Strategic Experimentation with Poisson Bandits*

Godfrey Keller† Sven Rady‡

This version: July 12, 2006
Preliminary and Incomplete

Abstract

This paper studies a game of strategic experimentation with two-armed bandits where the risky arm distributes lump-sum payoffs according to a Poisson process. The intensity of this process is either high or low, and unknown to the players. We consider Markov perfect equilibria with beliefs as the state variable. There is no equilibrium where all players use cut-off strategies, and all equilibria exhibit an 'encouragement effect' relative to the single-agent optimum. The fact that a success on the risky arm is not fully revealing limits the payoff asymmetry as well as the inefficiency of asymmetric equilibria. We construct the unique symmetric MPE and, for a sufficiently high difference between the two possible intensities, an asymmetric equilibrium that dominates the symmetric one in terms of aggregate payoffs.

Keywords: Strategic Experimentation, Two-Armed Bandit, Poisson Process, Bayesian Learning, Markov Perfect Equilibrium.

JEL Classification Numbers: C73, D83, O32.

*Our thanks for very helpful discussions and suggestions are owed to Martin Cripps and to seminar participants at the universities of Munich, Nürnberg, Oxford, Toulouse, Wisconsin–Madison and the 2002 CESifo Area Conference on Industrial Organization in Munich. We thank the Institut d'Economie Industrielle at the University of Toulouse and the Center for Economic Studies at the University of Munich for their hospitality. Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

†Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK.

‡Department of Economics, University of Munich, Kaulbachstr. 45, D-80539 Munich, Germany.

1 Introduction

When a new restaurant of unknown quality arrives in your neighborhood you can choose to visit it and risk getting a bad meal yourself; or you can wait until an acquaintance does and then find out about their meal. Furthermore, it may be difficult to determine the quality of the restaurant from one visit alone – it may take many visits to find out whether it is good or bad – so this is a dynamic problem in which the players can perform repeated costly experiments (visit the restaurant) or learn from the experimental observations of others. There are strategic issues in this: first because you can choose to free-ride on the costly information acquisition of your acquaintances (and they can on yours), and second because you can generate information now which may encourage others to share the burden of acquiring information in the future.

Such a game of *strategic experimentation* arises in a variety of economic contexts; besides consumer search (as in the restaurant example) or experimental consumption (of a new drug, for instance), firms' research and development activities are a prominent example. Academic researchers pursuing a common research agenda or simply working on a joint paper are also effectively engaged in strategic experimentation.

In the present paper we analyze a game of strategic experimentation based upon two-armed bandits with a safe arm that offers a known and constant flow payoff and a risky arm whose lump-sum payoffs are driven by a Poisson process of unknown intensity. The risky arm can be either 'good' or 'bad': if it is good, the lump-sums arrive more frequently than if it is bad. The players have replica two-armed bandits with all risky arms being of the same type (all good, or all bad), but with lump-sums arriving independently. Each player is endowed with one unit of a perfectly divisible resource per unit of time, and continually has to decide what fraction of this resource to allocate to each arm. Players observe each others' actions and outcomes, so information about the type of the risky arm is a public good.

With Poisson bandits, news arrives in a 'lumpy' fashion. For concreteness, we focus on a situation where this news is good. Examples would be the occasional 'breakthrough' in research and development, a completed research paper in a longer-term research agenda, or one of a sequence of crucial proofs in a paper. Beliefs jump to a more optimistic level whenever a 'news event' or 'success' occurs, whereas they gradually become more pessimistic in between such events.

A single success on the risky arm does not fully reveal its type. This stands in marked contrast to the experimentation model of Keller, Rady and Cripps (2005). There, a good risky arm also generates lump-sum payoffs according to some Poisson process, but a bad risky arm never generates any payoffs, so the belief jumps all the way to certainty as soon as the first lump-sum arrives, irrespective of the belief held immediately before. In the present model, there is never certainty about the state of the risky arm, and the belief immediately after a success depends non-linearly on the belief held immediately before. As a consequence, when the players in our model use Markov strategies with the posterior belief as the state variable, their value functions solve first-order differential-difference equations. These equations can be analyzed by elementary methods and admit closed-form solutions.

While all Markov perfect equilibria of the experimentation game are inefficient because of free-riding, we show that they always exhibit an 'encouragement effect': the presence of other players encourages at least one of them to continue experimenting with the risky arm at beliefs where a single agent would already have given up. This effect was first described by Bolton and Harris (1999) in a model where the risky arm yields a flow payoff with Brownian noise. Focusing on the symmetric MPE of their model, however, they obtain the encouragement effect for this particular equilibrium only. The first contribution of the present paper is therefore to establish this effect for *all* MPE of the Poisson model.

The unique symmetric MPE of the Poisson model shares the main features with its counterpart in the Bolton-Harris model. All players use the risky arm exclusively when they are sufficiently optimistic, the safe arm when they are sufficiently pessimistic, and both arms simultaneously at intermediate beliefs. Both the incentive to free-ride and the encouragement effect become stronger as the number of players increases, with the lower bound of the intermediate range of beliefs falling, and the upper bound rising. Finally, the acquisition of information is slowed down so severely near the lower bound of the intermediate range that the players' beliefs cannot reach this bound in finite time.

This strongly suggests that asymmetric equilibria where a last experimenter keeps the rate of information acquisition bounded away from zero at pessimistic beliefs ought to be more efficient than the symmetric one. Bolton and Harris (2000), which studies the undiscounted limit of the Brownian model, and Keller, Rady and Cripps (2005) confirm this by constructing a variety of asymmetric MPE that dominate the symmetric one in terms of aggregate payoffs. However, they do so in environments without the encouragement effect. In fact, the existence and structure of asymmetric equilibria *with* an encouragement effect has remained an open question in the literature so far.

The second contribution of the present paper is to provide several results on asymmetric Markov perfect equilibria in the Poisson model. First, we show that there is no MPE in which all players use cut-off strategies, i.e. use the risky arm exclusively when the probability they assign to the risky arm being good is above some cut-off, and the safe arm when it is below. The intuition behind this result is the same as in Keller, Rady and Cripps (2005).

Second, we show that the most inequitable (and least efficient) asymmetric MPE constructed for an arbitrary number of players in Keller, Rady and Cripps (2005) have no counterpart in the Poisson model. Thus, the fact that a success on the risky arm is not fully revealing limits the payoff asymmetry as well as the inefficiency of asymmetric equilibria.

Third, we construct an asymmetric MPE for an arbitrary number of players under the assumption that the frequency of lump-sums on a bad risky arm is sufficiently low. These equilibria generate a higher aggregate payoff than the symmetric MPE. They do so by combining behavior as in the symmetric equilibrium (at optimistic beliefs) with a different behavior (at more pessimistic beliefs) where players take turns, one at a time, to play the risky arm exclusively while all others free-ride. As in Keller, Rady and Cripps (2005), the gain in aggregate payoffs stems from the fact that, owing to this alternation, the intensity of experimentation is bounded away from zero immediately to the right of the belief where all experimentation stops for good.

The paper is organized as follows. Section 2 sets up the Poisson bandit model. Section 3 establishes the efficient benchmark where players coordinate in order to maximize joint expected payoffs. Section 4 introduces the strategic problem and shows that, because of free-riding, any equilibrium of the game leads to inefficiently low levels of experimentation. It also proves that there is no MPE where each player uses a cut-off strategy, and shows that any MPE exhibits the encouragement effect. Section 5 presents the unique symmetric MPE. Section 6 constructs asymmetric equilibria. Section 7 contains concluding remarks. Some of the proofs are relegated to the Appendix.

2 Poisson Bandits

The purpose of this section is to introduce the Poisson model of strategic experimentation. For mathematical details, see Presman (1990) or Presman and Sonin (1990).

Time $t \in [0, \infty)$ is continuous, and the discount rate is $r > 0$. There are $N \geq 1$ players, each of them endowed with one unit of a perfectly divisible resource per unit of time. Each player faces a two-armed bandit problem where she continually has to decide what fraction of the available resource to allocate to each arm. One arm S is 'safe' and yields a known expected flow payoff that is proportional to the fraction of the resource allocated to it. The other arm R is 'risky' and yields lump-sum payoffs at random times, the arrival rate of these payoffs being proportional to the fraction of the resource allocated to it. The risky arm can be either 'good' for all players or 'bad' for all players: if it is good, the lump-sums (or 'successes') arrive more frequently than if it is bad. Conditional on the type of the risky arm, the arrival of lump-sums is independent across players.

If a player allocates the fraction $k_t \in [0, 1]$ of the resource to R over an interval of time $[t, t + dt)$, and consequently the fraction $1 - k_t$ to S, then she receives the expected payoff $(1 - k_t)s\,dt$ from S, where $s > 0$ is a constant known to all players. Moreover, the probability that she receives a lump-sum payoff from R at some point in the interval is $k_t\lambda_\theta\,dt$, where $\theta = 1$ if R is good, $\theta = 0$ if R is bad, and $\lambda_1 > \lambda_0 > 0$ are constants known to all players. If a player uses a time-invariant allocation $k_t = k$, therefore, lump-sum payoffs arrive according to a Poisson process with intensity $k\lambda_\theta$.

Lump-sums are independent draws from a time-invariant distribution on $\mathbb{R}_{++}$ with a known mean h. When the fraction k_t of the resource is allocated to R on $[t, t + dt)$, therefore, the overall expected payoff increment conditional on θ is $[(1-k_t)s + k_t\lambda_\theta h]\,dt$. We assume that $\lambda_0 h < s < \lambda_1 h$, so each player strictly prefers R to S if R is good, and strictly prefers S to R if R is bad. At $t = 0$, however, players do not know whether the risky arm is good or bad; they start with a common prior belief about θ. Thereafter, all players observe each other's actions and outcomes, so they hold common posterior beliefs throughout time.

Given a player's actions $\{k_t\}_{t\geq 0}$ such that k_t is measurable with respect to the information available at time t, her total expected discounted payoff, expressed in per-period units, is

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

where the expectation is over both the random variable θ and the stochastic process $\{k_t\}$.

Let p_t denote the subjective probability at time t that players assign to the risky arm being good, so that given a player's action k_t on $[t, t+dt)$, her expected payoff increment conditional on all available information is $[(1-k_t)s + k_t\lambda(p_t)h]\, dt$ with

$$\lambda(p) = p\lambda_1 + (1-p)\lambda_0.$$

By the Law of Iterated Expectations, we can rewrite the above total payoff as

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

where the expectation is now over the stochastic processes $\{k_t\}$ and $\{p_t\}$. This highlights the potential for beliefs to serve as a state variable. Note that a player's payoff depends on others' actions only through their impact on the evolution of beliefs.

