

In summary, when $\frac{g}{s} > \frac{2r+\lambda}{r+\lambda}$, efficiency calls for complete learning, i.e., convergence of the posterior belief p to the truth with probability 1. When $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$, however, efficient learning can be incomplete.

Note finally that when $\frac{g}{s} = \frac{2r+\lambda}{r+\lambda}$, $\bar{p} = p^* = \frac{1}{2}$, and so both of the above results hold in this knife-edge case.

4 The Strategic Problem

Our solution concept is Markov perfect equilibrium, with the players' common posterior p as the state variable. As strategies of player $i = 1, 2$, we allow all functions $k_i : [0,1] \to \{0,1\}$ for which $k_i^{-1}(1)$ is the union of a finite number of (possibly degenerate) intervals. Given strategies k_1, k_2: $[0,1] \to \{0,1\}$, the payoff function of player i is

$$u_i(p) = \mathrm{E}\left[\int_0^\infty re^{-rt}\left\{k_i(p_t)g + [1 - k_i(p_t)]s\right\}\, dt\right]$$

if, starting from $p_0 = p$, the strategies induce a well-defined and unique law of motion for the posterior beliefs $\{p_t\}_{t>0}$; otherwise, $u_i(p) = -\infty$.[3]

Again proceeding as in Keller, Rady and Cripps (2005), we see that the following Bellman equation characterizes player 1's best responses against his opponent's strategy k_2:

$$u_1(p) = s + k_2(p)\beta_1(p, u_1) + \max_{k_1 \in \{0,1\}} k_1[b_1(p, u_1) - c_1(p)],$$

where $c_1(p) = s - pg$ is the opportunity cost player 1 has to bear when he plays risky, $b_1(p, u_1) = \frac{\lambda}{r}p[g - u_1(p) - (1-p)u_1'(p)]$ is the learning benefit player 1 accrues when he is playing risky, and $\beta_1(p, u_1) = \frac{\lambda}{r}(1-p)[s - u_1(p) + pu_1'(p)]$ is the learning benefit accruing to player 1 from player 2's playing risky.[4]

Analogously, the Bellman equation for player 2 is

$$u_2(p) = s + k_1(p)\beta_2(p, u_2) + \max_{k_2 \in \{0,1\}} k_2[b_2(p, u_2) - c_2(p)],$$

where $c_2(p) = s - (1-p)g$ is the opportunity cost player 2 has to bear when he plays risky, $b_2(p, u_2) = \frac{\lambda}{r}(1-p)[g - u_2(p) + pu_2'(p)]$ is the learning benefit player 2 accrues himself when playing risky, and $\beta_2(p, u_2) = \frac{\lambda}{r}p[s - u_2(p) - (1-p)u_2'(p)]$ is the learning benefit accruing to player 2 from player 1's playing risky.

[3]The law of motion is well-defined and unique if the differential equation $\dot{p} = -[k_1(p) - k_2(p)]\lambda p(1-p)$ possesses a unique global solution for any initial value on the unit interval. This has to be the case in equilibrium, of course.

[4]By standard results, player 1's payoff function from playing a best response against k_2 is once continuously differentiable on any open interval of beliefs where player 2's action is constant. At a belief where k_2 is discontinuous, $u_1'(p)$ must be understood as the one-sided derivative of u_1 in the direction implied by the law of motion of beliefs.

It is straightforward to obtain closed-form solutions for the payoff functions. If $k_1(p) = k_2(p) = 0$, the players' payoffs are $u_1(p) = u_2(p) = s$. If $k_1(p) = k_2(p) = 1$, the Bellman equations yield $u_1(p) = pg + \frac{\lambda}{\lambda+r}(1-p)s$ and $u_2(p) = u_1(1-p)$. On any interval where $k_1(p) = 1$ and $k_2(p) = 0$, u_1 and u_2 satisfy the ODEs

$$\begin{aligned} \lambda p(1-p)u_1'(p) + (r+\lambda p)u_1(p) &= (r+\lambda)pg, \\ \lambda p(1-p)u_2'(p) + (r+\lambda p)u_2(p) &= (r+\lambda p)s, \end{aligned}$$

which have the solutions $u_1(p) = pg + C_1(1-p)^{\frac{r+\lambda}{\lambda}}p^{-\frac{r}{\lambda}}$ and $u_2(p) = s + C_2(1-p)^{\frac{r+\lambda}{\lambda}}p^{-\frac{r}{\lambda}}$ with constants of integration C_1 and C_2, respectively. Finally, on any interval where $k_1(p) = 0$ and $k_2(p) = 1$, u_1 and u_2 solve

$$\begin{aligned} \lambda p(1-p)u_1'(p) - [r+\lambda(1-p)]u_1(p) &= -[r+\lambda(1-p)]s, \\ \lambda p(1-p)u_2'(p) - [r+\lambda(1-p)]u_2(p) &= -(r+\lambda)(1-p)g, \end{aligned}$$

which implies $u_1(p) = s + C_1 p^{\frac{r+\lambda}{\lambda}}(1-p)^{-\frac{r}{\lambda}}$ and $u_2(p) = (1-p)g + C_2 p^{\frac{r+\lambda}{\lambda}}(1-p)^{-\frac{r}{\lambda}}$.

5 Complete Learning

In this section, we shall show that whenever the planner's solution leads to complete learning, so will any MPE of the experimentation game. To this end, we first establish a lower bound on equilibrium payoffs.

From Keller, Rady and Cripps (2005), the optimal payoffs of player 1 and 2, if they were experimenting in isolation and hence applying the cutoffs p^* and $1-p^*$, respectively, would be

$$u_1^*(p) = \begin{cases} s & \text{if} \quad p \leq p^*, \\ pg + (s - p^*g)\left(\frac{1-p}{1-p^*}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{p}{p^*}\right)^{-\frac{r}{\lambda}} & \text{if} \quad p \geq p^* \end{cases}$$

and $u_2^*(p) = u_1^*(1-p)$. Since each player in the experimentation game always has the option to act as though he were a single player by just ignoring the additional signal he gets from the other player, it is quite intuitive that he cannot possibly do worse with the other player around than if he were by himself.[5] The following lemma confirms this intuition.

Lemma 5.1 *The value function of the respective single-agent problem constitutes a lower bound on each player's equilibrium value function in any Markov perfect equilibrium.*

[5]Clearly, this intuition carries over to the case where only a player's actions are observable, while his payoffs are private information. The results of this section are therefore robust to the introduction of this form of private information.

Now, if $\frac{g}{s} > \frac{2r+\lambda}{r+\lambda}$, then $p^* < \frac{1}{2} < 1 - p^*$, so at any belief p, Lemma 5.1 implies $u_1^*(p) > s$ or $u_2^*(p) > s$ or both. Thus, there cannot exist a p such that $k_1(p) = k_2(p) = 0$ be mutually best responses as this would mean $u_1(p) = u_2(p) = s$. This proves the following proposition:

Proposition 5.2 *If $\frac{g}{s} > \frac{2r+\lambda}{r+\lambda}$, learning will be complete in any Markov perfect equilibrium.*

Whenever efficiency calls for complete learning, therefore, learning will be complete in equilibrium. This complete learning result is in stark contrast to the benchmark problem of perfect positive correlation in Bolton and Harris (1999) and Keller, Rady and Cripps (2005), where there is always a range of beliefs for which learning will be incomplete. This thus confirms Dewatripont and Tirole's (1999) finding that two adversaries at loggerheads will perform better at (eventually) eliciting the truth than two partners whose interests are perfectly aligned. Indeed, provided the stakes are high enough, incomplete learning can be overcome by competition, our analysis shows.