To derive the law of motion of beliefs, suppose that over the interval of time $[t, t+dt)$ player $n = 1, \ldots, N$ allocates the fraction $k_{n,t}$ of the unit resource to her risky arm. Let $K_t = \sum_{n=1}^N k_{n,t}$. This sum measures how much of the N units of the resource is allocated to risky arms at a given time t – we will call this number the *intensity of experimentation*. If the risky arms are good, the probability of none of the players receiving a lump-sum payoff is $\prod_{n=1}^N (1 - k_{n,t}\lambda_1\, dt) = 1 - K_t\lambda_1\, dt$ (this is up to terms of the order $o(dt)$, which we can ignore here and in what follows); if the risky arms are bad, this probability is $1 - K_t\lambda_0\, dt$. When the players start with the common belief p_t at time t and none of them receives a lump-sum payoff in $[t, t+dt)$, therefore, the updated belief at the end of that time period is

$$p_t + dp_t = \frac{p_t\,(1 - K_t\lambda_1\, dt)}{(1-p_t)\,(1 - K_t\lambda_0\, dt) + p_t\,(1 - K_t\lambda_1\, dt)}$$

by Bayes' rule. Simplifying, we see that as long as no lump-sum arrives, the belief changes by $dp_t = -K_t \Delta\lambda\, p_t(1-p_t)\, dt$, where $\Delta\lambda = \lambda_1 - \lambda_0$. Once any of the players receives a lump-sum, the belief jumps up to

$$j(p) = \frac{\lambda_1 p}{\lambda(p)}\ ,$$

which is independent of the intensity of experimentation.

A myopic agent would simply weigh the short-run payoff from playing the safe arm, s, against what she expects from playing the risky arm, $\lambda(p)h$. So we define p^m as the belief that makes her indifferent between these choices,

$$p^m = \frac{s - \lambda_0 h}{\Delta\lambda\, h}\,.$$

For $p > p^m$ it is myopically optimal to play R; for $p < p^m$ it is myopically optimal to play S.

3 The Cooperative Problem

Suppose that the N players work cooperatively by jointly choosing the action profiles $\{(k_{1,t}, \ldots, k_{N,t})\}_{t \geq 0}$ so as to maximize the *average* expected payoff. If the current belief is p and the current actions are $(k_1, \ldots, k_N)$, the average expected payoff increment is $\left[(1 - \frac{K}{N})s + \frac{K}{N}\lambda(p)h\right] dt$ with $K = \sum_{n=1}^{N} k_n$, and the subjective probability of a breakthrough on at least one risky arm is $K\lambda(p)\, dt$. As the evolution of beliefs also merely depends on K, the cooperative solves an optimal control problem with state variable p and choice variable K.

By exactly the same arguments as in Keller, Rady and Cripps (2005), the value function for the cooperative, expressed as average payoff per agent, satisfies the Bellman equation

$$u(p) = s + \max_{K \in [0,N]} K \left\{ b(p,u) - c(p)/N \right\},$$

where

$$c(p) = s - \lambda(p)h$$

and

$$b(p,u) = [\lambda(p)\,(u(j(p)) - u(p)) - \Delta\lambda\, p(1-p)u'(p)]/r.$$

Clearly, $c(p)$ is the opportunity cost of playing R; the other term, $b(p,u)$, is the expected benefit of playing R, and has two parts: $\lambda(p)(u(j(p)) - u(p))$ is the expected improvement in the overall payoff should a success occur; $-\Delta\lambda\, p(1-p)u'(p)$, on the other hand, is the negative effect on the overall payoff should no success occur.[1]

If the shared opportunity cost of playing R exceeds the full expected benefit, the optimal choice is $K = 0$ (all agents use S exclusively), and $u(p) = s$. Otherwise, $K = N$ is optimal (all agents use R exclusively), and u satisfies the first-order ordinary differential-difference equation (henceforth ODDE)

$$\Delta\lambda\, p(1-p)u'(p) - \lambda(p)[u(j(p)) - u(p)] + \tfrac{r}{N}u(p) = \tfrac{r}{N}\lambda(p)h. \qquad (1)$$

[1]Note that infinitesimal changes of the belief are always downward, so strictly speaking only the left-hand derivative of the value function u matters here. While this turns out to be of no relevance to the cooperative case, we will indeed see equilibria of the strategic experimentation game where a player's payoff function is not of class C^1.

A particular solution to this equation is $u(p) = \lambda(p)h$, the expected per capita payoff from all agents using the risky arm forever.

The option value of being able to change to the safe arm is then captured by the solution to the homogeneous equation, for which we try $u_0(p) = (1-p)\left(\frac{1-p}{p}\right)^\mu$ for some $\mu > 0$ to be determined.[2] Now,

$$u_0'(p) = -\frac{\mu + p}{p(1-p)}\, u_0(p) \quad \text{and} \quad u_0(j(p)) = \frac{\lambda_0}{\lambda(p)}\left(\frac{\lambda_0}{\lambda_1}\right)^\mu u_0(p).$$

Inserting these into the homogeneous equation and simplifying leads to the requirement that

$$\tfrac{r}{N} + \lambda_0 - \mu\Delta\lambda = \lambda_0\left(\frac{\lambda_0}{\lambda_1}\right)^\mu. \tag{2}$$

As a function of μ, the left-hand side of (2) is a negatively sloped straight line which cuts the vertical axis at $\frac{r}{N}+\lambda_0$. The right-hand side is a decreasing exponential function which tends to 0 as $\mu \to +\infty$, tends to ∞ as $\mu \to -\infty$, and cuts the vertical axis at λ_0. Thus the above equation in μ has two solutions, one positive and one negative; we write μ_N for the positive solution, which obviously lies strictly between $\frac{r}{N\Delta\lambda}$ (the value of μ where the left-hand side of (2) equals λ_0) and $\frac{r}{N\Delta\lambda} + \frac{\lambda_0}{\Delta\lambda}$ (the value of μ where it equals 0). As the left-hand side of (2) rises with $\frac{r}{N}$, we also see that μ_N is increasing in the discount rate and decreasing in the number of agents.

The solution to the ODDE for the cooperative case is thus

$$V_N(p) = \lambda(p)h + C\,(1-p)\,\Omega(p)^{\mu_N}, \tag{3}$$

where

$$\Omega(p) = (1-p)/p$$

denotes the odds ratio at the belief p. At sufficiently optimistic beliefs, the option value of being able to change to the safe arm (captured by the second term of V_N) is positive, implying a positive constant of integration C and a convex solution V_N.

Proposition 1 (Cooperative solution) *In the N-agent cooperative problem, there is a cut-off belief p_N^* given by*

$$p_N^* = \frac{\mu_N(s-\lambda_0 h)}{(\mu_N+1)(\lambda_1 h - s) + \mu_N(s-\lambda_0 h)} \tag{4}$$

such that below the cut-off it is optimal for all to play S exclusively and above it is optimal for all to play R exclusively. The value function V_N^* for the N-agent cooperative is given by

$$V_N^*(p) = \lambda(p)h + [s - \lambda(p_N^*)h]\left(\frac{1-p}{1-p_N^*}\right)\left(\frac{\Omega(p)}{\Omega(p_N^*)}\right)^{\mu_N} \tag{5}$$

[2] This guess can be obtained by 'extrapolation' from the limiting case $\lambda_0 = 0$ studied in Keller, Rady and Cripps (2005). In this case, $j(p) = 1$ and $u(j(p)) = \lambda_1 h$, so (1) becomes a linear differential equation; the above function u_0 is easily seen to solve the corresponding homogeneous equation for $\mu = r/(N\lambda_1)$. A more systematic approach relies on a change of the independent variable from p to $\ln\frac{1-p}{p}$. This transforms (1) into a linear ODDE with constant delay to which results from Bellman and Cooke (1963) can be applied.

when $p > p_N^*$, and $V_N^*(p) = s$ otherwise.

PROOF: The expression for p_N^* and the constant of integration in (5) are obtained by imposing $V_N^*(p_N^*) = s$ (value matching) and $(V_N^*)'(p_N^*) = 0$ (smooth pasting). Then, $b(p, V_N^*)$ falls short of $c(p)/N$ to the left of p_N^*, coincides with it at p_N^*, and exceeds it to the right of p_N^*. So V_N^* solves the Bellman equation, with the maximum being achieved at the intensity of experimentation stated in the proposition. ■

This result confirms the findings of Rothschild (1974) and Presman (1990) for one agent, and Bolton and Harris (1999) and Keller, Rady and Cripps (2005) for several agents: the optimal strategy has a threshold where the agents change irrevocably from R to S; there are occasions where the agents make a mistake by changing from R to S although the risky action is actually better (R is good); the probability of mistakes decreases as the reward from the safe action decreases, as agents become more patient, and as the number of agents increases. In fact, as μ_N decreases with $\frac{r}{N}$, so does p_N^*. Finally, it is straightforward to show that each agent's payoff $V_N^*(p)$ increases in N over the range of beliefs where playing the risky arm is optimal.

The above proposition determines the *efficient* strategies. Note that the efficient intensity of experimentation exhibits a bang-bang feature, being N when the current belief is above p_N^*, and 0 when it is below. Thus, starting from a prior belief $p_0 > p_N^*$, the efficient intensity is maximal as long as successes occur frequently enough, and minimal after a sufficiently long spell without a success.

4 The Strategic Problem

From now on, we assume that there are $N \geq 2$ players acting non-cooperatively. We consider Markovian strategies with the common belief as the state variable. We first collect some results whose proofs carry over *verbatim* from Keller, Rady and Cripps (2005).

Bellman equation. Fix a belief p. With $k_n \in [0, 1]$ indicating player n's action at that belief and $K = \sum_{n=1}^N k_n$, let $K_{\neg n} = K - k_n$, which summarizes the actions of the other players. Player n's value function satisfies the Bellman equation

$$u_n(p) = s + K_{\neg n}\, b(p, u_n) + \max_{k_n \in [0,1]} k_n \left\{ b(p, u_n) - c(p) \right\},$$

where the second term on the right-hand side measures the benefit of the information generated by the other players. Arguing similarly to Bolton and Harris (1999), one shows easily that this benefit is always non-negative and that a player's value function is non-decreasing in the other players' intensity of experimentation.

Best responses. Player n's best response, k_n^*, is determined by comparing the opportunity cost of playing R with the expected *private* benefit. If $c(p) > b(p, u_n)$, then $k_n^* = 0$, and the Bellman equation implies $u_n(p) = s + K_{\neg n}\, b(p, u_n) < s + K_{\neg n}\, c(p)$. If

$c(p) = b(p, u_n)$, then k_n^* is arbitrary in $[0, 1]$ and $u_n(p) = s + K_{\neg n}\, c(p)$. Finally, if $c(p) < b(p, u_n)$, then $k_n^* = 1$ and $u_n(p) = s + (K_{\neg n} + 1)\, b(p, u_n) - c(p) > s + K_{\neg n}\, c(p)$. Thus, player n's best response depends on whether in the (p, u)-plane, the point $(p, u_n(p))$ lies below, on or above the line

$$\mathcal{D}_{K_{\neg n}} := \{(p, u) \in [0, 1] \times I\!\!R_+ : \ u = s + K_{\neg n}\, c(p)\}.$$

For $K_{\neg n} > 0$ this is a downward sloping diagonal that cuts the safe payoff line $u = s$ at $p = p^m$, the myopic cut-off; for $K_{\neg n} = 0$, it coincides with the safe payoff line.

ODDEs for the payoff functions. If $k_n^* = 0$ then, with $K = K_{\neg n}$, u_n satisfies the ODDE

$$K \Delta\lambda\, p(1 - p)u'(p) - K\lambda(p)[u(j(p)) - u(p)] + ru(p) = rs. \tag{6}$$

If $k_n^* = 1$ then, with $K = K_{\neg n} + 1$, u_n satisfies

$$K \Delta\lambda\, p(1 - p)u'(p) - K\lambda(p)[u(j(p)) - u(p)] + ru(p) = r\lambda(p)h. \tag{7}$$

If player n is indifferent between all allocations of the resource to S and R, u_n solves

$$\Delta\lambda\, p(1 - p)u'(p) - \lambda(p)[u(j(p)) - u(p)] = r\lambda(p)h - rs. \tag{8}$$

Bounds on payoffs. No profile of Markov strategies can generate an average payoff that exceeds V_N^*, and the payoff of a player using a best response to her opponents' strategies cannot fall below V_1^*. The upper bound follows immediately from the fact that the cooperative solution maximizes the average payoff. The intuition for the lower bound is that an agent can only benefit from the information generated by others.

Inefficiency of equilibria. All Markov perfect equilibria are inefficient. Along the efficient experimentation path, the benefit of an additional experiment tends to $1/N$ of its opportunity cost as p approaches p_N^*. A self-interested player compares the benefit of an additional experiment with the full opportunity cost and so has an incentive to deviate from the efficient path by using S instead of R.

The following result, which rules out equilibra where all players use cut-off strategies, also carries over from Keller, Rady and Cripps (2005).

Proposition 2 (No MPE in cut-off strategies) *In any Markov perfect equilibrium, at least one player uses a strategy that is not of the cut-off type.*

PROOF: Suppose to the contrary that there is an MPE where all players use a cut-off strategy. For $n = 1, \ldots, N$, let p_n denote the belief at which player n switches from using R exclusively to using S exclusively. Clearly, $p_n \leq p^m$ for all n. Without loss of generality, we can assume that $p_1 \leq p_2 \leq \ldots \leq p_{N-1} \leq p_N$. Moreover, we must have $p_1 < p^m$ since each player would have an incentive to deviate to the optimal strategy of a single player otherwise.

Suppose that $p_1 = p_2$. Immediately to the right of this cut-off, both u_1 and u_2 must then lie strictly below $\mathcal{D}_1$, so players 1 and 2 playing R are not best responses. This proves that $p_1 < p_2$.

Now, u_2 must lie below $\mathcal{D}_1$ immediately to the left of p_2 (as player 2 finds it optimal to free-ride on one opponent who plays R) and above $\mathcal{D}_1$ immediately to the right of p_2 (as player 2 finds it optimal to join in with at least one opponent who plays R), so u_2 crosses $\mathcal{D}_1$ at p_2. (In fact, one can iterate this argument to establish that all cut-offs are different, and that u_n crosses $\mathcal{D}_{n-1}$ at p_n.)

Since a player's value function is weakly increasing in the intensity of experimentation provided by the other players, we have $u_1 \leq u_2$, and so u_1 is either strictly below or exactly on $\mathcal{D}_1$ at p_2. In the first case, there is an interval $]p_2, p_2 + \epsilon[$ where player 1 (who is assumed to play R) is not responding optimally to the other players' combined intensity of experimentation $K_{\neg 1} = 1$. In the second case, $u_1 = u_2$ on $[p_2, 1]$ and $u_1'(p_2-) \geq u_2'(p_2-)$, hence $b(p_2, u_1) \leq b(p_2, u_2)$. But then, $u_2(p_2) = s + b(p_2, u_2) > s + b(p_2, u_1) - c(p_2) = u_1(p_2)$, a contradiction. ■

It is obvious that in any Markov perfect equilibrium, at least one player must be using the risky arm at any belief above p_1^*. The interesting question is whether experimentation continues below the single-agent optimum, i.e. whether there is an encouragement effect.

Proposition 3 (Encouragement effect) *In any Markov perfect equilibrium, at least one player experiments at some beliefs below p_1^*.*

PROOF: Suppose to the contrary that all players play S at all beliefs $p \leq p_1^*$. Then each player's payoff function satisfies $u_n(p_1^*) = s$ with the left-hand derivative $u_n'(p_1^*-) = 0$. For S to be optimal we must have $b(p_1^*, u_n) \leq c(p_1^*) = b(p_1^*, V_1^*)$, and hence $u_n(j(p_1^*)) \leq V_1^*(j(p_1^*))$, which must in fact hold as an equality. Thus, the difference $u_n - V_1^*$ assumes its minimum (of 0) at $j(p_1^*)$, which implies $u_n'(j(p_1^*)-) \leq (V_1^*)'(j(p_1^*)-)$. As $u_n(j^2(p_1^*)) \geq V_1^*(j^2(p_1^*))$, this implies $b(j(p_1^*), u_n) \geq b(j(p_1^*), V_1^*)$ and hence $b(j(p_1^*), u_n) > c(j(p_1^*))$. So all players must use R at the belief $j(p_1^*)$. By the ODDE for V_1^* and the explicit solution in Proposition 1, we have $b(j(p_1^*), V_1^*) = V_1^*(j(p_1^*)) - s + c(j(p_1^*)) = V_1^*(j(p_1^*)) - \lambda(j(p_1^*))h > 0$. Each player's Bellman equation now yields

$$\begin{aligned} u_n(j(p_1^*)) &= s + N\, b(j(p_1^*), u_n) - c(j(p_1^*)) \\ &\geq s + N\, b(j(p_1^*), V_1^*) - c(j(p_1^*)) \\ &> s + b(j(p_1^*), V_1^*) - c(j(p_1^*)) \\ &= V_1^*(j(p_1^*)), \end{aligned}$$

which contradicts the equality $u_n(j(p_1^*)) = V_1^*(j(p_1^*))$ derived earlier. ■

The idea behind the proof is that the only way that all experimentation could stop at p_1^* is for the 'jump-benefit' to be the same for each of the N players as for a lone agent, given the same opportunity cost and the same 'slide-disbenefit'; but this would imply that u_n and V_1^* matched in value not only at p_1^* but also at $j(p_1^*)$. This is not possible, since at $j(p_1^*)$ the benefit of a further jump up is no less and the disbenefit of a slide down is no worse for player n than for a lone agent, and if a lone agent has an incentive to experiment then so do each of the N players, the positive externality resulting in a higher value at $j(p_1^*)$.

We now turn to a more detailed investigation of Markov perfect equilibria.

5 Symmetric Equilibrium

A symmetric Markov perfect equilibrium admits three possible cases at any given belief. First, when all players play S exclusively, the common payoff is $u(p) = s$. Second, when all players play R exclusively, the common payoff function u satisfies (1), hence is of the form V_N given in (3). Third, when all players divide the resource strictly between S and R, the common payoff function solves (8).

In the (p, u)-plane, the diagonal $\mathcal{D}_{N-1}$ separates the region where all players use the risky arm exclusively from the region where they use both arms. Given the post-jump value $u(j(p))$, we have smooth pasting of the solutions to (1) and (8) along $\mathcal{D}_{N-1}$. Smooth pasting occurs also at the boundary of the region where all players use S exlusively with the region where they use both arms. In other words, u must be of class C^1. To see this, suppose we had a symmetric equilibrium with a payoff function that hits the level s at the belief $\tilde{p}$ with slope $u'(\tilde{p}+) > 0$. Then, at beliefs immediately to the right of $\tilde{p}$, we would have $b(p, u) = c(p)$ or

$$\lambda(p)[u(j(p)) - u(p)]/r = c(p) + \Delta\lambda\, p(1-p)u'(p)/r$$

implying

$$\lambda(\tilde{p})[u(j(\tilde{p})) - s]/r = c(\tilde{p}) + \Delta\lambda\, \tilde{p}(1-\tilde{p})u'(\tilde{p}+)/r > c(\tilde{p})$$

by continuity. Immediately to the left of $\tilde{p}$, continuity of $u(j(p))$ and the fact that $u'(p) = 0$ would then imply $b(p, u) = \lambda(p)[u(j(p)) - s]/r > c(p)$, so there would be an incentive to deviate from S to R.

Proposition 4 (Symmetric equilibrium) *The N-player experimentation game has a unique symmetric Markov perfect equilibrium with the common posterior belief as the state variable. The corresponding payoff function is the unique function $W_N^\dagger \colon [0,1] \to [s, \lambda_1 h]$ of class C^1 with the following properties: $W_N^\dagger(p) = s$ on an interval $[0, \tilde{p}_N]$ with $p_N^* < \tilde{p}_N < p_1^*$; $W_N^\dagger(p) > s$ on $]\tilde{p}_N, 1]$; $W_N^\dagger$ solves (8) on an interval $]\tilde{p}_N, p_N^\dagger[$ with $\tilde{p}_N < p_N^\dagger < p^m$, and (1) on $]p_N^\dagger, 1[$. The equilibrium strategy is continuous in the posterior belief and satisfies $k_N^\dagger(p) = 0$ for $p \leq \tilde{p}_N$,*

$$k_N^\dagger(p) = \frac{1}{N-1}\frac{W_N^\dagger(p) - s}{c(p)} \quad \in\ ]0,1[$$

for $\tilde{p}_N < p < p_N^\dagger$, and $k_N^\dagger(p) = 1$ for $p \geq p_N^\dagger$. $W_N^\dagger$ is strictly increasing on $[\tilde{p}_N, 1]$, and $k_N^\dagger$ on $[\tilde{p}_N, p_N^\dagger]$.