6 Markov Perfect Equilibria

Our next aim is to the characterize the Markov perfect equilibria of the experimentation game. We treat two MPE as identical if they lead to the same law of motion of posterior beliefs and the same payoff functions.

The profile of actions (k_1, k_2) must be $(0,0)$, $(0,1)$, $(1,0)$ or $(1,1)$ at any belief. For $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$, the profile $(1,1)$ cannot occur in equilibrium since it would imply an average payoff of $u_{11} < s$ at the relevant belief, giving at least one player a payoff below s, and hence below his single-agent optimum. For $\frac{g}{s} > \frac{2r+\lambda}{r+\lambda}$, on the other hand, the profile $(0,0)$ cannot occur since it would imply incomplete learning.

We say that the *transition* (k_1^-, k_2^-)—(k_1, k_2)—(k_1^+, k_2^+) occurs at the belief $\hat{p} \in]0,1[$ if $\lim_{p\uparrow\hat{p}}(k_1(p), k_2(p)) = (k_1^-, k_2^-)$, $(k_1(\hat{p}), k_2(\hat{p})) = (k_1, k_2)$, $\lim_{p\downarrow\hat{p}}(k_1(p), k_2(p)) = (k_1^+, k_2^+)$, and at least one of the sets $\{k_1^-, k_1, k_1^+\}$ and $\{k_2^-, k_2, k_2^+\}$ contains more than one element. Given our definition of strategies, each MPE has a finite number of transitions.

We first note that the transitions $(0,0)$—$(1,0)$—$(0,0)$ and $(0,0)$—$(0,1)$—$(0,0)$ can be ignored since the law of motion of beliefs and players' payoffs would be the same if both played safe at $\hat{p}$ and so no action changed there at all. Similarly, we can ignore the transitions $(1,1)$—$(1,0)$—$(1,1)$ and $(1,1)$—$(0,1)$—$(1,1)$ since the law of motion of beliefs and players' payoffs would be the same if both played risky at $\hat{p}$.

We further note that there can be no transitions $(0,1)$—$(1,0)$—(k_1^+, k_2^+) in equilibrium. In fact, the law of motion conditional on there being no breakthrough would read $\dot{p} = \lambda p(1-p)$ on $[\hat{p} - \epsilon, \hat{p}[$, and $\dot{p} = -\lambda p(1-p)$ at $\hat{p}$. Such a change in sign precludes the existence of a solution with initial value $\hat{p}$, and so the law of motion of our state variable would not be well-defined.[6] The same argument rules out transitions (k_1^-, k_2^-)—$(0,1)$—$(1,0)$.

We call all transitions that cannot be ignored or have not been ruled out so far *admissible*. Among these, we first consider transitions where one player's action does not change. To this end, we note that the cutoff belief above which a myopic player 1 (i.e., a player who is only interested in maximizing current payoffs) would play risky is $p^m = \frac{s}{g}$. A myopic player 2 would play risky at beliefs below the cutoff $1 - p^m$. Invoking the standard principles of value matching and smooth pasting, we obtain the following result.

Lemma 6.1 *The following statements hold for all admissible transitions in any Markov perfect equilibrium:* $(k_1^-, 0)$—$(k_1, 0)$—$(k_1^+, 0)$ *can only occur at the belief* p^*, $(0, k_2^-)$—$(0, k_2)$—$(0, k_2^+)$ *only at* $1 - p^*$, $(k_1^-, 1)$—$(k_1, 1)$—$(k_1^+, 1)$ *only at* p^m, *and* $(1, k_2^-)$—$(1, k_2)$—$(1, k_2^+)$ *only at* $1 - p^m$.

While it is intuitive that a player would apply the single-agent cutoff rule against an opponent who plays safe and thus provides no information, it is surprising that the myopic cutoff determines equilibrium behavior against an opponent who plays risky. Technically, this result is due to the fact that along player 1's payoff function for $k_1 = k_2 = 1$, his learning benefit from playing risky vanishes:

$$b_1(p, u_1) = \frac{\lambda}{r} p \left[g - \left(pg + \frac{\lambda}{\lambda + r}(1-p)s \right) - (1-p) \left(g - \frac{\lambda}{\lambda + r} s \right) \right] = 0,$$

and so $k_1 = 1$ is optimal against $k_2 = 1$ if and only if $c_1(p) \leq 0$, that is, $p \geq p^m$. This is best understood by recalling the law of motion of beliefs in the absence of a success on either arm, $\dot{p} = -(k_1 - k_2)\lambda p(1-p)$, which tells us that if both players are playing risky, the state variable does not budge. In other words, all a player does by chiming in in his opponent's experimentation is to keep the belief, his action and his continuation value constant until the first success occurs. But this can only be optimal if he reaps maximal current payoffs while he waits for the resolution of uncertainty. So his playing the risky arm must be myopically optimal.

In the following lemma, we pin down the conditions under which some of the remaining admissible transitions may occur in equilibrium.

[6] Similar problems have already been treated in the decision-theoretic literature. To guarantee a well-defined law of motion of posterior beliefs, Presman (1990) allows for simultaneous use of both arms, i.e. for experimentation intensities $k_t \in [0,1]$.

Lemma 6.2 *The following statements hold for all Markov perfect equilibria. (i) The transition* $(0,1)$—$(0,0)$—$(1,0)$ *can only occur if* $\frac{g}{s} = \frac{2r+\lambda}{r+\lambda}$ *and only at belief* $\frac{1}{2}$*. (ii) The transition* $(1,0)$—$(0,0)$—$(0,1)$ *can only occur if* $\frac{g}{s} \leq \frac{2r+\lambda}{r+\lambda}$ *and only at beliefs in* $[1-p^*, p^*]$*. (iii) The transition* $(0,1)$—$(1,1)$—$(1,0)$ *can only occur if* $\frac{2r+\lambda}{r+\lambda} \leq \frac{g}{s} \leq 2$ *and only at beliefs in* $[1-p^m, p^m]$*. (iv) The transition* $(1,0)$—$(1,1)$—$(0,1)$ *can only occur if* $\frac{g}{s} \geq 2$ *and only at beliefs in* $[p^m, 1-p^m]$*.*

The only admissible transitions that we have not addressed yet are those of the form $(1,0)$—$(0,1)$—(k_1^+, k_2^+) and (k_1^-, k_2^-)—$(1,0)$—$(0,1)$. We will see later that they cannot occur in equilibrium.

The structure of Markov perfect equilibria depends on the relative position of the possible transition points, which in turn depends on the stakes involved, i.e. on the ratio $\frac{g}{s}$. For expositional reasons, we shall first analyze the case of very low and that of very high stakes.

6.1 Low Stakes

The low-stakes case is defined by the inequality $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$. In this case, $1-p^m < 1-p^* < \frac{1}{2} < p^* < p^m$.

Proposition 6.3 *When* $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$*, the unique Markov perfect equilibrium coincides with the planner's solution. That is, player 1 plays risky at beliefs above* p^**, and safe below* p^**, while player 2 plays risky at beliefs below* $1-p^*$*, and safe above* $1-p^*$*. Player 1's behavior at* p^* *and player 2's behavior at* $1-p^*$ *are of no consequence. The pertaining value functions are those of the respective single-agent problems,* u_1^* *and* u_2^**.*

Figure 3 illustrates this result.[7] The players' average payoff function coincides with the planner's value function as stated in Proposition 3.1 and shown in Figure 1.

Why we should have efficiency in this case is intuitively quite clear, as the planner basically lets players behave as though they were single players. As $p^* > \frac{1}{2}$, there is no spillover from a player behaving like a single agent on the other player's optimization problem. Hence the latter's best response calls for behaving like a single player as well. Thus, there is no conflict between social and private incentives.

The law of motion for the belief and the probability of the players' eventually finding out the true state of the world are thus the same as in the planner's solution for low stakes.

[7] In this and all subsequent figures, the thick solid line depicts the value function of player 1, the thin solid line that of player 2, and the dotted line the players' average payoff function.

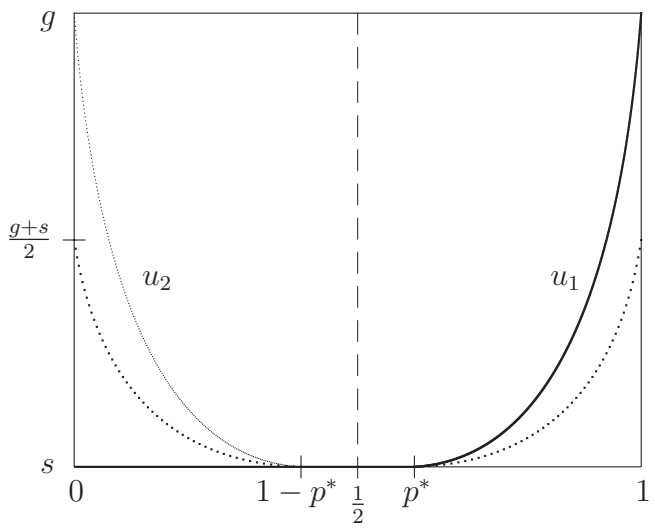

Figure 3: The equilibrium value functions for $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$.

6.2 High Stakes

The high-stakes case is defined by the inequality $\frac{g}{s} > 2$. In this case, $p^* < p^m < \frac{1}{2} < 1-p^m < 1-p^*$.

Proposition 6.4 *When $\frac{g}{s} > 2$, the game has a unique Markov perfect equilibrium, in which both players behave myopically. That is, player 1 plays risky at beliefs above p^m, and safe below p^m, while player 2 plays risky at beliefs below $1-p^m$, and safe above $1-p^m$. Player 1's behavior at p^m and player 2's behavior at $1-p^m$ are of no consequence. The pertaining value functions are*

$$u_1(p) = \begin{cases} s + \frac{\lambda}{\lambda+r}(1-p^m)s\left(\frac{p}{p^m}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{1-p}{1-p^m}\right)^{-\frac{r}{\lambda}} & \text{if } \ p \le p^m, \\ pg + \frac{\lambda}{\lambda+r}(1-p)s & \text{if } \ p^m \le p \le 1-p^m, \\ pg + \frac{\lambda}{\lambda+r}p^m s\left(\frac{1-p}{p^m}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{p}{1-p^m}\right)^{-\frac{r}{\lambda}} & \text{if } \ p \ge 1-p^m \end{cases}$$

and $u_2(p) = u_1(1-p)$.

Thus, the unique equilibrium calls for both players' behaving myopically. This is best understood by recalling from our discussion above that individual optimality calls for myopic behavior whenever one's opponent is playing risky. When the stakes are high, players' myopic cutoff beliefs are more pessimistic than $p = \frac{1}{2}$, so the relevant intervals overlap.

Figure 4 illustrates this result. Player 1's value function has a kink at $1-p^m$, where player 2 changes action. Symmetrically, player 2's value function has a kink at p^m, where

player 1 changes action. As a consequence, the average payoff function has a kink both at p^m and at $1-p^m$. That it dips below the level u_{11} close to these kinks is evidence of the inefficiency of equilibrium. We will return to this point in Section 6.4 below.

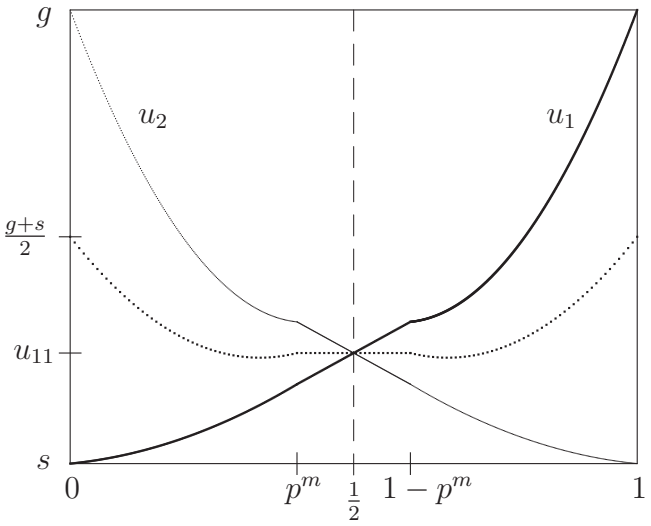

Figure 4: The equilibrium value functions for $\frac{g}{s} > 2$.

Arguing exactly as after Proposition 3.2, it is straightforward to see that learning will be complete, as predicted by Proposition 5.2.

6.3 Intermediate Stakes

This case is defined by the condition that $\frac{2r+\lambda}{r+\lambda} < \frac{g}{s} < 2$. In this case, $p^* < \frac{1}{2} < p^m$.

When the stakes are intermediate in size, equilibrium is not unique; rather there is a continuum of equilibria, as the following proposition shows.

Proposition 6.5 *When $\frac{2r+\lambda}{r+\lambda} < \frac{g}{s} < 2$, there is a continuum of Markov perfect equilibria. Each of them is characterized by a unique belief $\hat{p} \in [\max\{1-p^m, p^*\}, \min\{p^m, 1-p^*\}]$ such that player 1 plays risky at all beliefs $p \geq \hat{p}$, and player 2 at all beliefs $p \leq \hat{p}$. The pertaining value functions are given by*

$$u_1(p) = \begin{cases} s + \left[\hat{p}g + \frac{\lambda}{\lambda+r}(1-\hat{p})s - s\right]\left(\frac{p}{\hat{p}}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{1-p}{1-\hat{p}}\right)^{-\frac{r}{\lambda}} & \text{if } p \leq \hat{p} \\ pg + \frac{\lambda}{\lambda+r}(1-\hat{p})s\left(\frac{1-p}{1-\hat{p}}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{p}{\hat{p}}\right)^{-\frac{r}{\lambda}} & \text{if } p \geq \hat{p} \end{cases}$$

for player 1, and

$$u_2(p) = \begin{cases} (1-p)g + \frac{\lambda}{\lambda+r}\hat{p}s\left(\frac{p}{\hat{p}}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{1-p}{1-\hat{p}}\right)^{-\frac{r}{\lambda}} & \text{if } p \le \hat{p} \\ s + \left[(1-\hat{p})g + \frac{\lambda}{\lambda+r}\hat{p}s - s\right]\left(\frac{1-p}{1-\hat{p}}\right)^{\frac{r+\lambda}{\lambda}}\left(\frac{p}{\hat{p}}\right)^{-\frac{r}{\lambda}} & \text{if } p \ge \hat{p} \end{cases}$$

for player 2.