PROOF: We just sketch the construction of the symmetric equilibrium here; for details and the proof of uniqueness, see the Appendix.

Varying the point of intersection with the diagonal $\mathcal{D}_{N-1}$, one first constructs a family of candidate value functions that solve the ODDE (1) (N players using R exclusively) above $\mathcal{D}_{N-1}$, and the ODDE (8) (indifference between R and S) below. Using an intermediate-value argument, one then establishes the existence of one such function that reaches the level s with zero slope as we move down from $p = p^m$ to lower beliefs. This function is easily seen to solve each player's Bellman equation. Finally, the identity $u_n(p) = s + K_{\neg n}\, c(p)$ uniquely determines the common intensity of experimentation in the range of beliefs where the value function lies below $\mathcal{D}_{N-1}$ but above the level s. ■

The symmetric equilibrium of the Poisson model shares the main features with its counterpart in the Brownian model of Bolton and Harris (1999). First, because of the incentive to free-ride, experimentation stops for good inefficiently early ($\tilde{p}_N$ is above the cooperative cut-off p_N^*), and the intensity of experimentation is inefficiently low at any belief between p_N^* and $p_N^\dagger$. Second, there is the encouragement effect ($\tilde{p}_N$ is below the single-agent cut-off p_1^*). Third, both the incentive to free-ride and the encouragement effect become stronger as the number of players increases.[3] Fourth, the acquisition of information is slowed down so severely near $\tilde{p}_N$ that the players' beliefs cannot reach this cut-off in finite time.

Corollary 1 *Starting from a prior belief above $\tilde{p}_N$, the players' common posterior belief never reaches this cut-off in the symmetric Markov perfect equilibrium.*

PROOF: Close to the right of $\tilde{p}_N$, the dynamics of the belief p given no success are

$$dp = -\Delta\lambda\, \frac{N}{N-1}\, \frac{W_N^\dagger(p) - s}{c(p)}\, p\,(1-p)\, dt.$$

(A success merely causes a delay before the belief decays to near $\tilde{p}_N$ again.) As $W_N^\dagger$ is of class C^2 to the right of $\tilde{p}_N$ with $W_N^\dagger(\tilde{p}_N) = s$, $(W_N^\dagger)'(\tilde{p}_N) = 0$ and $(W_N^\dagger)''(\tilde{p}_N+) \geq 0$, we can find a positive constant C such that

$$\Delta\lambda\, \frac{N}{N-1}\, \frac{W_N^\dagger(p) - s}{c(p)}\, p\,(1-p) < C\,(p - \tilde{p}_N)^2$$

in a neighborhood of $\tilde{p}_N$. Starting from an initial belief $p_0 > \tilde{p}_N$ in this neighborhood, consider the dynamics $dp = -C\,(p-\tilde{p}_N)^2\, dt$. The solution with initial value p_0,

$$p_t = \tilde{p}_N + \frac{1}{Ct + (p_0 - \tilde{p}_N)^{-1}}\,,$$

does not reach $\tilde{p}_N$ in finite time. Since the modified dynamics decrease faster than the original ones, this result carries over to the true evolution of beliefs. ■

[3] As N increases, the lower cut-off $\tilde{p}_N$ falls, the upper cut-off $p_N^\dagger$ rises, and each player obtains a higher payoff at all beliefs where the risky arm is used some of the time.

What differentiates the Poisson model from the Brownian one is that the above results can be obtained by elementary methods. In fact, we can represent the payoff function $W_N^\dagger$ in closed form up to some constants of integration that are implicitly determined by the cut-off $p_N^\dagger$.

Corollary 2 *Define intervals* $I_0 = [p_N^\dagger, 1]$ *and* $I_i = [j^{-i}(p_N^\dagger), j^{-(i-1)}(p_N^\dagger)[$ *for* $i = 1, 2, \ldots$ *If* $\mu_N \neq \lambda_0/\Delta\lambda$,[4] *then*

$$\begin{aligned} W_N^\dagger(p) &= \left(\lambda_1 h + \tfrac{r}{\lambda_1}(\lambda_1 h - s)\, i\right) p + \left(\lambda_0 h + \tfrac{r}{\lambda_0}(\lambda_0 h - s)\, i\right)(1-p) \\ &\quad + C^{(0)} \left(\frac{\lambda_0\,(\lambda_0/\lambda_1)^{\mu_N}}{\lambda_0 - \mu_N \Delta\lambda}\right)^i (1-p)\Omega(p)^{\mu_N} \\ &\quad + \sum_{n=0}^{i-1} \frac{C^{(i-n)}}{n!} \left(-\frac{\lambda_0\,(\lambda_0/\lambda_1)^{\lambda_0/\Delta\lambda}}{\Delta\lambda} \ln\left[(\lambda_0/\lambda_1)^{n-1}\,\Omega(p)\right]\right)^n (1-p)\Omega(p)^{\lambda_0/\Delta\lambda} \end{aligned}$$

on $I_i \cap \{p : W_N^\dagger(p) > s\}$ *for some constants* $C^{(i-n)}$ $(n = 0, \ldots i-1)$, *chosen to ensure continuity of* $W_N^\dagger$. *The constant* $C^{(0)}$ *ensuring that* $W_N^\dagger(p_N^\dagger) \in \mathcal{D}_{N-1}$ *is given by*

$$C^{(0)} = N\, c(p_N^\dagger)\,(1 - p_N^\dagger)^{-1}\,\Omega(p_N^\dagger)^{-\mu_N}.$$

PROOF: See the Appendix. The proof shows how the constants $C^{(i)}$ can be calculated recursively given $C^{(0)}$. ■

6 Asymmetric Equilibria

We call an equilibrium *simple* if the action profile $(k_1, \ldots, k_N)$ is always an element of $\{0,1\}^N$. We say that an equilibrium has a *last experimenter* if all experimentation stops at some belief $\bar{p}$ and there is a player n and an $\epsilon > 0$ such that $k_n = 1$ and $K_{\neg n} = 0$ on $]\bar{p}, \bar{p} + \epsilon[$.

For the case $\lambda_0 = 0$, Keller, Rady and Cripps (2005) construct simple asymmetric equilibria with a last experimenter whose payoff is strictly below his opponents' at all beliefs in $]\bar{p}, 1[$. For $N = 2$, they specify a continuum of such equilibria. Surprisingly, no such MPE exists when $\lambda_0 > 0$.

Proposition 5 (Last experimenter) *Consider a simple Markov perfect equilibrium of the* N*-player experimentation game that has a last experimenter. Let* $\bar{p}$ *denote the belief at which this player stops using* R *for ever. Then her payoff at the belief* $j(\bar{p})$ *is at least as high as that of any of her opponents.*

[4] The proof makes it obvious how one has to modify this result in the knife-edge case where $\mu_N = \lambda_0/\Delta\lambda$.

Proof: Optimal behaviour of the last experimenter (player 1, say) requires $c(\bar{p}) = b(\bar{p}, u_1) = \lambda(\bar{p})\,[u_1(j(\bar{p})) - s]/r$. If there were another player (player 2, say) with $u_2(j(\bar{p})) > u_1(j(\bar{p}))$, we would have $b(\bar{p}, u_2) = \lambda(\bar{p})\,[u_2(j(\bar{p})) - s]/r > c(\bar{p})$. So player 2 would act suboptimally on a nonempty interval of beliefs around $\bar{p}$. ■

When $\lambda_0 = 0$, Proposition 5 still holds but does not impose any restriction because all players' values jump to the same level, $\lambda_1 h$, when a lump-sum arrives. When $\lambda_0 > 0$, it rules out the most inequitable (and least efficient) equilibria constructed in Keller, Rady and Cripps (2005). Thus, the fact that a success on the risky arm is not fully revealing limits the asymmetry in the burden of experimentation (and hence in the payoffs) as well as the inefficiency that can arise in equilibrium.

Proposition 6.1 of Keller, Rady and Cripps (2005) allows for a simple two-player equilibrium where both players achieve exactly the same payoffs to the right of $\mathcal{D}_1$ and swap roles once below it. This equilibrium maximizes average payoffs over all simple two-player MPE with a last experimenter, and hence is the least complex equilibrium with that property. Our next aim is to show that this equilibrium survives when λ_0 is positive but small.

Figure 1 illustrates the best response correspondence for $N = 2$ and payoffs in the equilibrium we aim to construct. The faint straight line is $\mathcal{D}_1$, the solid kinked line the myopic payoff. The solid curves are the graphs of the players' value functions.

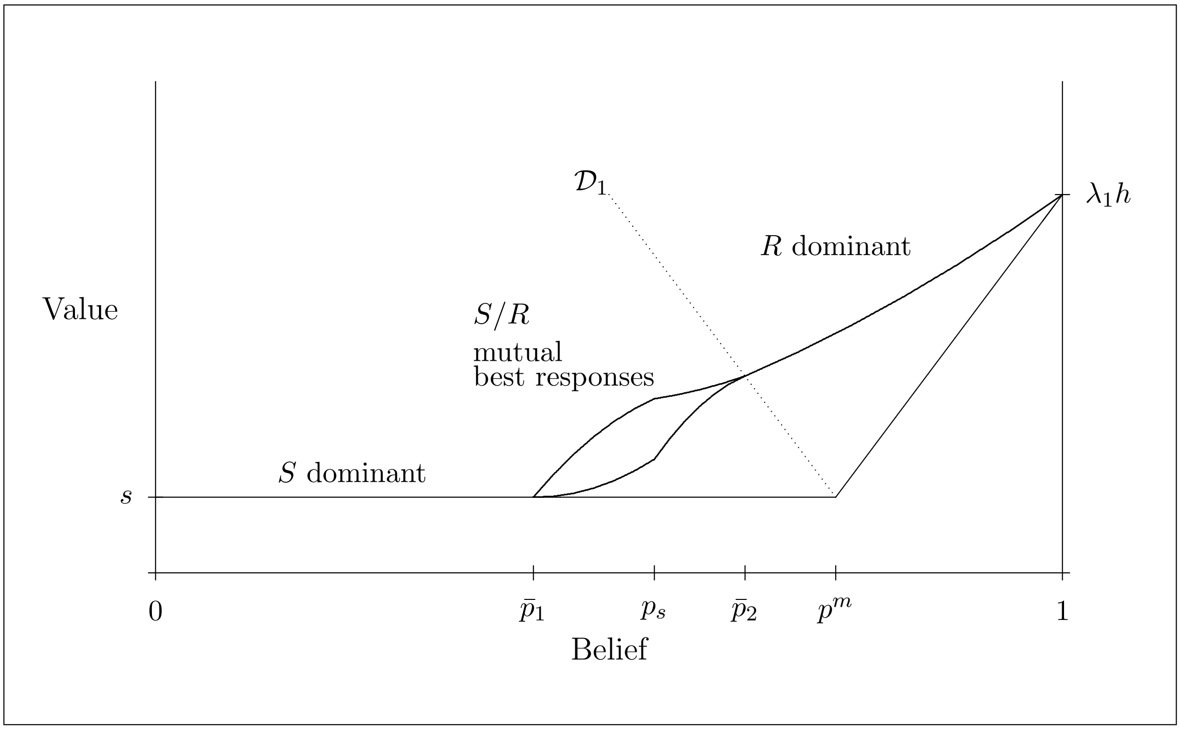

Figure 1: Best responses for $N = 2$, and payoffs in a simple asymmetric equilibrium

Proposition 6 (Simple asymmetric equilibrium for $N = 2$) *For λ_0 sufficiently close to 0, the two-player experimentation game admits a simple Markov perfect equilibrium with the following features. There are two cut-offs, $\bar{p}_1$ and $\bar{p}_2$, and a belief, p_s, with $p_2^* < \bar{p}_1 < p_s < \bar{p}_2 < p^m$, such that: on $]\bar{p}_2, 1]$, both players play R and have a common payoff function; on $]p_s, \bar{p}_2]$, player 1 plays S and player 2 plays R; on $]\bar{p}_1, p_s]$, player 1 plays R and player 2 plays S; on $[0, \bar{p}_1]$, they both play S.*

PROOF: We just sketch the main steps in the proof here. Details can be found in the Appendix.

We first choose λ_0 small enough so that $j(p_2^*) \geq p^m$, which ensures that after a success both players play R. Second, we construct the two players' average payoff function in the purported MPE, using an approach similar to the proof of Proposition 4; this determines $\bar{p}_1$ and $\bar{p}_2$. The 'jump-benefit' at $\bar{p}_1$ is such that both players are indeed indifferent between R and S at that belief. Third, an intermediate-value argument establishes the existence of a belief p_s such that the strategies described in the proposition give rise to payoff functions compatible with the average payoff constructed in the first step. Finally, we derive monotonicity properties for the payoff functions which allow us to show that for λ_0 small enough, the two strategies are mutually best responses. ■

We can improve the average equilibrium payoff for the two players by considering an MPE with an intermediate phase where the overall intensity of experimentation falls gradually from 2 to 1, rather than dropping in one step.

Proposition 7 (Best MPE with a last experimenter for $N = 2$) *For λ_0 sufficiently close to 0, the two-player experimentation game admits a Markov perfect equilibrium with the following properties. There are three cut-offs, $p_1^\sharp$, $p_2^\ddagger$ and $p_2^\sharp$, and a belief, $p_s^\sharp$, with $p_2^* < p_1^\sharp < p_s^\sharp < p_2^\ddagger < p_2^\sharp < p^m$, such that: on $[p_2^\ddagger, 1]$, both players have a common payoff function; on $[p_2^\sharp, 1]$, both players play R; on $]p_2^\ddagger, p_2^\sharp[$, both players allocate identical fractions of the unit resource to R, and this fraction lies strictly between $\frac{1}{2}$ and 1; on $]p_s^\sharp, p_2^\ddagger]$, player 1 plays S and player 2 plays R; on $]p_1^\sharp, p_s^\sharp]$, player 1 plays R and player 2 plays S; on $[0, p_1^\sharp]$, they both play S. This equilibrium achieves the highest average payoff of all two-player MPE with a last experimenter, and its average payoff is strictly higher than that of the symmetric MPE at all beliefs in $]p_1^\sharp, 1[$.*

PROOF: The construction of the equilibrium is the same as in the proof of Proposition 6 except for the fact that the average (and each players') payoff function solves the indifference ODDE (8) between the diagonals $\mathcal{D}_{1/2}$ and $\mathcal{D}_1$ (the cut-off $p_2^\ddagger$ is the belief where this payoff function crosses $\mathcal{D}_{1/2}$). We therefore omit the details.

The statements about average payoffs are shown as follows. ■

The idea that each player plays R on precisely one interval of beliefs when the overall intensity of experimentation is one generalizes to more than two players. This allows us to construct asymmetric equilibria for an arbitrary number of players.

Proposition 8 (Asymmetric equilibrium for arbitrary N) *For λ_0 sufficiently close to 0, the N-player experimentation game admits a Markov perfect equilibrium with three cut-offs, $p^\sharp_{1|N}$, $p^\ddagger_N$ and $p^\sharp_N$, where $p^*_N < p^\sharp_{1|N} < p^\ddagger_N < p^\sharp_N < p^m$, such that: on $[p^\ddagger_N, 1]$, all players have a common payoff function; on $[p^\sharp_N, 1]$, all players play R; on $]p^\ddagger_N, p^\sharp_N[$, all players allocate identical fractions of the unit resource to R, and this fraction lies strictly between $\frac{1}{N}$ and 1; the interval $]p^\sharp_{1|N}, p^\ddagger_N]$ splits into N subintervals $I_1, \ldots, I_N$, with player n using R on subinterval I_n, and S on all others; on $[0, p^\sharp_{1|N}]$, all players play S. The average payoff in this equilibrium is strictly higher than in the symmetric MPE at all beliefs in $]p^\sharp_{1|N}, 1[$.*

Proof: The construction of the equilibrium is a straightforward generalization of the corresponding steps in the proofs of Propositions 6 and 7. Now, the average (and each players') payoff function solves the indifference ODDE (8) between the diagonals $\mathcal{D}_{1-1/N}$ and $\mathcal{D}_1$, and the cut-off $p^\ddagger_N$ is the belief where this payoff function crosses $\mathcal{D}_{1-1/N}$. The comparison with the symmetric MPE is obtained precisely as in the proof of Proposition 7. Therefore we again omit the details. ■

7 Concluding Remarks

The asymmetric equilibria that we constructed in the Poisson framework raise the question whether similar equilibria exist in the Brownian model of Bolton and Harris (1999). There are at least two possible approaches to this question. One could try to adapt the elementary constructive method that we used here. Alternatively, one could study a sequence of Poisson models with increasing intensities that converges in distribution to a model with Brownian noise. We intend to explore both avenues in future work.

Appendix

Proof of Proposition 4

We first show that there is at most one symmetric MPE. Suppose therefore that we have two symmetric equilibria with different payoff functions u and $\hat{u}$, respectively. Without loss of generality, let $u - \hat{u}$ assume a strictly positive global maximum at the belief p, which by necessity must lie in the open unit interval. At this belief, the left-hand derivatives satisfy $u'(p-) \geq \hat{u}'(p-)$, and $u(j(p)) - \hat{u}(j(p)) \leq u(p) - \hat{u}(p)$, so $b(p, u) \leq b(p, \hat{u})$. We cannot have both $u(p)$ and $\hat{u}(p)$ above $\mathcal{D}_{N-1}$ since in this region both u and $\hat{u}$ are of the form (3) and the difference $u - \hat{u}$ is strictly decreasing to the right of $\mathcal{D}_{N-1}$. Further, if $u(p)$ is above $\mathcal{D}_{N-1}$ and $\hat{u}(p)$ is on or below, then $b(p, u) > c(p) = b(p, \hat{u})$ in contradiction to what we derived before. Consequently, we must have both $u(p)$ and $\hat{u}(p)$ on or below $\mathcal{D}_{N-1}$, so $b(p, u) = c(p) = b(p, \hat{u})$. This in turn yields $u(j(p)) - \hat{u}(j(p)) = u(p) - \hat{u}(p)$, so the difference $u - \hat{u}$ is also at its maximum at the belief $j(p)$. Iterating the argument until we get to the right of p^m (and hence to the right of $\mathcal{D}_{N-1}$), we obtain the desired contradiction.

Turning to the construction of a symmetric MPE, let $\bar{p}_{N,N-1}$ denote the belief where the graph of V_N^* cuts $\mathcal{D}_{N-1}$, and $\bar{p}_{1,N-1}$ denote the belief where the graph of V_1^* cuts $\mathcal{D}_{N-1}$. By continuity, there is an open interval $I \supset [\bar{p}_{N,N-1}, \bar{p}_{1,N-1}]$ such that for all $\bar{p} \in I$, the unique solution to (1) that crosses $\mathcal{D}_{N-1}$ at the belief $\bar{p}$ has strictly positive slope there.

Fix a belief $\bar{p} \in I$ and let $(\bar{p}, \bar{u})$ be the corresponding point on the diagonal $\mathcal{D}_{N-1}$. On $[\bar{p}, 1]$, we define $u^{(0)}$ as the unique solution to (1) that assumes the value $\bar{u}$ at belief $\bar{p}$. Now consider the ordinary differential equation

$$\Delta\lambda\, p(1-p)u'(p) + \lambda(p)u(p) = r\lambda(p)h - rs + \lambda(p)u^{(0)}(j(p)). \tag{A.1}$$

Standard results imply that this ODE has a unique solution $u^{(1)}$ on $[j^{-1}(\bar{p}), \bar{p}]$ with $u^{(1)}(\bar{p}) = u^{(0)}(\bar{p})$ and, by construction, $(u^{(1)})'(\bar{p}) = (u^{(0)})'(\bar{p})$.