Amongst the continuum of equilibria characterized in Proposition 6.5, there is a unique symmetric one, given by $\hat{p} = \frac{1}{2}$. Figure 5 illustrates this equilibrium. Both players' value functions and their average are kinked at $p = \frac{1}{2}$, where both players change action. At any belief except $p = \frac{1}{2}$, the average payoff function is below the planner's solution; if the initial belief is $p_0 = \frac{1}{2}$, however, the efficient average payoff u_{11} is achieved.

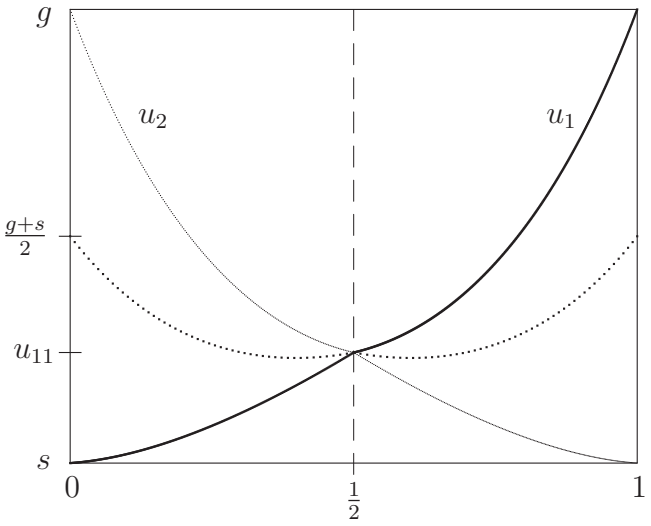

Figure 5: The value functions in the unique symmetric equilibrium for $\frac{2r+\lambda}{r+\lambda} < \frac{g}{s} < 2$.

For arbitrary $\hat{p}$, the dynamics of beliefs in the absence of a breakthrough are given by

$$\dot{p} = \begin{cases} \lambda p(1-p) & \text{if } \ p < \hat{p}, \\ 0 & \text{if } \ p = \hat{p}, \\ -\lambda p(1-p) & \text{if } \ p > \hat{p}. \end{cases}$$

As predicted by Proposition 5.2, learning is complete in all these equilibria.

6.4 Efficiency vs. Myopia

As we have pointed out already, when the stakes are low, players do not interfere with each other's optimization problem and behave as though they were all by themselves. We have

seen that this kind of behavior is also efficient.

If stakes are high, however, we have seen that players behave myopically. This implies that in the unique MPE, experimentation is at efficient levels except on $[\bar{p}, p^m] \cup [1-p^m, 1-\bar{p}]$, the union of two non-empty and non-degenerate intervals, where experimentation is inefficiently low. Put differently, there is a region of beliefs where one player free-rides on the other player's experimentation, which is inefficient from a social point of view.

In the case of intermediate stakes, equilibrium behavior changes gradually from efficiency to myopia. Indeed, as is easily verified, if $\frac{2r+\lambda}{r+\lambda} < \frac{g}{s} \leq \frac{2r+\lambda}{2(r+\lambda)} + \sqrt{\frac{(2r+\lambda)^2}{4(r+\lambda)^2} + \frac{\lambda}{r+\lambda}}$, then the lower bound on the equilibrium cutoff $\hat{p}$ satisfies $\max\{p^*, 1-p^m\} \leq \bar{p}$. Now, if the players' initial belief is $p_0 > \bar{p}$, the equilibrium with $\hat{p} = \bar{p}$ achieves efficiency as the only beliefs that are reached with positive probability under the equilibrium strategies are given by the set $\{0,1\} \cup [p_0, \hat{p}]$, and the equilibrium strategies prescribe the efficient actions at all of these beliefs. Similarly, for $p_0 < 1-\bar{p}$, efficiency is achieved by the equilibrium with $\hat{p} = 1-\bar{p}$. Finally, if $\bar{p} \leq p_0 \leq 1-\bar{p}$, efficiency is achieved by the equilibrium with $\hat{p} = p_0$, since this ensures that only beliefs in $\{0, p_0, 1\}$ are reached with positive probability.

If $\frac{2r+\lambda}{2(r+\lambda)} + \sqrt{\frac{(2r+\lambda)^2}{4(r+\lambda)^2} + \frac{\lambda}{r+\lambda}} < \frac{g}{s} < 2$, then $p^* < \bar{p} < 1-p^m$. Now, suppose $\bar{p} \leq p_0 < 1-p^m$. Equilibrium uniquely calls for $(k_1, k_2)(p) = (0,1)$ for all $p \leq 1-p^m$, whereas efficiency would require $(k_1, k_2)(p) = (1,1)$ whenever $\bar{p} < p \leq 1-\bar{p}$. Thus, equilibrium implies inefficient play on the interval $]\bar{p}, 1-p^m[$ which is reached with positive probability given the initial belief p_0.

Combined with our results for low and high stakes, these arguments establish the following proposition.

Proposition 6.6 *If $\frac{g}{s} \leq \frac{2r+\lambda}{2(r+\lambda)} + \sqrt{\frac{(2r+\lambda)^2}{4(r+\lambda)^2} + \frac{\lambda}{r+\lambda}}$, then for each initial belief, there exists a Markov perfect equilibrium that achieves the efficient outcome. If $\frac{g}{s} > \frac{2r+\lambda}{2(r+\lambda)} + \sqrt{\frac{(2r+\lambda)^2}{4(r+\lambda)^2} + \frac{\lambda}{r+\lambda}}$, there are initial beliefs under which the efficient outcome cannot be reached in equilibrium.*

If $1 + \sqrt{\frac{r}{r+\lambda}} \leq \frac{g}{s} < 2$, then $p^m \leq 1-p^*$. In this situation, setting $\hat{p} = p^m$ ($\hat{p} = 1-p^m$) yields an equilibrium where only player 1 (player 2) behaves myopically, while the other player bears the entire burden of experimentation by himself, something he is only willing to do provided the stakes involved exceed the threshold of $1 + \sqrt{\frac{r}{r+\lambda}}$. In view of our findings for low and high stakes, this establishes the following result.

Proposition 6.7 *If $\frac{g}{s} \geq 1 + \sqrt{\frac{r}{r+\lambda}}$, there exists a Markov perfect equilibrium where at least*

one of the players behaves myopically. If $\frac{g}{s} < 1 + \sqrt{\frac{r}{r+\lambda}}$*, no player behaves myopically in equilibrium.*

Note that for certain parameter values, namely if $1+\sqrt{\frac{r}{r+\lambda}} \leq \frac{g}{s} \leq \frac{2r+\lambda}{2(r+\lambda)}+\sqrt{\frac{(2r+\lambda)^2}{4(r+\lambda)^2}+\frac{\lambda}{r+\lambda}}$, equilibria where one player behaves myopically co-exist with equilibria that achieve efficiency given the initial belief.

7 Conclusion

We have analyzed a game of strategic experimentation in continuous time where players' interests are diametrically opposed. We have found that, in very sharp contrast to the case where players' interests are perfectly aligned, all the equilibria are of the cutoff type, and that for a large subset of parameters, equilibrium is unique. When the stakes are low, equilibrium behavior is efficient, whereas for high stakes players behave myopically.

In our analysis, we have restricted ourselves to what in the literature has been termed "pure strategy equilibria" (by Bolton and Harris, 1999, and 2000) or "'simple equilibria" (by Keller, Rady and Cripps, 2005, and Keller and Rady, 2007). Our results on efficiency, as well as our complete learning result, are robust to an extension of the strategy space where players are allowed to choose experimentation intensities from the entire unit interval.

Our finding that incomplete learning, which hitherto has been a staple result of the strategic two-armed bandit literature, can be overcome by competition, may be interesting as a building block for more applied models with a richer structure. For instance, it may constitute a microfoundation for the empirical fact that democracy and decentralization will foster investments in risky R&D and innovation. For example, Aghion, Alesina, and Trebbi (2007) were able to show empirically that democracy and political rights enhance the growth of technologically more advanced sectors, which rely more heavily on innovation and R&D. They showed to boot that an important channel by which democracy spurs the growth of the more advanced sectors is freedom of entry, which obviously encourages competition.

Appendix

Proof of Proposition 3.1

The policy (k_1, k_2) implies a well-defined law of motion for the posterior belief. The function u satisfies value matching and smooth pasting at p^* and $1-p^*$, hence is of class C^1. It is strictly decreasing on $[0, 1-p^*]$ and strictly increasing on $[p^*, 1]$. Moreover, $u = s + B_2 - \frac{c_2}{2}$ on $[0, 1-p^*]$, $u = s$ on $[1-p^*, p^*]$, and $u = s + B_1 - \frac{c_1}{2}$ on $[p^*, 1]$ (we drop the arguments for simplicity), which shows that u is indeed the planner's payoff function from (k_1, k_2).

To show that u and this policy (k_1, k_2) solve the planner's Bellman equation, and hence that (k_1, k_2) is optimal, it is enough to establish that $B_1 < \frac{c_1}{2}$ and $B_2 > \frac{c_2}{2}$ on $]0, 1-p^*[$, $B_1 < \frac{c_1}{2}$ and $B_2 < \frac{c_2}{2}$ on $]1-p^*, p^*[$, and $B_1 > \frac{c_1}{2}$ and $B_2 < \frac{c_2}{2}$ on $]p^*, 1[$. Consider this last interval. There, $u = s + B_1 - \frac{c_1}{2}$ and $u > s$ (by monotonicity of u) immediately imply $B_1 > \frac{c_1}{2}$. Next, $B_2 = \frac{\lambda}{r}[\frac{g+s}{2} - u] - B_1 = \frac{\lambda}{r}[\frac{g+s}{2} - u] - u + s - \frac{c_1}{2}$; this is smaller than $\frac{c_2}{2}$ if and only if $u > u_{11}$, which holds here since $u > s$ and $s > u_{11}$. The other two intervals are treated in a similar way. ∎

Proof of Proposition 3.2

The proof proceeds along the same lines as the previous one and is therefore omitted. ∎

For $p \in [0, 1]$, we now define

$$w_1(p) = pg + \frac{\lambda}{\lambda + r}(1-p)s \quad \text{and} \quad w_2(p) = w_1(1-p).$$

Furthermore, we define the players' expected full-information payoffs:

$$\overline{u}_1(p) = pg + (1-p)s \quad \text{and} \quad \overline{u}_2(p) = (1-p)g + ps.$$

We then get the following lemma:

Lemma A.1 *$k_1(p) = 1$ is a best response to $k_2(p) = 1$ on some non-degenerate interval of beliefs if and only if $u_1(p) \leq w_1(p)$ on that interval. Similarly, $k_2(p) = 1$ is a best response to $k_1(p) = 1$ on some non-degenerate interval if and only if $u_2(p) \geq w_2(p)$ on that interval.*

Proof: We first note that $b_1(p, u_1) = \frac{\lambda}{r}[\overline{u}_1(p) - u_1(p)] - \beta_1(p, u_1)$, and, analogously, that $b_2(p, u_2) = \frac{\lambda}{r}[\overline{u}_2(p) - u_2(p)] - \beta_2(p, u_2)$. We also note that a necessary and sufficient condition for $k_1(p) = 0$ to be a best response to $k_2(p) = 1$ on some interval of beliefs is that $u_1(p) = s + \beta_1(p, u_1)$ and $b_1(p, u_1) \leq c_1(p)$, which in turn requires that $c_1(p) \geq \frac{\lambda}{r}[\overline{u}_1(p) - u_1(p)] - [u_1(p) - s]$, which is the same as $u_1(p) \geq w_1(p)$. Analogously, a necessary and sufficient condition for $k_2(p) = 0$ to be a best response to $k_1(p) = 1$ on some interval of beliefs is that $u_2(p) \geq w_2(p)$. ∎

Proof of Lemma 5.1

Let u_1 be player 1's equilibrium value function in some MPE with equilibrium strategies (k_1, k_2). Write $b_1^*(p) = b_1(p, u_1^*)$, and $\beta_1^*(p) = \beta_1(p, u_1^*)$. Henceforth, we shall suppress arguments whenever

this is convenient. Since p^* is the single-agent cutoff belief for player 1, we have $u_1^* = s$ for $p \leq p^*$ and $u_1^* = s + b_1^* - c_1 = pg + b_1^*$ for $p > p^*$. Thus, if $p \leq p^*$, the claim obviously holds as s is a lower bound on u_1.

Now, let $p > p^*$. Then, noting that $b_1^* = u_1^* - pg$, we have $\beta_1^* = \frac{\lambda}{r}[\overline{u}_1 - u_1^*] - (u_1^* - gp)$. Thus, $\beta_1^* > 0$ if and only if $u_1^* < pg + \frac{\lambda}{\lambda+r}(1-p)s = w_1$. Noting that $w_1(p^*) = u_1^*(p^*) = s$, $w_1(1) = u_1^*(1) = g$, and that w_1 is linear whereas u_1^* is strictly convex in p, we conclude that $u_1^* < w_1$ and hence $\beta_1^* > 0$ on $]p^*, 1[$. As a consequence, we have $u_1^* = pg + b_1^* \leq gp + k_2\beta_1^* + b_1^*$ on $[p^*, 1]$.

Now, suppose $u_1 < u_1^*$ at some belief. Since s is a lower bound on u_1, this implies existence of a belief strictly greater than p^* where $u_1 < u_1^*$ and $u_1' \leq (u_1^*)'$. This immediately yields $b_1 > b_1^* > c_1$, so that we must have $k_1 = 1$ and $u_1 = pg + k_2\beta_1 + b_1$ at the belief in question. But now,

$$u_1 - u_1^* \geq pg + k_2\beta_1 + b_1 - (pg + k_2\beta_1^* + b_1^*) = (1-k_2)(b_1 - b_1^*) + k_2\left[\frac{\lambda}{r}(u_1^* - u_1)\right] > 0,$$

a contradiction.