Iterating this step, we construct functions $u^{(i+1)}$ defined on $[j^{-(i+1)}(\bar{p}), j^{-i}(\bar{p})]$ for $i = 1, 2, 3, \ldots$ by choosing $u^{(i+1)}$ as the unique solution of the ODE

$$\Delta\lambda\, p(1-p)u'(p) + \lambda(p)u(p) = r\lambda(p)h - rs + \lambda(p)u^{(i)}(j(p)) \tag{A.2}$$

subject to the condition $u^{(i+1)}(j^{-i}(\bar{p})) = u^{(i)}(j^{-i}(\bar{p}))$. Setting $u_{\bar{p}}(p) = u^{(i)}(p)$ whenever $j^{-(i+1)}(\bar{p}) \leq p < j^{-i}(\bar{p})$, we thus obtain a function $u_{\bar{p}}$ of class C^1 on $]0, 1]$ that solves (8) to the left of $\bar{p}$, and (1) to the right of $\bar{p}$. Standard results imply that $u_{\bar{p}}$ depends in a continuous fashion on $\bar{p}$. In particular, $M(\bar{p})$, the minimum of $u_{\bar{p}}$ on $[p_N^*, p^m]$, is continuous in $\bar{p}$.

For $\bar{p} \in I$ with $\bar{p} < \bar{p}_{N,N-1}$, the function $u_{\bar{p}}$ lies strictly above V_N^* on at least $[\bar{p}, 1[$. If $u_{\bar{p}}$ and V_N^* assumed the same value at some belief $p_\ell \in [p_N^*, \bar{p}[$, then the restriction of $u_{\bar{p}} - V_N^*$ to $[p_\ell, 1]$ would have a strictly positive global maximum at some belief $p_r \in]p_\ell, 1[$. In fact, we would have $p_r \in]p_\ell, \bar{p}[$ since $u_{\bar{p}} - V_N^*$, being the difference of two functions of the form (3), has a strictly negative first derivative on $[\bar{p}, 1[$. As $(u_{\bar{p}})'(p_r) = (V_N^*)'(p_r)$ and $u_{\bar{p}}(j(p_r)) - V_N^*(j(p_r)) \leq u_{\bar{p}}(p_r) - V_N^*(p_r)$, we would thus have $b(p_r, V_N^*) \geq b(p_r, u_{\bar{p}}) = c(p_r)$, hence $V_N^*(p_r) = s + Nb(p_r, V_N^*) - c(p_r) \geq s + (N-1)c(p_r)$, which is inconsistent with the fact that V_N^* is strictly below $\mathcal{D}_{N-1}$ at p_r. Consequently, $u_{\bar{p}}$ lies strictly above V_N^* on $[p_N^*, 1[$.

By continuity, $\bar{u}_N$, the function $u_{\bar{p}}$ obtained for $\bar{p} = \bar{p}_{N,N-1}$, lies weakly above V_N^* on $[p_N^*, 1]$. While the two functions are identical on $[\bar{p}_{N,N-1}, 1]$ by construction, they cannot be identical on the whole of $[p_N^*, \bar{p}_{N,N-1}[$ as V_N^* does not solve (A.1) immediately to the left of $\bar{p}_{N,N-1}$, for example. Arguing exactly as in the previous paragraph, we see that the restriction of $\bar{u}_N - V_N^*$ to $[p_N^*, 1]$ must assume its strictly positive global maximum at p_N^*. This establishes $\bar{u}_N(p_N^*) > V_N^*(p_N^*) = s$. As $V_N^*(p) > s$ for $p > p_N^*$, we thus have $\bar{u}_N > s$ on $[p_N^*, 1]$, hence $M(\bar{p}_{N,N-1}) > s$.

For $\bar{p} \in I$ with $\bar{p} > \bar{p}_{1,N-1}$, the function $u_{\bar{p}}$ lies strictly below V_1^* in a neighbourhood of $\bar{p}$. If $u_{\bar{p}}$ and V_1^* assumed the same value at some belief $p_\ell \in [p_1^*, \bar{p}[$, then the restriction of $V_1^* - u_{\bar{p}}$ to $[p_\ell, 1]$ would have a strictly positive global maximum at a belief $p_r \in]p_\ell, 1[$. As $(V_1^*)'(p_r) = (u_{\bar{p}})'(p_r)$ and $V_1^*(j(p_r)) - u_{\bar{p}}(j(p_r)) \leq V_1^*(p_r) - u_{\bar{p}}(p_r)$, we would thus have $b(p_r, u_{\bar{p}}) \geq b(p_r, V_1^*)$. As $s < V_1^*(p_r) = s + b(p_r, V_1^*) - c(p_r)$, this would imply $b(p_r, u_{\bar{p}}) > c(p_r)$ and $p_r > \bar{p}$. But then $u_{\bar{p}}(p_r) = s + Nb(p_r, u_{\bar{p}}) - c(p_r) > s + b(p_r, V_1^*) - c(p_r) = V_1^*(p_r)$, which is a contradiction. Consequently, $u_{\bar{p}}$ lies strictly below V_1^* on $[p_1^*, \bar{p}]$.

By continuity, $\bar{u}_{1,N}$, the function $u_{\bar{p}}$ obtained for $\bar{p} = \bar{p}_{1,N-1}$, lies weakly below V_1^* on $[p_1^*, \bar{p}_{1,N-1}]$. While the two functions are identical at $\bar{p}_{1,N-1}$ by construction, they cannot be identical on the whole of $[p_1^*, \bar{p}_{1,N-1}[$. Arguing exactly as in the previous paragraph, we see that the restriction of $V_1^* - \bar{u}_{1,N}$ to $[p_1^*, 1]$ must assume its strictly positive global maximum at p_1^*. In particular, $\bar{u}_{1,N}(p_1^*) < V_1^*(p_1^*) = s$, hence $M(\bar{p}_{1,N-1}) < s$.

So there exists a $p_N^\dagger \in]\bar{p}_{N,N-1}, \bar{p}_{1,N-1}[$ such that $M(p_N^\dagger) = s$. With $u^\dagger$ denoting the solution $u_{\bar{p}}$ corresponding to $\bar{p} = p_N^\dagger$, let $\tilde{p}_N$ be the highest belief in $[p_N^*, p^m]$ at which $u^\dagger$ assumes the value s. By construction, $\tilde{p}_N < p_N^\dagger < p^m$. Define the function $W_N^\dagger$ by $W_N^\dagger(p) = s$ on $[0, \tilde{p}_N]$ and by $W_N^\dagger(p) = u^\dagger(p) > s$ on $]\tilde{p}_N, 1]$. This is the common payoff function when all players use the strategy $k_N^\dagger$ described in the proposition. As a consequence, $W_N^\dagger \leq V_N^*$ and in particular $\tilde{p}_N \geq p_N^*$.

If we had $\tilde{p}_N = p_N^*$, then $W_N^\dagger(p_N^*) = s = V_N^*(p_N^*)$, $W_N^\dagger(j(p_N^*)) \leq V_N^*(j(p_N^*))$ and $(W_N^\dagger)'(p_N^*-) = 0 = (V_N^*)'(p_N^*)$, implying $b(p_N^*, V_N^*) \geq b(p_N^*, W_N^\dagger)$. As $b(p_N^*, V_N^*) = c(p_N^*)/N$, $b(p_N^*, W_N^\dagger) = c(p_N^*)$ and $c(p_N^*) > 0$, this is a contradiction. So we have $p_N^* < \tilde{p}_N < p^m$, hence $(W_N^\dagger)'(\tilde{p}_N+) = (u^\dagger)'(\tilde{p}_N) = 0$ because the minimum of $u^\dagger$ on $[p_N^*, p^m]$ is achieved at an interior point. Thus, the function $W_N^\dagger$ is of class C^1.

It is straightforward to check from the explicit representation of $W_N^\dagger$ above $\mathcal{D}_{N-1}$ that this function is strictly convex and strictly increasing on $[p_N^\dagger, 1]$. Suppose $W_N^\dagger$ is not strictly increasing on $[\tilde{p}_N, p_N^\dagger]$. Then it must assume both a local minimum and a local maximum in the interior of that interval, and there exist beliefs $p' < p''$ in $]\tilde{p}_N, p_N^\dagger[$ such that $(W_N^\dagger)'(p') = (W_N^\dagger)'(p'') = 0$, $W_N^\dagger(p') \geq W_N^\dagger(p'')$, and $W_N^\dagger$ is weakly decreasing on $[p', p'']$ and strictly increasing on $[p'', 1]$. We now have $b(p', W_N^\dagger) = \lambda(p')[W_N^\dagger(j(p')) - W_N^\dagger(p')]/r = c(p') > 0$, hence $W_N^\dagger(j(p')) > W_N^\dagger(p')$ and $j(p') > p''$. As a consequence, $W_N^\dagger(j(p'')) > W_N^\dagger(j(p'))$ and $b(p'', W_N^\dagger) = \lambda(p'')[W_N^\dagger(j(p'')) - W_N^\dagger(p'')]/r > \lambda(p')[W_N^\dagger(j(p')) - W_N^\dagger(p')]/r = c(p') > c(p'')$, which is a contradiction. This establishes that $W_N^\dagger$ is strictly increasing on $[\tilde{p}_N, 1]$, and $k_N^\dagger$ is strictly increasing on $[\tilde{p}_N, p_N^\dagger]$.

We thus have $b(p, W_N^\dagger) > c(p)$ on $]p_N^\dagger, 1]$, $b(p, W_N^\dagger) = c(p)$ on $[\tilde{p}_N, p_N^\dagger]$, and, because of the monotonicity of $W_N^\dagger$ on $[\tilde{p}_N, 1]$, $b(p, W_N^\dagger) < c(p)$ on $[0, \tilde{p}_N[$. So all players using the strategy $k_N^\dagger$ constitutes an equilibrium. Finally, $\tilde{p}_N < p_1^*$ by Proposition 3. ■

Proof of Corollary 2

With $u^{(0)}(p) = \lambda_1 hp + \lambda_0 h(1-p) + C^{(0)}\,(1-p)\,\Omega(p)^\mu$ (see (3)), we seek a sequence of functions $u^{(i+1)}$ for $i = 0, 1, \ldots$, defined recursively as solutions to the ODE (A.2). Let $\alpha = \lambda_0/\Delta\lambda$, and, for $i \geq 0$, let

$$u^{(i)}(p) = d_1^{(i)}p + d_0^{(i)}(1-p) + m^{(i)}(1-p)\Omega(p)^\mu + (1-p)\Omega(p)^\alpha \sum_{n=0}^{i-1} l^{(i-n)} \left(\ln\left[(\lambda_0/\lambda_1)^{n-1}\,\Omega(p)\right]\right)^n$$

where $d_1^{(i)}, d_0^{(i)}, m^{(i)}, l^{(i-n)}$ are constants to be determined – we will show that the functions $u^{(i)}$ form just such a sequence. Clearly we need

$$d_1^{(0)} = \lambda_1 h, \quad d_0^{(0)} = \lambda_0 h, \quad \text{and} \quad m^{(0)} = C^{(0)}$$

with $C^{(0)}$ being the constant that fixes payoffs above the diagonal where everyone plays R. The final (summed) term in the above equation defining $u^{(i)}$ is vacuous for $i = 0$.