An analogous argument applies for player 2's equilibrium value function u_2. ∎

Proof of Lemma 6.1

At each of these admissible transitions, we must have value matching and smooth pasting for the player who changes his action. For example, suppose that there is a transition $(0,0)$—$(0,0)$—$(1,0)$ at the belief $\hat{p}$. Then the value function of player 1 must satisfy $u_1(\hat{p}) = s$, $u_1'(\hat{p}) = 0$ and $\lambda\hat{p}(1-\hat{p})u_1'(\hat{p}) + (r + \lambda\hat{p})u_1(\hat{p}) = (r+\lambda)\hat{p}g$ by the ODE for $(k_1, k_2) = (1,0)$. Substituting for $u_1(\hat{p})$ and $u_1'(\hat{p})$ and solving yields $\hat{p} = \frac{rs}{(r+\lambda)g - \lambda s} = p^*$. The other transitions are dealt with in the same way. ∎

Proof of Lemma 6.2

Suppose the transition $(0,1)$—$(0,0)$—$(1,0)$ occurs at belief $\hat{p}$. This implies $u_1(\hat{p}) = u_2(\hat{p}) = s$. Moreover, to the left of $\hat{p}$, player 2's value function solves the ODE for $k_1 = 0$ and $k_2 = 1$, which, by continuity of u_2, implies $\lambda\hat{p}(1-\hat{p})u_2'(\hat{p}-) = [r + \lambda(1-\hat{p})]s - (r+\lambda)(1-\hat{p})g$, where $u_2'(\hat{p}-) = \lim_{p\uparrow\hat{p}} u_2'(p)$. Now, if $\hat{p} > 1 - p^*$, then $u_2'(\hat{p}-) > 0$ and so $u_2(p) < s$ immediately to the left of $\hat{p}$ – a contradiction. So we must have $\hat{p} \leq 1 - p^*$. To the right of $\hat{p}$, player 1's value function solves the ODE for $k_1 = 1$ and $k_2 = 0$, which implies $\lambda\hat{p}(1-\hat{p})u_1'(\hat{p}+) = (r+\lambda)\hat{p}g - (r + \lambda\hat{p})s$, where $u_1'(\hat{p}+) = \lim_{p\downarrow\hat{p}} u_1'(p)$. If $\hat{p} < p^*$, then $u_1'(\hat{p}+) < 0$ and $u_1(p) < s$ immediately to the right of $\hat{p}$ – a contradiction again. So we must have $\hat{p} \in [p^*, 1-p^*]$. But $p^* \leq 1 - p^*$ if and only if $\frac{g}{s} \geq \frac{2r+\lambda}{r+\lambda}$. By Lemma 5.1, however, the existence of a belief $\hat{p}$ such that $u_1(\hat{p}) = u_2(\hat{p}) = s$ requires $\frac{g}{s} \leq \frac{2r+\lambda}{r+\lambda}$. Thus, we must have $\frac{g}{s} = \frac{2r+\lambda}{r+\lambda}$ and so $p^* = 1 - p^* = \frac{1}{2}$. This proves statement (i).

Next, suppose the transition $(1,0)$—$(0,0)$—$(0,1)$ occurs at belief $\hat{p}$. This implies $u_1(\hat{p}) = u_2(\hat{p}) = s$. Now, player 2's value function solves the ODE for $k_1 = 0$ and $k_2 = 1$ to the right of $\hat{p}$, and so we find $u_2'(\hat{p}+) < 0$ whenever $\hat{p} < 1 - p^*$. Player 1's value function solves the ODE for $k_1 = 1$

and $k_2 = 0$ to the left of $\hat{p}$, and so $u_1'(\hat{p}-) > 0$ whenever $\hat{p} > p^*$. So we must have $\hat{p} \in [p^*, 1-p^*]$, which requires $\frac{g}{s} \leq \frac{2r+\lambda}{r+\lambda}$. This proves statement (ii).

Suppose now that the transition $(0,1)$—$(1,1)$—$(1,0)$ occurs at belief $\hat{p}$. This implies $u_1(\hat{p}) = w_1(\hat{p})$ and $u_2(\hat{p}) = w_2(\hat{p})$. To the right of $\hat{p}$, player 2' value function solves the ODE for $k_1 = 1$ and $k_2 = 0$, which implies

$$u_2'(\hat{p}+) = \frac{r+\lambda\hat{p}}{\lambda\hat{p}(1-\hat{p})}\left[\frac{r+\lambda(1-\hat{p})}{r+\lambda}s - (1-\hat{p})g\right].$$

Now, if $\hat{p} < 1-p^m$, then $u_2'(\hat{p}+) < w_2'(\hat{p})$ and so $u_2 < w_2$ to the immediate right of $\hat{p}$, implying by Lemma A.1 that $k_2 = 0$ is *not* a best response to $k_1 = 1$ there – a contradiction. Thus, we must have $\hat{p} \geq 1-p^m$. To the left of $\hat{p}$, player 1's value function solves the ODE for $k_1 = 0$ and $k_2 = 1$, which implies

$$u_1'(\hat{p}-) = \frac{r+\lambda(1-\hat{p})}{\lambda\hat{p}(1-\hat{p})}\left[\hat{p}g - \frac{r+\lambda\hat{p}}{r+\lambda}s\right].$$

If $\hat{p} > p^m$, then $u_1'(\hat{p}-) > w_1'(\hat{p})$ and so $u_1 < w_1$ to the immediate left of $\hat{p}$ – another contradiction to Lemma A.1. So we must have $\hat{p} \in [1-p^m, p^m]$, which requires $\frac{g}{s} \leq 2$. Furthermore, we note that the existence of a belief $\hat{p}$ such that $k_1(\hat{p}) = k_2(\hat{p}) = 1$ requires $u_{11} \geq s$ and hence $\frac{g}{s} \geq \frac{2r+\lambda}{r+\lambda}$. This proves statement (iii).

Finally, suppose the transition $(1,0)$—$(1,1)$—$(0,1)$ occurs at belief $\hat{p}$. Again, this implies $u_1(\hat{p}) = w_1(\hat{p})$ and $u_2(\hat{p}) = w_2(\hat{p})$. Now, player 2's value function solves the ODE for $k_1 = 1$ and $k_2 = 0$ to the left of $\hat{p}$, and so we find $u_2'(\hat{p}-) > w_2'(\hat{p})$ whenever $\hat{p} > 1-p^m$. Player 1's value function solves the ODE for $k_1 = 0$ and $k_2 = 1$ to the right of $\hat{p}$, and so $u_1'(\hat{p}+) < w_1'(\hat{p})$ whenever $\hat{p} < p^m$. Thus we must have $\hat{p} \in [p^m, 1-p^m]$, which requires $\frac{g}{s} \geq 2$. This proves statement (iv). ∎

Proof of Proposition 6.3

The policies k_1 and k_2 induce a well-defined law of motion for the posterior belief. The functions u_1 and u_2 are of class C^1 with u_2 strictly decreasing on $[0, 1-p^*]$ and u_1 strictly increasing on $[p^*, 1]$. As $u_2 = s + b_2 - c_2$ on $[0, 1-p^*]$ and $u_1 = s + b_1 - c_1$ on $[p^*, 1]$ (we drop the arguments for simplicity), u_1 and u_2 are indeed the players' payoff functions for (k_1, k_2).

To show that u_1 and the policy k_1 solve player 1's Bellman equation given player 2's strategy k_2, and hence that k_1 is a best response to k_2, it is enough to establish that $b_1 < c_1$ on $]0, p^*[$ and $b_1 > c_1$ on $]p^*, 1[$. On this last interval, $u = s + b_1 - c_1$ and $u_1 > s$ (by monotonicity of u_1) immediately imply $b_1 > c_1$. On $]0, p^*[$, we have $u_1 = s$ and $u_1' = 0$, hence $b_1 - c_1 = \frac{\lambda}{r}p(g-s) - (s - pg) = \frac{(r+\lambda)g - \lambda s}{r}p - s < 0$. As $u_2(p) = u_1(1-p)$ and $k_2(p) = k_1(1-p)$, the previous steps also imply $b_2 > c_2$ on $]0, 1-p^*[$ and $b_2 < c_2$ on $]1-p^*, 1[$, which completes the proof that (k_1, k_2) constitutes an equilibrium.