First note that:

$$\begin{aligned} u^{(i)}(j(p)) &= d_1^{(i)}\frac{\lambda_1}{\lambda(p)}\,p + d_0^{(i)}\frac{\lambda_0}{\lambda(p)}\,(1-p) + m^{(i)}\frac{\lambda_0}{\lambda(p)}\left(\frac{\lambda_0}{\lambda_1}\right)^\mu (1-p)\Omega(p)^\mu \\ &\quad + \frac{\lambda_0}{\lambda(p)}\left(\frac{\lambda_0}{\lambda_1}\right)^\alpha (1-p)\Omega(p)^\alpha \sum_{n=0}^{i-1} l^{(i-n)} \left(\ln\left[(\lambda_0/\lambda_1)^n\,\Omega(p)\right]\right)^n \end{aligned}$$

in which case the right-hand side of (A.2) becomes:

$$G^{(i)}(p) = D_1^{(i)}p + D_0^{(i)}(1-p) + M^{(i)}(1-p)\Omega(p)^\mu + (1-p)\Omega(p)^\alpha \sum_{n=0}^{i-1} L^{(i-n)} \left(\ln\left[(\lambda_0/\lambda_1)^n\,\Omega(p)\right]\right)^n$$

where

$$D_1^{(i)} = d_1^{(i)}\lambda_1 + r(\lambda_1 h - s), \quad D_0^{(i)} = d_0^{(i)}\lambda_0 + r(\lambda_0 h - s)$$

and

$$M^{(i)} = m^{(i)}\lambda_0\left(\lambda_0/\lambda_1\right)^\mu, \quad L^{(i-n)} = l^{(i-n)}\lambda_0\left(\lambda_0/\lambda_1\right)^\alpha.$$

The homogeneous equation, $\Delta\lambda\, p(1-p)u'(p) + \lambda(p)u(p) = 0$, has the solution

$$u_0(p) = (1-p)\Omega(p)^\alpha.$$

Using the method of variation of constants, we now write $u(p) = a(p)u_0(p)$ so that

$$\Delta\lambda\, p(1-p)u'(p) + \lambda(p)u(p) = \Delta\lambda\, p(1-p)u_0(p)a'(p).$$

The ODE thus transforms into the following equation for the first derivative of the unknown function a:

$$\begin{aligned} \Delta\lambda\, a'(p) &= \frac{G^{(i)}(p)}{p(1-p)\,u_0(p)} \\ &= D_1^{(i)}\,\Omega(p)^{-\alpha}(1-p)^{-2} + D_0^{(i)}\,\Omega(p)^{-\alpha+1}(1-p)^{-2} + M^{(i)}\,\Omega(p)^{\mu-\alpha+1}(1-p)^{-2} \\ &\quad + \Omega(p)(1-p)^{-2}\sum_{n=0}^{i-1} L^{(i-n)}\left(\ln\left[(\lambda_0/\lambda_1)^n\,\Omega(p)\right]\right)^n. \end{aligned}$$

Make the substitution $\omega = \Omega(p)$ and define $A(\omega) = a(p)$, so $a'(p) = -A'(\omega)/p^2$. Then

$$-\Delta\lambda\, A'(\omega) = D_1^{(i)}\,\omega^{-\alpha-2} + D_0^{(i)}\,\omega^{-\alpha-1} + M^{(i)}\,\omega^{\mu-\alpha-1} + \omega^{-1}\sum_{n=0}^{i-1} L^{(i-n)}\left(\ln\left[(\lambda_0/\lambda_1)^n\,\omega\right]\right)^n,$$

so

$$
\begin{aligned}
A(\omega) \quad = \quad & \frac{D_1^{(i)}}{\lambda_1}\,\omega^{-\alpha-1} + \frac{D_0^{(i)}}{\lambda_0}\,\omega^{-\alpha} + \frac{M^{(i)}}{\lambda_0 - \mu\Delta\lambda}\,\omega^{\mu-\alpha} \\
& - \sum_{n=0}^{i-1} \frac{L^{(i-n)}}{(n+1)\Delta\lambda}\,(\ln\left[(\lambda_0/\lambda_1)^n\,\omega\right])^{n+1} + C^{(i+1)},
\end{aligned}
$$

where $C^{(i+1)}$ is a constant of integration (and assuming $\mu \neq \alpha$, else we have another logarithmic term). Multiplying by $u_0(p) = (1-p)\,\omega^\alpha$ and substituting $\omega = \Omega(p)$ leads to

$$
\begin{aligned}
u^{(i+1)}(p) \quad = \quad & \frac{D_1^{(i)}}{\lambda_1}\,p + \frac{D_0^{(i)}}{\lambda_0}\,(1-p) + \frac{M^{(i)}}{\lambda_0 - \mu\Delta\lambda}\,(1-p)\Omega(p)^\mu \\
& + (1-p)\Omega(p)^\alpha \sum_{n=1}^{i} \frac{-L^{(i+1-n)}}{n\,\Delta\lambda}\left(\ln\left[(\lambda_0/\lambda_1)^{n-1}\,\Omega(p)\right]\right)^n + (1-p)\Omega(p)^\alpha C^{(i+1)}.
\end{aligned}
$$

The above iterative step shows that

$$
d_1^{(i+1)} = d_1^{(i)} + \tfrac{r}{\lambda_1}(\lambda_1 h - s), \quad d_0^{(i+1)} = d_0^{(i)} + \tfrac{r}{\lambda_0}(\lambda_0 h - s), \quad \text{and} \quad m^{(i+1)} = m^{(i)}\left(\frac{\lambda_0\,(\lambda_0/\lambda_1)^\mu}{\lambda_0 - \mu\Delta\lambda}\right)
$$

and so, in general,

$$
d_1^{(i)} = \lambda_1 h + \tfrac{r}{\lambda_1}(\lambda_1 h - s)\,i\,, \quad d_0^{(i)} = \lambda_0 h + \tfrac{r}{\lambda_0}(\lambda_0 h - s)\,i\,, \quad \text{and} \quad m^{(i)} = C^{(0)}\left(\frac{\lambda_0\,(\lambda_0/\lambda_1)^\mu}{\lambda_0 - \mu\Delta\lambda}\right)^i.
$$

After a little algebra, we find that the constants in the summation are given by:

$$
l^{(i-n)} = \frac{C^{(i-n)}}{n!}\left(-\frac{\lambda_0\,(\lambda_0/\lambda_1)^\alpha}{\Delta\lambda}\right)^n \qquad \text{for } n = 0, \ldots, i-1.
$$

The constants $C^{(i-n)}$ ($n = 0, \ldots i-1$) are chosen to ensure continuity. In particular, writing $\hat{\jmath}^{-i}$ for $j^{-i}(p^\dagger)$, $C^{(i+1)}$ is chosen such that $u^{(i+1)}(\hat{\jmath}^{-i}) = u^{(i)}(\hat{\jmath}^{-i})$ for $i \geq 0$, and satisfies:

$$
\begin{aligned}
& C^{(i+1)}\,(1-\hat{\jmath}^{-i})\Omega(\hat{\jmath}^{-i})^\alpha \\
& = \; -\tfrac{r}{\lambda_1}(\lambda_1 h - s)\,\hat{\jmath}^{-i} - \tfrac{r}{\lambda_0}(\lambda_0 h - s)\,(1-\hat{\jmath}^{-i}) \\
& \quad + C^{(0)}\left(1 - \frac{\lambda_0\,(\lambda_0/\lambda_1)^\mu}{\lambda_0 - \mu\Delta\lambda}\right)\left(\frac{\lambda_0\,(\lambda_0/\lambda_1)^\mu}{\lambda_0 - \mu\Delta\lambda}\right)^i (1-\hat{\jmath}^{-i})\Omega(\hat{\jmath}^{-i})^\mu \\
& \quad + \left\{\sum_{n=0}^{i-1} C^{(i-n)}\left[\frac{1}{n!}\left(-\frac{\lambda_0\,(\lambda_0/\lambda_1)^\alpha}{\Delta\lambda}\ln\left[(\lambda_0/\lambda_1)^{n-1}\,\Omega(\hat{\jmath}^{-i})\right]\right)^n \right.\right. \\
& \qquad\qquad \left.\left. - \frac{1}{(n+1)!}\left(-\frac{\lambda_0\,(\lambda_0/\lambda_1)^\alpha}{\Delta\lambda}\ln\left[(\lambda_0/\lambda_1)^n\,\Omega(\hat{\jmath}^{-i})\right]\right)^{n+1}\right]\right\}(1-\hat{\jmath}^{-i})\Omega(\hat{\jmath}^{-i})^\alpha.
\end{aligned}
$$

■

Proof of Proposition 6

We choose λ_0 small enough so that $j(p_2^*) \geq p^m$.[5] As in the proof of Proposition 4, let $\bar{p}_{2,1}$ denote the belief where the graph of V_2^* cuts $\mathcal{D}_1$, $\bar{p}_{1,1}$ the belief where the graph of V_1^* cuts $\mathcal{D}_1$, and $I \supset [\bar{p}_{2,1}, \bar{p}_{1,1}]$ an open interval such that for all $\bar{p} \in I$, the unique solution to (1) with $N = 2$ that crosses $\mathcal{D}_1$ at the belief $\bar{p}$ has strictly positive slope there.

Fix a belief $\bar{p} \in I$. On $[\bar{p}, 1]$, we define $u_{\bar{p}}$ as the unique solution to (1) with $N = 2$ that starts on $\mathcal{D}_1$ at $\bar{p}$. On $[p_2^*, \bar{p}[$, we define $u_{\bar{p}}$ as the unique solution to the ODE

$$\Delta\lambda\, p(1-p)u'(p) + [\lambda(p) + r]u(p) = \tfrac{r}{2}[s + \lambda(p)h] + \lambda(p)u_{\bar{p}}(j(p))$$

that ends on $\mathcal{D}_1$ at $\bar{p}$. By construction, $u_{\bar{p}}$ is continuous, $u_{\bar{p}}(p) = s + 2b(p, u_{\bar{p}}) - c(p)$ on $]\bar{p}, 1]$, and $u_{\bar{p}}(p) = s + b(p, u_{\bar{p}}) - \frac{1}{2}c(p)$ on $]p_2^*, \bar{p}]$.