For uniqueness, we note that, as $g > s$, $k_1(1) = k_2(0) = 1$ and $k_1(0) = k_2(1) = 0$ in any MPE. Recall that as $u_{11} < s$, the action profile $(k_1, k_2) = (1,1)$ cannot be part of an MPE since this would involve a payoff strictly below s for at least one player at some belief. Of the transitions considered in Lemma 6.2, only $(1,0)$—$(0,0)$—$(0,1)$ could happen in this case, and it could only occur at some

belief $\hat{p} \in [1-p^*, p^*]$. Moreover, besides the transitions considered in Lemma 6.1, only the following transitions could potentially arise here: $(1,0)$—$(0,1)$—$(0,0)$, $(1,0)$—$(0,1)$—$(0,1)$, $(1,0)$—$(1,0)$—$(0,1)$, $(0,0)$—$(1,0)$—$(0,1)$. It thus follows that in any MPE, players can only transition out of $(0,1) = (k_1(0), k_2(0))$ at belief $1-p^*$, and have to move into $(0,0)$ to the immediate right of $1-p^*$. As $(k_1(1), k_2(1)) = (1,0)$, players cannot transition back into $(0,1)$ to the right of $1-p^*$. Since the only way for the players to transition from $(0,0)$ into $(1,0)$ would be via smooth pasting at p^*, $(1,0)$—$(0,1)$—$(0,0)$ could only happen to the right of p^*, implying $k_1(1) = 0$ – a contradiction. Since $(1,0)$—$(0,0)$—$(1,0)$ could potentially only happen at p^*, and $(0,1)$—$(0,0)$—$(0,1)$ only at $1-p^*$, these two transitions cannot occur either. ∎

Proof of Proposition 6.4

The policies k_1 and k_2 induce a well-defined law of motion for the posterior belief. The functions u_1 and u_2 are of class C^1 except at $1-p^m$ and p^m, respectively, where their first derivative jumps downward; u_1 is strictly increasing, u_2 strictly decreasing. Moreover, $u_1 = s+\beta_1$ and $u_2 = s+b_2-c_2$ on $[0, p^m]$, $u_1 = s+\beta_1+b_1-c_1$ and $u_2 = s+\beta_2+b_2-c_2$ on $[p^m, 1-p^m]$, and $u_1 = s+b_1-c_1$ and $u_2 = s+\beta_2$ on $[1-p^m, 1]$. So u_1 and u_2 are indeed the players' payoff functions for (k_1, k_2).

To show that u_1 and the policy k_1 solve player 1's Bellman equation given player 2's strategy k_2, and hence that k_1 is a best response to k_2, it is enough to establish that $b_1 < c_1$ on $]0, p^m[$ and $b_1 > c_1$ on $]p^m, 1[$. On $]1-p^m, 1[$, $u_1 = s+b_1-c_1$ and $u_1 > s$ (by monotonicity of u_1) immediately imply $b_1 > c_1$. On $]p^m, 1-p^m[$, we have $b_1 = 0 > c_1$. On $]0, p^m[$, $u_1 = s+\beta_1$ and $b_1+\beta_1 = \frac{\lambda}{r}[pg+(1-p)s-u_1]$ imply $b_1 - c_1 = \frac{\lambda}{r}[pg+(1-p)s] - (1+\frac{\lambda}{r})u_1 + pg$. This is strictly smaller than 0 if and only if $u_1 > w_1$, which is easily verified for the interval under consideration. As $u_2(p) = u_1(1-p)$ and $k_2(p) = k_1(1-p)$, the previous steps also imply $b_2 > c_2$ on $]0, 1-p^m[$ and $b_2 < c_2$ on $]1-p^m, 1[$, which completes the proof that (k_1, k_2) constitutes an equilibrium.

For uniqueness, we note that, as $g > s$, $k_1(1) = k_2(0) = 1$ and $k_1(0) = k_2(1) = 0$ in any MPE. Recall that the action profile $(k_1, k_2) = (0,0)$ cannot be part of an MPE since it would imply incomplete learning. It thus follows immediately from Lemmas 6.1 and 6.2 that the only way for players to transition out of $(0,1)$ is for them to switch to $(1,1)$ at p^m. Thus, players cannot transition back into $(0,1)$ to the right of p^m. Therefore, again using Lemma 6.1, the only way to transition out of $(1,1)$ is to switch to $(1,0)$ at $1-p^m$. Hence, players cannot transition back to $(1,1)$ or $(0,1)$ to the right of $1-p^m$. As, by Lemma 6.1, the transitions $(1,0)$—$(1,1)$—$(1,0)$ and $(1,1)$—$(1,0)$—$(1,1)$ could potentially only happen at $1-p^m$, and $(0,1)$—$(1,1)$—$(0,1)$ and $(1,1)$—$(0,1)$—$(1,1)$ only at p^m, these transitions cannot occur either. ∎

Proof of Proposition 6.5

The policies k_1 and k_2 induce a well-defined law of motion for the posterior belief with an absorbing state at $\hat{p}$. The functions u_1 and u_2 are of class C^1 except at $\hat{p}$, where their first derivatives jump; u_1 is strictly increasing, u_2 strictly decreasing. Moreover, $u_1 = s+\beta_1$ and $u_2 = s+b_2-c_2$ on $[0, \hat{p}[$, u_1 and u_2 coincide with w_1 and w_2, respectively, at $\hat{p}$, and $u_1 = s+b_1-c_1$ and $u_2 = s+\beta_2$ on

$]\hat{p}, 1]$. So u_1 and u_2 are indeed the players' payoff functions for (k_1, k_2).

As $u_1 > w_1$ and $u_2 > s$ on $[0, \hat{p}[$, we have $b_1 < c_1$ and $b_2 > c_2$ on this interval. Similarly, as $u_1 > s$ and $u_2 > w_2$ on $]\hat{p}, 1]$, we have $b_1 > c_1$ and $b_2 < c_2$ there. Finally, continuity of u_1 and u_2 implies $b_1(\hat{p}-, u_1) = c_1(\hat{p}) < b_1(\hat{p}+, u_1)$ and $b_2(\hat{p}+, u_2) = c_2(\hat{p}) < b_2(\hat{p}-, u_2)$. Now consider player 1 at the belief $\hat{p}$. Whatever action he chooses, the belief can only drift upward conditional on there being no success, so it is the right-hand derivative of u_1, and hence the right limit $b_1(\hat{p}+, u_1)$, that matters in the Bellman equation. As this limit is strictly below $c_1(\hat{p})$, player 1's strict best response at $\hat{p}$ is to play risky. An analogous argument works for player 2. This establishes that k_1 and k_2 are mutual best responses at all beliefs.