Proceeding as in the proof of Proposition 4, one establishes the existence of a $\bar{p} \in]\bar{p}_{2,1}, \bar{p}_{1,1}[$ such that the corresponding function $u_{\bar{p}}$ has an interior global minimum equal to s at some belief $\breve{p} \in]p_2^*, \bar{p}[$. As $u'_{\bar{p}}(\breve{p}) = 0$, we have $\lambda(\breve{p})\,[u_{\bar{p}}(j(\breve{p})) - s]/r = b(\breve{p}, u_{\bar{p}}) = \frac{1}{2}c(\breve{p}) < c(\breve{p})$. For $\bar{p} = p^m$, on the other hand, the corresponding function $u_{\bar{p}}$ assumes value s at p^m. As its slope there is strictly positive and $c(p^m) = 0$, we have $\lambda(p^m)\,[u_{\bar{p}}(j(p^m)) - s]/r > b(p^m, u_{\bar{p}}) = \frac{1}{2}c(p^m) = c(p^m)$. By continuity of $u_{\bar{p}}$ with respect to $\bar{p}$, there exists $\bar{p}_2 \in]\bar{p}_{2,1}, p^m[$ such that the function $u_{\bar{p}}$ obtained for $\bar{p} = \bar{p}_2$ has the following property: there is a belief $\bar{p}_1 \in]\breve{p}, \bar{p}_2[$ such that $u_{\bar{p}}(\bar{p}_1) = s$, $u_{\bar{p}}(p) > s$ for $p > \bar{p}_1$, and $\lambda(\bar{p}_1)\,[u_{\bar{p}}(j(\bar{p}_1)) - s]/r = c(\bar{p}_1)$.

We define a function $\bar{u}$ on $[0, 1]$ by taking $\bar{u}$ equal to the function $u_{\bar{p}}$ just determined on $[\bar{p}_1, 1]$, and $\bar{u} = s$ everywhere else. (This is the average payoff function associated with the strategies described in the proposition.) Next, we define

$$\bar{b}(p, u) = [\lambda(p)\,(\bar{u}(j(p)) - u(p)) - \Delta\lambda\, p(1-p)u'(p)]/r$$

for any left-differentiable real-valued function u on $]0, 1]$. (This is the benefit of experimentation when the value after a success is given by the payoff function $\bar{u}$.) For arbitrary but fixed $p_s \in [\bar{p}_1, \bar{p}_2]$, we consider the continuous function u_1 on $[\bar{p}_1, \bar{p}_2]$ that is uniquely determined by the following properties: $u_1(\bar{p}_1) = s$; $u_1(p) = s + \bar{b}(p, u_1) - c(p)$ on $]\bar{p}_1, p_s]$; and $u_1(p) = s + \bar{b}(p, u_1)$ on $]p_s, \bar{p}_2]$.

For $p_s = \bar{p}_1$, we have $u_1'(\bar{p}_1+) > \bar{u}'(\bar{p}_1+)$ since $\bar{b}(p, u_1)$ and $\bar{b}(p, \bar{u})$ tend to 0 and $c(\bar{p}_1)/2$, respectively, as p tends to $\bar{p}_1$ from above, while $\lambda(p)\,[\bar{u}(j(p)) - u_1(p)]/r$ and $\lambda(p)\,[\bar{u}(j(p)) - \bar{u}(p)]/r$ tend to one and the same limit. Thus, $u_1(p) > \bar{u}(p)$ immediately to the right of $\bar{p}_1$. Now, the inequality $u_1(\bar{p}_2) \leq \bar{u}(\bar{p}_2)$ would imply the existence of a belief $p' \in]\bar{p}_1, \bar{p}_2[$ such that $u_1(p') = \bar{u}(p')$ and $u_1'(p') \leq \bar{u}'(p')$. As $u_1(p') = s + \bar{b}(p', u_1)$ and $\bar{u}(p') = s + b(p', \bar{u}) - \frac{1}{2}c(p')$, we would have $\frac{1}{2}c(p') = b(p', \bar{u}) - \bar{b}(p', u_1) = -\Delta\lambda\, p'(1-p')[\bar{u}'(p') - u_1'(p')]/r \leq 0$ – a contradiction. So $p_s = \bar{p}_1$ implies $u_1(\bar{p}_2) > \bar{u}(\bar{p}_2)$.

In exactly the same way, one shows that $p_s = \bar{p}_2$ implies $u_1(\bar{p}_2) < \bar{u}(\bar{p}_2)$. By continuity, there is a belief $p_s \in]\bar{p}_1, \bar{p}_2[$ such that the function u_1 obtained for this p_s satisfies $u_1(\bar{p}_2) = \bar{u}(\bar{p}_2)$, that is, meets $\bar{u}$ on $\mathcal{D}_1$. We extend the domain of definition of this function by setting $u_1(p) = \bar{u}(p)$ on $[0, \bar{p}_1[$ and $]\bar{p}_2, 1]$. We also define a function u_2 on $[0, 1]$ by setting $u_2(p) = 2\bar{u}(p) - u_1(p)$. Clearly, u_1 and u_2 are the payoff functions associated with the pair of strategies described in the proposition.

Both functions are strictly increasing on $[\bar{p}_1, 1]$. The explicit representation (3) makes this obvious on $[\bar{p}_2, 1]$. Suppose therefore that u_1 is not strictly increasing on $[\bar{p}_1, \bar{p}_2]$. Then there

[5] A necessary and sufficient condition is that $\mu_2/(\mu_2+1) \geq \lambda_0/\lambda_1$, or equivalently that $\mu_2 \geq \lambda_0/\Delta\lambda$. Using (2), this holds iff $\lambda_0(\lambda_0/\lambda_1)^{\lambda_0/\Delta\lambda} \leq r/2$; clearly, since $\lambda_0/\lambda_1 < 1$ and $\lambda_0/\Delta\lambda \geq 0$, a sufficient condition for this last inequality is that $\lambda_0 \leq r/2$.

exist beliefs $p' < p''$ in $[\bar{p}_1, \bar{p}_2[$ such that $u_1'(p'-) \geq 0$, $u_1'(p''-) \leq 0$, u_1 is weakly decreasing on $[p', p'']$, and strictly increasing on $[p'', 1]$. As $j(p') > \bar{p}_2$, this implies in particular that $b(p', u_1) < b(p'', u_1)$. There are three cases now, each of them leading to a contradiction with u_1 being weakly decreasing on $[p', p'']$: (i) if $p'' \leq p_s$, then $u_1(p') = s + b(p', u_1) - c(p') < s + b(p'', u_1) - c(p'') = u_1(p'')$; (ii) if $p' > p_s$, then $u_1(p') = s + b(p', u_1) < s + b(p'', u_1) = u_1(p'')$; (iii) if $p' \leq p_s < p''$, then $u_1(p') = s + b(p', u_1) - c(p') < s + b(p'', u_1) = u_1(p'')$.

Suppose next that there exist beliefs $p' < p''$ in $[\bar{p}_1, \bar{p}_2[$ such that $u_2'(p'-) \geq 0$, $u_2'(p''-) \leq 0$, u_2 is weakly decreasing on $[p', p'']$, and strictly increasing on $[p'', 1]$. Again, this implies $b(p', u_2) < b(p'', u_2)$. Now, there are four cases, the first three of which quickly lead to a contradiction with u_2 being weakly decreasing on $[p', p'']$: (i) if $p'' \leq p_s$, then $u_2(p') = s + b(p', u_2) < s + b(p'', u_2) = u_2(p'')$; (ii) if $p' > p_s$, then $u_2(p') = s + b(p', u_2) - c(p') < s + b(p'', u_2) - c(p'') = u_2(p'')$; (iii) if $p' < p_s < p''$, then $b(p', u_2) < b(p_s, u_2)$ and so $u_2(p') = s + b(p', u_2) < s + b(p_s, u_2) = u_2(p_s)$; (iv) if $p' = p_s$, then $u_2(p') = s + b(p', u_2)$ and $u_2(p'') = s + b(p'', u_2) - c(p'')$, which is very well compatible with $b(p', u_2) < b(p'', u_2)$ and $u_2(p') \geq u_2(p'')$. So either u_2 is strictly increasing on $[\bar{p}_1, \bar{p}_2]$, or it has a local maximum (with a kink) in p_s. By necessity, this local maximum is below $\mathcal{D}_1$.

The two strategies are now easily seen to be mutually best responses. First, both payoff functions are above $\mathcal{D}_1$ on $]\bar{p}_2, 1]$. Second, they are above s and below $\mathcal{D}_1$ on $]\bar{p}_1, \bar{p}_2[$. Third, we have $b(\bar{p}_1, u_n) = \lambda(\bar{p}_1)\,[u_n(j(\bar{p}_1)) - s]/r = \lambda(\bar{p}_1)\,[\bar{u}(j(\bar{p}_1)) - s]/r = c(\bar{p}_1)$ for $n = 1, 2$ as the left-hand derivative of u_n at $\bar{p}_1$ is zero. Both functions are strictly increasing on $[\bar{p}_2, 1]$. Consequently, for $p \in [p_2^*, \bar{p}_1[$, $u_n(j(p))$ is strictly increasing, and also $c(p)$ is strictly decreasing, therefore $b(p, u_n) < c(p)$ on this interval. As u_1 is strictly increasing on $[\bar{p}_1, 1]$, the same argument actually yields $b(p, u_1) < c(p)$ on the whole of $[0, \bar{p}_1[$. Finally, for $p \in [0, p_2^*[$, $b(p, u_2) \leq \lambda(p)\,[\lambda_1 h - s]/r < \lambda(p_2^*)\,[\lambda_1 h - s]/r$ and this last term tends to $c(p_2^*)/2$ as λ_0 tends to 0; since $c(p_2^*)/2 < c(p)/2 < c(p)$, we have $b(p, u_2) < c(p)$ on this interval as well, for λ_0 sufficiently small. ■

References

BELLMAN, R. and K.L. COOKE (1963): *Differential-Difference Equations.* New York: Academic Press.

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.

KELLER, G., S. RADY and M. CRIPPS (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, **73**, 39–68.

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

PRESMAN, E.L. and I.N. SONIN (1990): *Sequential Control with Incomplete Information: The Bayesian Approach to Multi-Armed Bandit Problems.* London: Academic Press.