To see that there are no other equilibria, note again that in any MPE $(k_1, k_2)(1) = (1, 0)$ and $(k_1, k_2)(0) = (0, 1)$. Moreover, by Lemmas 6.1 and 6.2, there might potentially be two ways of transitioning out of $(0, 1)$, namely either via $(0, 1)$—$(1, 1)$—$(1, 0)$, which by Lemma 6.2 can only happen at points in the interval $[1 - p^m, p^m]$, or via $(0, 1)$—$(0, 1)$—$(1, 1)$ or $(0, 1)$—$(1, 1)$—$(1, 1)$, which by Lemma 6.1 can only happen at p^m. Now, suppose that there exists an MPE where players transition from $(0, 1)$ into $(1, 1)$ at p^m. To the right of p^m, players cannot transition back into $(0, 1)$ as, to the right of p^m, there is no way for them to transition out of $(0, 1)$ again. Moreover, they can only transition from $(1, 1)$ to $(1, 0)$ via $(1, 1)$—$(1, 1)$—$(1, 0)$ or $(1, 1)$—$(1, 0)$—$(1, 0)$, both of which can only happen at $1 - p^m < p^m$. Thus, in such an MPE, we must have $(k_1, k_2)(1) = (1, 1)$ – a contradiction.

Therefore, in any MPE, there exists a belief $\hat{p} \in [1 - p^m, p^m]$ at which a transition of the form $(0, 1)$—$(1, 1)$—$(1, 0)$ occurs. Now, the only ways for the players to transition out of $(1, 0)$ to the right of $\hat{p}$ are the following: $(1, 0)$—$(1, 0)$—$(0, 1)$, $(1, 0)$—$(0, 1)$—$(0, 1)$ or $(1, 0)$—$(0, 1)$—$(1, 1)$, since by Lemma 6.1 $(1, 0)$—$(1, 0)$—$(1, 1)$ and $(1, 0)$—$(1, 1)$—$(1, 1)$ can only occur at $1 - p^m$, and Lemma 6.2(iv) rules out $(1, 0)$—$(1, 1)$—$(0, 1)$.

Suppose therefore there exists a belief $p^\dagger > \hat{p}$ at which a transition of the form $(1, 0)$—$(0, 1)$—$(1, 1)$ occurs. By continuity of the value function, this implies that $u_1(p^\dagger) = p^\dagger g + C(1 - p^\dagger)^{\frac{r+\lambda}{\lambda}}(p^\dagger)^{-\frac{r}{\lambda}} = w_1(p^\dagger)$. By the same token, we have $u_1(\hat{p}) = \hat{p}g + C(1 - \hat{p})^{\frac{r+\lambda}{\lambda}}\hat{p}^{-\frac{r}{\lambda}} = w_1(\hat{p})$. Both equations can only hold if $\left(\frac{p^\dagger}{1-p^\dagger}\right)^{\frac{r}{\lambda}} \frac{\lambda}{\lambda+r}s = \left(\frac{\hat{p}}{1-\hat{p}}\right)^{\frac{r}{\lambda}} \frac{\lambda}{\lambda+r}s$ – a contradiction because the function $p \mapsto \frac{p}{1-p}$ is strictly monotone.

Now, suppose there exists a $p^\ddagger > \hat{p}$ where a transition either of the form $(1, 0)$—$(0, 1)$—$(0, 1)$ or $(1, 0)$—$(1, 0)$—$(0, 1)$ occurs. Take $p^\ddagger$ to be the smallest such belief. By continuity and Lemma A.1, we have $u_1(p^\ddagger) \geq w_1(p^\ddagger)$. Yet we also have that $u_1(\hat{p}) = w_1(\hat{p})$ – a contradiction since w_1 is linear, and on any non-degenerate interval where the action profile $(1, 0)$ prevails, player 1's value function is easily seen to be strictly convex.

Hence we have shown that in any MPE, players will not transition out of $(1, 0)$ to the right of $\hat{p}$. This already implies uniqueness of $\hat{p}$. The only transitions that remain to be ruled out are $(1, 0)$—$(1, 1)$—$(1, 0)$ and $(0, 1)$—$(1, 1)$—$(0, 1)$, which by Lemma 6.1 can only occur at the beliefs $1 - p^m$ and p^m, respectively. If $(1, 0)$—$(1, 1)$—$(1, 0)$ did occur at $1 - p^m$, $(1, 0)$ would be played to the left of $1 - p^m$ – a contradiction. And if $(0, 1)$—$(1, 1)$—$(0, 1)$ occurred at p^m, $(0, 1)$ would be played to the right of p^m – another contradiction.

Thus, we have shown that there is exactly one transition in any MPE, occurring at a belief $\hat{p} \in [1-p^m, p^m]$. For the case where $p^m > 1-p^*$, we shall now show that in fact $\hat{p} \in [p^*, 1-p^*]$. Indeed, suppose that $\hat{p} < p^*$. Then, $\hat{p}g + \frac{\lambda}{\lambda+r}(1-\hat{p})s < s$ and, by the explicit expression for player 1's value function, $u_1 < s$ on $]0, \hat{p}[$, which is incompatible with player 1 playing a best response. By an analogous argument, we can rule out $\hat{p} > 1-p^*$. ∎

References

Acemoglu, D., P. Aghion, C. Lelarge, J. v. Reenen and F. Zilibotti (2006): "Technology, Information and the Decentralization of the Firm," CEPR Discussion Paper No. 5678.

Aghion, P., R. Blundell, R. Griffith, P. Howitt and S. Prantl (2006): "The Effects of Entry on Incumbent Innovation and Productivity," NBER Working Paper No. W12027.

Aghion, P. and P. Howitt (2006): "Growth Policy: A Unifying Framework (Joseph Schumpeter Lecture)," *Journal of the European Economic Association*, **Vol. 4 No. 2–3**, 269–314.

Aghion, P., A. Alesina and F. Trebbi (2007): "Democracy, Technology, and Growth," NBER Working Paper No. W13180.

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

Bolton, P. and C. Harris (2000): "Strategic Experimentation: the Undiscounted Case," in *Incentives, Organizations and Public Economics – Papers in Honour of Sir James Mirrlees*, ed. by P.J. Hammond and G.D. Myles. Oxford: Oxford University Press, 53–68.

Camargo, B. (2007): "Good News and Bad News in Two–Armed Bandits," *Journal of Economic Theory*, **135**, 558–566.

Dewatripont, M. and J. Tirole (1999): "Advocates," *Journal of Political Economy*, **107**, 1–39.

Hopenhayn, H. and F. Squintani (2006): "Preemption Games with Private Information," working paper, University of California at Los Angeles and University College London.

Keller, G. and S. Rady (2007): "Strategic Experimentation with Poisson Bandits," working paper, University of Oxford and University of Munich.

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

Moscarini, G. and F. Squintani (2004): "Competitive Experimentation with Private Information," Cowles Foundation Discussion Paper No. 1489.

Murto, P. and J. Välimäki (2006): "Learning in a Model of Exit," Helsinki Center of Economic Research Working Paper No. 110.

Pastorino, E. (2005): "Essays on Careers in Firms," Ph.D. Dissertation, University of Pennsylvania.

Presman, E.L. (1990): "Poisson Version of the Two-Armed Bandit Problem with Discounting," *Theory of Probability and its Applications*, **35**, 307–317.

Rosenberg, D., E. Solan and N. Vieille (2007a): "Social Learning in One–Armed Bandit Problems," *Econometrica*, **75**, 1591–1611.

Rosenberg, D. and E. Solan, N. Vieille (2007b): "Informational Externalities and Emergence of Consensus," working paper, Université Paris Nord, Tel Aviv University and HEC.